

# EVALUATION OF SIMULATION MODELS IN AGRICULTURE AND BIOLOGY: CONCLUSIONS OF A WORKSHOP

F. W. T. PENNING DE VRIES

*Department of Theoretical Production Ecology, Agricultural University, Wageningen,  
The Netherlands*

## SUMMARY

*A workshop on evaluation of simulation models was held in Wageningen. The object of the meeting was to discuss problems and points of view concerning evaluation of complex models. Most of the workshop was devoted to discussions. At the end, there was a considerable consensus on many points. This paper summarises the major conclusions.*

## THE WORKSHOP

A workshop on 'evaluation of simulation models in agriculture and biology' was held in Wageningen from 14 to 19 December, 1975. The object of the meeting was to discuss problems and points of view concerning evaluation of models that are used to simulate dynamic systems. There was a need for such an exchange of opinions since much experience has developed almost independently by a few groups and individuals over the last decade. Congresses and symposia give little opportunity to discuss this field adequately, and exchange of papers is not enough. Most of the workshop was therefore devoted to discussions, no formal contributions being prepared or presented. About twenty-five modellers (see Appendix) experienced in simulating agricultural and biological systems, participated in the workshop. Their background ranged from applied mathematics and physics to crop science, agricultural engineering and ecology.

In the first session, a survey was made of possible themes, and these were grouped under five headings:

- (1) requirements for distribution and acceptance of models among potential users;
- (2) model objectives and problems of complexity;
- (3) use of sensitivity analysis;
- (4) parameterisation and formulation of basic processes;
- (5) comparison of model and real world in terms of results and behaviour.

Each session was introduced by one volunteer participant and chaired by another; a reporter took notes. The last day was used to evaluate the notes and to elaborate some remaining problems.

At the end of the week there was considerable, but no full consensus on many points and it was considered useful to publish a brief report on the meeting for other workers in the field of dynamic simulation. This paper summarises the major conclusions of the workshop as seen by the author. A draft of this paper was read and commented upon by most participants. However, the opinions expressed below may not be shared by all of them.

#### TERMINOLOGY

Discussing terminology, it was felt that the word 'evaluation' is the broadest term to describe the action of judging the value of a model. 'Evaluation' may be used for checking internal consistency and units used in a computer program, for comparison of model output with real world data and for judgement of practical utility. 'Validation' is used as evaluation with emphasis on usefulness and relevance of the model, whilst 'verification' is used for evaluation with emphasis on truthfulness. These terms are sometimes applied to different stages of modelling activities. 'Falsification' can be used to describe the statistical judgement of agreement of model output with real world data. Other terms were suggested also for specific forms of evaluation. The workshop preferred not to attempt to standardise the definitions of these words, but suggested that it is important to describe such terms clearly whenever they are used in communications.

#### THE SCIENTIFIC, PREDICTIVE AND INSTRUCTIVE VALUE OF A MODEL

Many models are scientifically interesting, others are utilised to predict and their results are applied in other fields, and some models are largely used for instructive purposes. But all models have some value in each of these fields.

A scientifically interesting model contributes to our understanding of the real world because it helps to integrate the relevant processes of the system studied and to bridge areas and levels of knowledge. It helps also to test hypotheses, to generate alternative ones and to suggest experiments to falsify them. Briefly, the scientific value of a model is its contribution to the development of science. A good predictive model simulates accurately the behaviour of a part of the real world in situations where its behaviour has been, or has not yet been, observed. It is therefore a good instrument to apply scientific knowledge in practice. It should predict reasonably well over a range of boundary conditions, to provide its users with alternative solutions of a problem. Good scientific models are often too detailed or too

speculative for those who want to apply them, whilst models used for predictive or management purposes are often too trivial or too crude to challenge scientific interest.

Applicability is not a useful term to characterise models. All models are applied, but for different purposes: to develop science, to predict correct results or to instruct people. The only models that are not applied are those that are filed away.

#### OBJECTIVES OF MODELLING

The workshop devoted much time to the objectives of modelling. Setting the objectives of a model generally takes much time, particularly in large-scale, multidisciplinary research efforts. Formulating the goal of the work, identifying the system and its boundaries, setting priorities of development and determining the accuracy desired in preparation for the analytical scientific work determine the type of model and its uses, and thus the type of evaluation procedure that should be used. Setting of objectives was estimated to occupy 5% to 50% of the total time spent in large-scale predictive modelling efforts. Most of this time is spent early on in the effort, but some of it later, for refining and redefining the original goals. When setting model objectives, it may become apparent that simulation is not, or not yet, the most appropriate method for studying a particular system.

