ISSUES IN MODEL VALIDATION

Andrzej Lewandowski
Technical University, Warsaw, Poland

RR-82-37
November 1982

Reprinted from *Angewandte Systemanalyse*, volume 3 number 1 (1982)

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
Laxenburg, Austria

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.

Reprinted with permission from *Angewandte Systemanalyse* 3(1):2–11, 1982.

FOREWORD

During the years 1980–1981 the System and Decision Sciences Area at IIASA has been involved in research clarifying the issue of model validation and verification. Apart from its task force meeting in October 1980 and a summer meeting in 1981 held on this topic, Andrzej Lewandowski, from the Technical University, Warsaw, Poland, has written an article that attempts to classify concepts and issues related to this broad and not yet fully focused area of research.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences Area

Issues in model validation

by A. Lewandowski

Dokumentation

Lewandowski, A.: **Probleme bei der Modellvalidierung.**
Angewandte Systemanalyse Band 3 Heft 1 (1982). S. 2–11, 3 Bilder, 34 Lit.Ang.

Schlagwörter: Validierung/Modellklassen/Methodologie der Validierung

Zusammenfassung

Man ist sich einig, daß die Validierung einer der wichtigsten Schritte bei der Aufstellung eines mathematischen Modells ist. Leider besteht eine Kluft zwischen generellen Empfehlungen für die Validierung und der Modellierungspraxis.
Zur teilweisen Überbrückung dieser Kluft stellt der Autor eine Modellklassifikation vor, welche für die Validierungspraxis von Wichtigkeit ist. Auf der Basis dieser Klassifikation könnte eine Bestimmung solcher Modellklassen möglich sein, für welche die vorhandenen Validierungshilfsmittel geeignet sind.
Außerdem wird eine Übersicht über bestehende Ideen zu einer Methologie der Validierung gegeben.

Issues in model validation

Abstract

There is a common agreement between mathematical modelers that the validation stage is one of the most important ones in the model-building process. Unfortunately, there exists a gap between the general advices and the modeling practice.
To bridge this gap partially, the author introduces a model taxonomy which is important for the validation practice. On the basis of this taxonomy it would be possible to select the groups of models for which the given validation tools are appropriate.
The review of existing ideas dealing with validation methodology is also presented.

Problèmes rencontrés dans la validation des modèles

Résumé

Les mathématiciens sont d'accord pour reconnaître que la validation représente l'une des étapes les plus importantes dans la construction de modèles mathématiques. Malheureusement, il existe un fossé entre les recommandations générales et la pratique de la modélisation.
Afin de combler partiellement ce fossé, l'auteur propose une classification des modèles qui constitue une aide importante dans la procédure de validation. Sur la base de cette classification, il serait possible de sélectionner les groupes de modèles pour lesquels les moyens de validation actuels sont appropriés.
En outre, l'auteur présente un aperçu général des idées existantes sur la méthodologie de la validation.

1 Introduction

It is commonly agreed between modeling methodologists that model validation is one of the most important stages in the model building process. Many papers addressing this subject have been published and an SCS Technical Committee on Model Credibility has been established in order to generalize and summarize the experiences in this field [28]. However, at the present stage of research there are almost no suggestions concerning concrete methods of validation. Practically all authors only discuss definitions of validation – not methods. The number of papers dealing with methods of model validation is also rather limited.
The reason for this gap between methodological consciousness and the practice of model building seems to be obvious – the discussion stays at too high a level of abstraction. In general, all authors consider "model" as a description of reality, and on this level of concretization it is only possible to generate rather general statements, frequently true but without operational meaning. The author of this paper believes that, in order to examine validation methods, it is necessary to specify more precisely the model under consideration, the properties of the model, the modeling techniques, and, most importantly, the purpose of the model.
The aim of this paper, therefore, is to present a classification of models and an analysis of the modeling process from the point of view of model validation. At his stage of the investigation, however, it is not yet possible to design, nor to analyze, methods of validation. Our goal is to design a framework for model validation as a first

and important step in establishing a model validation methodology.

2 Validation: definitions

There are various definitions for model validation, but all are very similar and have been summarized by the SCS Technical Committee on Model Credibility [28]. This set of modeling methodology definitions and concepts is quite precise and clear:

> . . . *(model validation is) substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy constistent with the intended application of the model.*

This definition also coincides very well with the definitions given by, for example, Naylor [18] and Mihram [20]. The most interesting consideration of validation methodology, however, can be found in Mankin et al. [19], where a more formal definition is given:

> . . . *a model is valid if its behaviour corresponds to system behaviour under all conditions of interest. A model is considered invalid if we can devise an experiment in which the model outputs disagree with system measurements within the specified area of interest . . .*

Similar notions have also been investigated by Beck [4]. A somewhat broader notion is that of usefulness: "...a model is useful if it accurately represents some of the system behavior and useless if it does not" (Mankin et al. [19]).

Model validity can be related to model reliability and adequacy:

- reliability is defined as the fraction of the model outputs which corresponds correctly to system outputs;
- adequacy is the fraction of system outputs which can be modeled correctly.

In the definitions formulated above, "model output" should be understood in a rather general sense and "output" means the result of the modeling experiment.

Since the last two concepts have more definite operational meaning and can be relatively easily measured and computed, they can be treated as more practical tools for model testing and choosing between alternative models. These more qualitative model validity measures imply application possibilities of more advanced techniqes, for example, statistical hypothesis testing (Greig [9]). Hence, there is now a good terminological background for model validation in the sense that we know generally what model validation means. There remains open, however, the problem of how to validate a given model.

3 Model attributes

A large number of *model attributes* can be listed, but only three of them seem to be interesting for model validation purposes. The first attribute can be called *model background* which gives information on the natural and behavioral background of the model. This attribute determines to what extent basic consideration and natural laws have been applied when building the model. Hard models with a natural background are built on the basis of well established natural laws, for example, such precise and well-defined concepts as mass or energy balances, variational mechanical principles, etc. In other words, the validity of these models can be judged on the basis of well-known and accepted theories. This type of validity consideration can be called *internal validation*, and consists of checking the preservation of the basic laws which have been used when building the model. Models of electrical circuits, technological processes, and selected environmental problems (water quality) are examples of hard models with natural backgrounds.

At the other end of the spectrum we have soft models with behavioral background. They are formulated on the basis of more inductive analysis of system behavior – without such *a priori* knowledge of natural laws governing the system under consideration. In many important practical cases we must hypothesize when dealing with system behavior, either because of the complexity of the system, large numbers of factors, or because of an insufficient level of basic knowledge dealing with the phenomena being modeled. This situation frequently arises in the modeling of social, environmental, or economic systems. Similar considerations have been performed by Kalman [11]:

> . . . *the usual procedure of making a model of a system is obvious. A catalog of known facts and data is compiled and equations are written down by taking into account all available quantitative information . . . An absolutely essential assumption for this process to work is that the "laws" governing physical phenomena are independent of the system context . . . Oversimplifying a bit, no matter what system is built, who builds it, how it is built, and why it is built, Ohm's law is immutable. The essential feature of economics is that this is simply not so . . . There are no "laws" in economics as this term is understood in physics, because economics is a system-determined science . . .*

Similar concepts of hard and soft models have also been introduced by Beck [4], but his definition is a little bit broader.

The second model attribute relates to the *logical type of the model.* One can consider two types of models –

causal and *descriptive*. Causal models can be built if one can distinguish between cause and effect and the input and output variables in the system consideration. According to Zadeh's [34] terminology, these models should be called "oriented models". Descriptive (or nonoriented) models are built on the basis of correlation analysis, without distinguishing between inputs and outputs. Correlation analysis makes it possible to test the dependence between various variables, but cannot give conclusive evidence about cause and effect. Independent information on natural laws and logical relations governing the system under consideration is needed to establish a causal relationship. Most of the econometric and regression-based models belong to this class. A typical example is a model of dependence between the weight and height of individuals in a population. There is a strong correlation between these variables, but what is cause and effect, what is input and output?

This second attribute is rather important from the point of view of validation methodology: causal models can be subjected to *simulation experiments*, while such experiments are not possible in the case of descriptive models. In other words we can experiment with modeling to answer what will happen with a specific input signal. This kind of experiment cannot be performed for the model mentioned above. It is possible, however, to use a *formally* obtained relationship between height and weight (usually in the form of a linear equation) but such an experiment is not very sensible.

The third attribute, called the *interpretative type* of model, is related to the way in which the modeling results are interpreted. Here we can distinguish between probabilistic and nonprobabilistic (or deterministic) approaches to model interpretation, although there are also other ways of including uncertainty in model interpretation (e.g., the fuzzy approach). It is necessary to stress here that

- the same model can be interpreted in both ways. For example, we can use a linear model estimated on basis of least squares analysis, and interpret the results in terms of a probabilistic analysis, or compare only judgementally the numbers obtained from measurements and from the model. Thus, the interpretative type of the model depends on the methods of analysis, rather than on the form of the model.
- the interpretative type of the model does not depend on the nature of the real world. The assumptions about the deterministic or indeterministic nature of the real world is a purely philosophical hypothesis and has nothing to do with the type of models we use: we can describe a deterministic world using probabilistic models and vice versa.

The interpretative type of models automatically determines the possible tools for model validation. The only difficulty relates to the necessity of specifying assumptions about the model environment. In fact, when using probabilistic models it is also necessary to build models of the environment of the base model, for example, statistical properties of measurement errors. It is then necessary to validate these additional models, which, of course, causes further technical difficulties.

In the case of deterministic models, the situation is even more difficult: there are no formal methods of model validity analysis. The only possibilities here are sensitivity analysis and heuristic methods (visual inspection of the results, judgmental estimation, etc.). Model adequacy can then be tested only in a qualitative way. We are now able to characterize the model in terms of the attributes formulated above, and hopefully can suggest tools for model validation connected with every attribute. Possible situations are presented in Figure 1. Let us briefly consider the existing com-

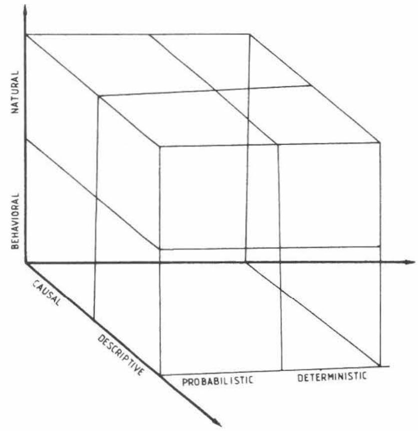

Figure 1: Model attributes and classes

binations (eight possibilities). Some of these combinations seem to be empty, for example, it does not seem possible to build a natural and descriptive model, or to build a descriptive and determinstic one. The suggestions dealing with possible validation tools, however, can be formulated rather automatically, on the basis of previous considerations. These suggestions have been collected in Figure 2. It can be seen, for example, that for a natural, causal, and deterministic model one can use an internal validity approach based on a simulation approach supported by sensitivity analysis and judgmental evaluation. If the last attribute is "probabilistic" we can also use internal validity based on simulation techniques but using probabilistic methods to interpret the results [15].

These statements seem to be rather general and, of course, do not constitute a solution to the problem,

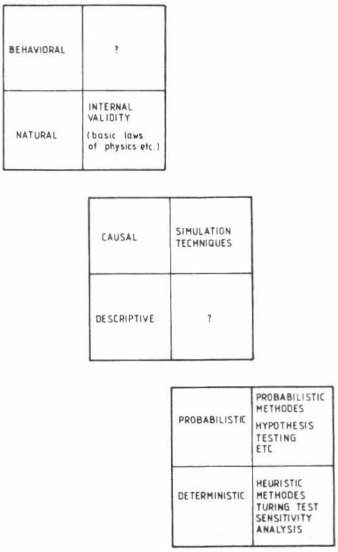

Figure 2: Model classes and existing validation methodologies

but provide instead guidelines for the solution of a concrete problem. Moreover, for some combinations of model attributes there are no existing tools for model validation. Thus, on the basis of these investigations, we can see what kind of methods should be used in future and what classes of validation techniques are interesting from the practical point of view. It is necessary to point out here the model attributes listed above are incomplete. It is, of course, possible to formulate many other attributes but they are not so important from the point of view of model validation; however they do have influence on the validation process, and for this reason we shall call them "secondary attributes". In this way we obtain two model classification levels. It is also necessary to point out that these attributes can be essential at the early model building stage to determine possible technical tools for the modeling. These secondary attributes consist of the following:

- linear – nonlinear
- time constant – time dependent
- continuous time – discrete time
- dynamic – static.

4 System attributes

The model is only the first component in the validation process. The second component is the system or the real world. Clearly, system attributes and their relationship to model attributes will influence the validation methodology.

The first attribute we shall consider is the *experimental type of the system*. This attribute determines which kind of experiments can be performed with the system. Three possible situations may occur:

(1) The system is a design abstraction, not yet existing in the real world and there is no experimental basis for modeling. This kind of situation arises very frequently in engineering problems when determining new systems: modeling is then used to test complicated projects. As the real system does not exist, there is no "reality" which can correspond to the model. In every realistic situation however, there is a correspondence with reality; practically every new system under construction consists of components already applied in other existing systems. This means that the model consists of *submodels* which have previously been tested. A good example is chemical engineering modeling where new technology connects a series of apparatus (reactors, distillation columns, mixers, etc.). Models of such apparatus are well known and in this case we are able to *extrapolate* our knowledge. Models consisting of well-validated submodels will probably be valid, and this kind of approach can be called *component validation.*

(2) The system exists in the real world, but it is not possible to make active experiments. This is the situation which arises most frequently. It occurs in economic and social system modeling, and environmental and technological problems. The "reality" in this case is a data record which in most instances is too short and of too low a quality. This situation makes things rather difficult from the point of view of model validation. Because of a small data base, typical statistical methods frequently cannot be applied. A possible solution is to apply the extended model concept developed by Wierzbicki [31]. The extended model is built starting with the basic model in question and supplementing it by models of possible differences between the basic model and reality from *a priori* knowledge of system properties and partially validated by existing measurements. The extended model is then treated as the "real world" for evaluation and verification of the simplified model. This concept has been applied with success in the modeling of technological processes (in chemical engineering, gas and water transmission systems). The author also believes it is possible to apply this concept to environmental systems modeling (e.g., water quality problems) or even economic systems.

(3) The system exists in the real world and it is possible to make a series of active experiments. This is the best situation, of course, but it occurs very rarely. In this case we have good support for model validation; it is

possible to generate as much data as necessary, to apply experiment design techniques, and so on. Statistical methods can be applied as well as those described in the literature (for example of Turing test and extensions, see [27]; for hypothesis testing, see [9]). Possible situations in the model validation process are shown in Figure 3.

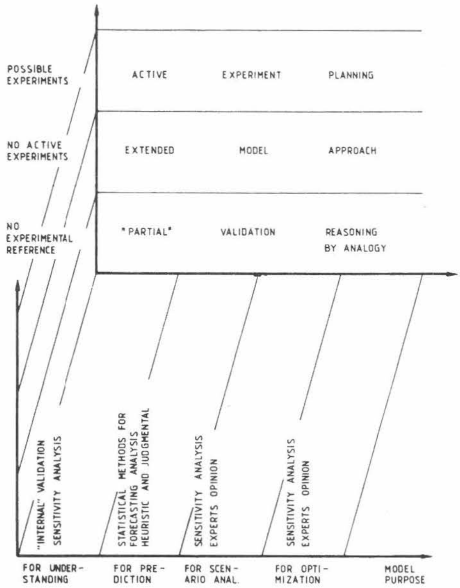

Figure 3: Possible situations in the model validation process

5 Validation attributes

Let us now consider the validation process. It is obvious that this process depends both on the model and system attributes and that it is necessary to combine them; some combinations, however, limit the number of possible validation approaches. It is not possible, for example, to use statistical methods for analyzing the validity of a deterministic model. Model type, however, is only one of the important attributes of the validation procedure. Two other important aspects are the model purpose and the relationship between the model and the real world.

Many authors point out that the model validation process should be goal-oriented, however, it is not an easy task explaining what this statement means. Let us consider possible situations:

Modeling for understanding

In many instances, the only modeling goal is to understand the system structure and its behavior better. The modeler can perform simulation experiments, he can "play" with the model in order to observe what will happen in certain situations. One of the most important advantages of such experiments is the fact that it is then possible to view the internal structure of the model and see the processes "inside" the investigated phenomena. This kind of investigation is especially popular in physics and astrophysical research, and has also been utulized in ecological research [19].

The main problem that arises with validation is the relationship between the *structure of the process and the structure of the model*. According to the terminology introduced above, the internal validity (or model testing "part-by-part") should be performed in this case. One other factor can also be important: that the model should pose a level of "internal stability" with respect to data. Sensitivity analysis is then recommended for checking this property. "Sensitivity" should be understood here in a rather broad sense. During the modeling process we make a number of assumptions dealing with the external world (system neighborhood), model structure and model parameters, and one of the goals should be the exploration of the influence that these assumptions have on model behavior. It is necessary to mention here that a single simulation run without more exact analysis is of little practical value from the point of view of understanding the system. The importance of sensitivity testing has been described well by Quade [21]:

> *Ordinarily there is no unique, "best" set of assumptions in modelling, but a variety of possibilities, each of which has some basis for support. A good system study will include sensitivity tests on the assumptions in order to find out which ones really affect the outcome and to what extent. This enables the analyst to determine where further investigation of assumptions is needed and to call attention to the decisionmaker to possible danger that might be present . . .*

Similar ideas are also considered in Quade and Findeisen [8]. There are many formal tools for sensitivity analysis and basic concepts have been considered by Tomovic [30] and Wierzbicki [31]. Especially interesting is the general framework for sensitivity analysis developed by Wierzbicki and his concept of basic and extended models. There are also a number of good examples of model sensitivity analysis, especially in ecosystem modeling (see, for example, Rosk and Harmsen [23]). A lot of research in this direction has been performed at IIASA: sensitivity analysis for energy models (Konno and Srinivasan [12]; Suzuki and Schrattenholzer [26]), for demographic models (Arthur [1]; Willekens [32]) as well as some more general investigations (Stehfest [25]). There are, of course, many other excellent works available in the literature (see, for example, Thornton et al.

[29]) but because of lack of space these will not be considered in detail here.
It is necessary to point out here, however, that the existing methods of sensitivity analysis are only local and parametric. This means that it is rather difficult to investigate large deviations of parameters and structural changes in the model. All methods are also only applicable to models continuously depending on parameters – there is no way to analyze sensitivity in a discontinous case. In the non-differentiable case for large parameter variations, estimation of Lipshitz constant might be a help; however, there are only a few theoretical papers on model sensitivity that deal with this question and the theoretical basis is as yet not fully advanced.

Modeling for forecasting

This is one of the most frequent situations, and probably the most difficult one from the point of view of validation approach. The particular situation has been considered by Beck [4] and Mankin et al. [19]. The main difficulty arises from the fact that a well-validated model, in the sense that the model responses correspond very well to the system outputs, does not necessarily reflect the future behavior of the system well. The reasons seem to be rather *obvious* in that there can be an essential non-stationarity in the system environment, or that there are some additional input variables which are not considered in the model. In both cases the model is evidently inadequate although it may happen that factors not considered in the model manifest their presence only during the forecasting (model utilization) period. Makin et al. [19] have therefore introduced a concept of *model usefulness and model reliability*. According to their terminology, a *valid* model has no behavior which does not correspond to system behavior, and a *useful* model predicts some system behavior correctly. It is obvious, however, that although generally a valid model is useful this may not always be the case. There still remains the problem, however, of how to determine the usefulness of the model, and, of course, it is not possible to do it *a priori*. In the case of statistical model interpretation, validation of forecasting models is understood better, and we can use these tools to determine the model usefulness. Moreover, by applying the Bayes approach it is possible to determine the confidence intervals for predicted system behavior. Pioneering work has been performed by Box and Jenkins [3] and their methodology is a good example of general modeling methodology. As a final test for the usefulness of the model they consider the statistical properties of the prediction error. Another criterion for model validation has been considered by Kashyap and Rao [13] and in every case they assume that the qualitiy of prediction is the main criterion for model quality analysis. In this case, however, it is necessary to assume that prediction will be performed many times, and only in this case we can apply probabilistic methods to analyze the quality of the prediction – and consequently the quality of the model.
A different situation arises frequently in the case of economic forecasting where we have a very short data series and a prediction is only made once. This is complicated and only a few rather heuristic methods have been developed. Introductory work on this subject has been made by Waszkiewicz [33] where some new validation criteria for forecasting methods have been formulated and analyzed.

Model for scenario analysis

Scenario analysis models simulate the future behavior of a system on the basis of a judgmentally chosen set of assumptions, called scenarios and the time horizon here may be rather long, say, 100 years. The World Global Models and the IIASA Energy Models are good examples of this type of model, and in this case there is no accepted methodology for model validation.
An additional difficulty connected with scenarios is the fact that they are also models, models of the neighborhood of the system being modeled, and these models should also be validated. As yet, there are only a few works dealing with this problem, and much more research in this direction is needed. A critical analysis of the existing modeling approaches for scenario-analysis has recently been made by Kalman [11]. He analyzes the world models of Forrester and Meadows from a systems theorist point of view. In his opinion

> *. . . the model consists of a system of nonlinear difference equations which are analyzed by simulation. It is a well-known fact that in such a system almost anything can happen . . . Unless there is an "organizing principle" for writing down these equations and thereby a priori controlling their properties, rather complicated and erratic behavior may be expected on general theoretical grounds. Such an organizing principle is not available from theoretical economics and the naive faith that the equations (might) "represent" reality is certainly not good enough . . .*

Kalman also stresses the role of sensitivity analysis as a validation tool in scenario model analysis:

> *. . . (they observed) that small variations in the assumed parameters and initial conditions result in gross changes in observed behavior. Since these*

parameter variations of the order of 2–10 percent are much smaller than the reasonable uncertainties in their values on economic growth (of the order of 30–100 percent), the value of the Meadow exercise is utterly destroyed. Any general conclusion from the model must be rejected because the behavior of the model is just not robust enough under parameter uncertainty ...

A critique of the existing methodology of scenario analysis has also been performed by Scolnik [22], and Dubovsky and Pirogov [6]. Practically, sensitivity analysis is the only method for validating these models. In a case mentioned by Kalman, this analysis has shown *nonadequacy* of the model. However, there are a number of other works available where sensitivity analysis applied to scenario models does not give such a pessimistic conclusion (e. g. [12, 26, 24]).

Despite these efforts and the understanding partially given by them, we must conclude that the methodology for validation of scenario models does not, as yet, exist.

Optimization models

There are three basic types of models where optimization methods can be applied, and in every case the role of optimization is quite different; thus, different methods for model validation should be applied.

The first situation occurs when the phenomena being modeled can be described in terms of *variational principle* – where minimization (or maximization) of something is a basic principle of system behavior. A typical example is the minimization of energy in mechanical or electrical systems; every system operates in such a way as to minimize the total energy accumulation. In this situation instead of writing down all the equations and then solving them, we can formulate the function by describing the total energy which depends on the system variables. Then, minimization of this functional solves the problem and we obtain the variables at the point of equilibrium. This approach has been investigated by many authors (e. g. Kurman [16]). The role of optimization is evident: it is only a tool for solving the model, while the model itself belongs to one of the previously mentioned classes.

The second situation occurs when we want to make some experiments whith the model to determine the possible model responses. In many situations, optimization methods are good tools to perform this task. Usually we can formulate an objective function (sometimes also called the performance index). While using appropriate parameterization and optimization procedures it is possible to investigate system responses. It is necessary to point out, however, that very often a single objective function has no economic or other practical meaning and should be considered more as a technical tool for diminishing the number of investigated parameters. Clearly, it is more convenient to operate with low numbers of objective function parameters than with a large number of model solutions or trajectories. In this situation, a more straightforward approach is to specify many objective functions with good economic, or other practical, interpretations and apply one of the existing multiple-objective optimization methods. A reference point optimization method developed by Wierzbicki [31] is a very useful tool for analyzing possible solutions to optimization models with many objective functions. This approach has recently been applied to several IIASA models, see, for example, the investigation of the Finnish forest and wood industry sectors [14]. In this case, the use of the optimization approach also does not reflect directly on validation methodology because optimization is only used here as a tool for model analysis.

The third situation is essentially different from the previous ones in that a model is used to determine an optimal system operation and the resulting decisions are then applied to the real system. These kind of models are called *decision and control* models. It is necessary to stress from the beginning one important fact which very often is only implicitly understood: in the case of decision and control models, we deal, in fact, with two models

- the model of the system being optimized and
- the objective function model.

This distinction is important as it is related to the following observations:

- Solutions obtained in a decision and control model are often very sensitive to the form of the objective functions; practically, the objective function determines the solution of the problem.
- The objective function model is only an approximation of the real costs in many cases (especially in social sciences and ecology) and it is not possible to express all the aspects of the system operation in the same (monetary) units.

It is also necessary, therefore, to validate the objective function model. Essential methodological difficulties arise when considering the relationship between a decision and control model and a real system. Practically, the first goal of decision and control modeling is to improve the *system operation*, that is, to optimize the value of the real objective function, measured on the real system. This causes several problems, one of them being that it is not always possible to measure real values of objective functions. A second, and important, problem is that the properties of the pair model/real

system depend on the structural properties of the connection between the model and the system, and on the method of applying computed decisions to the real system.
One of the possible ways of validating decision and control models is to utilize the knowledge of an experienced system operator (a manager, a dispatcher, or a similar expert familiar with system behavior). In practice, this knowledge is quite substantial and such experts usually have no difficulty in evaluating computed solutions. There are also more formal approaches of taking expert opinion into account, i.e., multiobjective methods, developed, for example, by Raiffa and Kenney [17] and the methods proposed by Eremin and Mazurov [7] among others. A valid decision model can be defined in this case as a model whose solutions do not contradict with the expert's opinions. An extensive analysis of the relationship between the "model/ real system" pair can be found in [31] but so far the results have only been applied to control engineering problems. However this methodology is universal and could also be applied in other fields. The fundamental concept in this methodology is the distinction between basic and extended models, mentioned earlier, and supplemented with a rather extensive sensitivity analysis.

6 Validation process

Validation is not a single act, it is a process. It follows from the fact that model building is an iterative procedure. It is possible, however, to separate this process into stages, connected strictly with the stages of model building.
In the first stage of model building it is necessary to determine the model type, what its basic attributes are and what its relation to the system being modeled is. This stages of modeling and consequently the detail analysis of the assumptions made (which can be called *initial verification* or *hypothesis verification*) is especially important as any mistakes are costly and time consuming. For example, at this stages important aspects such as the possible application of the discrete time model to the continuous time system, static models for a dynamic system, etc., are discussed. In any case, however, the initial assumption should be very carefully analyzed taking into account the purpose and possible future applications of the model being developed.
In the second stage of model building, when the model is being formulated and computerized, it is necessary to validate the "model itself", that is, without taking into account the modeling purpose. One of the questions at this stage is the relationship between the computerized model and the conceptual model obtained in the first stage. In other words, the correspondence between the model, the initial knowledge of the modeled phenomena and the expected model behavior should be checked. According to Hermann terminology this stage of model verification can be called *face validity:* " . . . face validity is a surface or initial impression of a simulation or game's realism" [10]. From the methodological point of view, however, this is not really validation: this stage should rather be called a test of reasonable credibility of the model. In many cases, information can be obtained from experts (or managers) that could judge whether the model is reasonable. In other cases more formal methods can also be used.
The third stage of validation depends strictly on the purpose of modeling, and for this reason this stage can be called *essential validation*. Possible questions arising from this stage have already been considered in a previous section and will not be repeated here. It is useful, however, to stress the difference between "face validity" and "essential validity". Consider for example, a model for predicting future system outputs. Face validation is concerned with the correspondence of model outputs to past historical data, where essential validation is concerned with the quality of prediction. It is obvious that we cannot expect good predictions from the model which has been rejected at the face validation stage; however, a positive face validation cannot guarantee good quality predictions. Face validation can be interpreted as a sieve for the selection of models before further, more complicated stages of validation are performed.

7 Conclusions

In this work, a framework for model validation has been proposed. The main conclusion is that the problem of model validation can be more strictly defined by analyzing in more detail the model itself and the purpose of modeling. On the basis of this analysis it is possible, in many specific cases, to propose appropriate tools for model validation. The problem still remains, however, of putting these tools to the best use. Moreover, in many important cases such tools do not exist, or are insufficiently developed. In the author's opinion, a more detailed analysis of possible situations, appropriate tools, and their use is an interesting and important direction to take in model validation research.

References

[1] *Arthur, W. B.:* The Analysis of Causal Linkages in Demographic Theory. WP-80-110, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1980.

[2] *Aitchinson, J., Dunsmore, I. R.:* Statistical Predicition Analysis. Cambridge University Press, Cambridge 1975.

[3] *Box, G. E. P., Jenkins, G. M.:* Time Series Analysis, Forecasting and Control. Holden-Day Publishing, San Francisco 1970.

[4] *Beck, M. B.:* Hard or Soft Environmental Systems? WP-80-25, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1980.

[5] *Burns, J. E.:* Error Analysis of Nonlinear Simulations – Application to World Dynamics. IEEE Transactions on SME, SMC-5 (3) 1975.

[6] *Dubovsky, S. V., Pirogov, G. G.:* Critical Survey of Some Global Modeling Approaches. In: Global and Large Scale System Models. Proceedings of the Center for Advanced Studies, International Summer Seminar, Dubrovnik, Yugoslavia, August 1978. Springer Verlag 1979.

[7] *Eremin, I. I., Mazurov, W. D.:* Nonstationary Processes in Mathematical Programming. Nauka, Moscow 1979 (in Russian).

[8] *Findeisen, W., Quade, E. S.:* The Method of Applied Systems Analysis – Finding a Solution. WP-80-62, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1980.

[9] *Greig, I. A. D.:* Validation, Statistical Testing and the Decisions to Model. Simulation, August (1979).

[10] *Hermann, C. F.:* Validation Problems in Games and Simulations with Special Reference to Models in International Politics. Behavioral Science 12 (1967).

[11] *Kalman, R. E.:* A System Theoretic Critique of Dynamic Economic Models. In: Global and Large Scale System Models. Proceedings of the Center for Advanced Studies, International Summer Seminar, Dubrovnik, Yugoslavia, August 1978. Springer Verlag 1979.

[12] *Konno, H., Srinivasan, T. N.:* The Haefele-Manne Model on Reactor Strategies – Some Sensitivity Analysis. RM-74-19, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1974.

[13] *Kashyap, R. L., Rao, A. R.:* Dynamic Stochastic Models from Empirical Data. Academic Press, New York 1976.

[14] *Kallio, M., Lewandowski, A., Orchard-Hays, W.:* An Implementation of the Reference Point Approach for Multiobjective Optimization. WP-80-35, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1980.

[15] *Klejinen, M. P. C.:* Statistical Techniques in Simulation. M. Dekker Inc., New York 1974.

[16] *Kurman, K.:* Control Theory. WNT, Warsaw 1975 (in Polish).

[17] *Keeney, R. L., Raiffa, H.:* Decisions with Multiple Objectives – Preference and Value Tradeoffs. Wiley, New York 1976.

[18] *Naylor, T. H.:* Simulation and Validation. Proceedings of the Sixth IFORS International Conference on Operations Research, Dublin, August 1972.

[19] *Mankin, J. B., O'Neill, R. V., Shugart, H. H., Rust, B. W.:* The Importance of Validation in Ecosystem Analysis. Proceedings of Simulation Councils. New Directions in the Analysis of Ecological Systems 5 (1975).

[20] *Mirham, G. A.:* Simulation Modeling. International Journal of General Systems 1 (1974).

[21] *Quade, E. S., Bonder, W. I.:* Systems Analysis and Policy Planning, Elsevier 1968.

[22] *Scolnik, H. D.:* A Critical Review of Some Global Models. Proceedings of the Sixth IIASA Global Modeling Conference, Laxenburg (Austria) 1978.

[23] *Rose, M. E., Harmsen, R.:* Using Sensitivity Analysis to Simplify Ecosystem Models. Simulation, July (1978).

[24] *Schroeder, W. F., Svenny, R. E., Alfred, L. E.:* Readings in Urban Dynamics. Wright-Allen Press, Cambridge 1970.

[25] *Stehfest, H.:* Decision Theoretical Remarks on Sensitivity Analysis. RR-75-3, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1975.

[26] *Suzuki, A., Schrattenholzer, L.:* Sensitivity Analysis on Hydrogen Utilization Factor of the Haefele-Manne Model. RM-74-30, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1974.

[27] *Schruben, L. W.:* Establishing the Credibility of Simulations. Simulation, March (1980).

[28] *SCS Technical Committee:* Terminology for Model Credibility – Reports of the SCS Technical Committees. Simulation, March (1979).

[29] *Thornton, K. W., Lessen, A. S., Ford, D. E., Stirgens, C. A.:* Improving Simulation Through Sensitivity Analysis. Simulation, May (1979).

[30] *Tomovic, R., Vukobratovic, M.:* General Sensitivity Theory. Elsevier, New York 1970.

[31] *Wierzbicki, A. P.:* Models and Sensitivity in Control Systems. WNT, Warsaw 1970 (in Polish).

[32] *Willekens, F.:* Sensitivity Analysis. RM-76-49, International Institute for Applied Systems Analysis, Laxenburg (Austria) 1976.

[33] *Waszkiewicz, L.:* Analysis of Forecasting Methods. PWE, Warsaw 1976 (in Polish).

[34] *Zadeh, L. A., Desoer, C. A.:* Linear System Theory. McGraw-Hill 1963.

Dr. Eng. Andrzej Lewandowski
Technical University of Warsaw
Institute of Automatic Control
Nowowiejska 15/19
00–665 Warsaw/Poland
formerly:
International Institute
for Applied Systems Analysis
A 2361 Laxenburg/Austria