Objectives of modelling were distinguished as (a) those for which the model is or will be designed, (b) those of the modelling effort and (c) personal objectives of the individuals involved. In models developed for scientific purposes objectives are usually not specified beyond those of acquiring a better understanding of the system studied. Objectives of models used for managerial purposes are sometimes better defined; there may be a requirement, for example, to simulate the behaviour of some critical variables in certain situations with a 5% to 20% accuracy for most of the simulated time span. Some objectives may turn out to be too difficult to achieve and are either not met or adjusted. The effort, however, may still have been valuable because of the knowledge gained. It seems that the value of the modelling effort, *per se*, must be emphasised when it is difficult to oversee the scientific field in which the exercise is to occur. The scientific value of a model seems therefore particularly related to the modelling effort, whilst its practical utility in management depends particularly on the quality of the model finally obtained.

Like anyone else, those involved in modelling have personal objectives. An important and legitimate personal objective is the formal appreciation of one's work, independent of that of colleagues. Considerable friction can develop if specialists are asked to answer specific questions which are then collected and published by a modeller. It is in general not only necessary that the importance of the problem attacked by a modelling team is recognised by its members, but also that the research of each scientist in the team is relevant for his scientific discipline as such, so

that it can lead to valuable, independent publications. The pure modeller, often a computer expert or a mathematician, must know how to communicate with the team specialists in their scientific language, and he has the burden of explaining to them the formulations of the model.

#### TRANSPARENCY AND COMPLEXITY

In electrical engineering, a 'black box' implies a system with a well known transfer function. In biology and agriculture, this term is sometimes used to label an unknown mechanism that relates outputs of a system to its inputs. Such 'black boxes' form the frontier of knowledge in empirical sciences and challenge the specialist. If these black boxes are related to one another in a way that is widely accepted by scientists, that system may be called 'transparent' by analogy. All variables in a model of a transparent system have a biological, physical or chemical meaning. The more transparent a system is, the more suitable is its model for making predictions. 'Black boxes' must always be described and incorporated in models in such a way that the subsystem or process represented is accessible for experimentation. Computer programs of simulation models are, in principle, transparent, since their basic operations are known; their formulation may be clumsy, however, and thus difficult to comprehend or evaluate.

Adjusting parameter values of a model of a black box with few inputs such that the expected outputs are obtained is a form of curve fitting. The resulting equation gives no clue to mechanism or generality. If something is known about the contents of the black box, which may then be called a 'grey box', this information should be used for adjusting and selecting parameter values. This process of parameterisation, which may be called 'tuning', gives little or no insight into the mechanism or the generality of the solution, but is often unavoidable in early stages of model development. Sensitivity analysis of modelled subsystems may help to indicate the accuracy of parameter values required. 'Tuning' of models of grey boxes with many inputs and outputs is at least cumbersome. The maximum number of parameters to be 'tuned' simultaneously depends on the knowledge of that subsystem and the amount and quality of real-world data available, but is often considerably smaller than the maximum of about ten, which is now technically possible.

It may be advantageous not to implement all known details in a model of a transparent system. This is obviously the case in most instructive models, but it can be a useful procedure in scientific and predictive models also, when their principal results are of a very much lower level of detail than the basic information in the model. The use of models on one level of detail to generate data or a simpler model for use on a less detailed level is often called a hierarchical approach to model building.

The size of a model may increase because its objectives are broadened, or because it is elaborated. In the first case, the number of parameters usually increases and the sensitivity of model behaviour to each parameter decreases. Elaborating the model of a system implies the formulation of more structure. The more transparent a system is, the more structure and the fewer parameters its model usually contains. A thorough knowledge of a complex real world system, and thus a large model of it, is always required before the model can be simplified reliably for use by others. The simpler a model that still accomplishes its purpose, the better it is for those who want to apply it in other fields or higher up in the model hierarchy. Hence, the model attains its maximal scientific value while it is being elaborated, while its value for application elsewhere increases during subsequent simplification.

#### COMPARISON OF MODEL OUTPUT WITH REAL-WORLD DATA

The first thorough test of a model is often the comparison of the behaviour of its output with that observed of the real system in an analogous situation. This behaviour includes, for instance, the general shape of the time course of variables, the presence of discontinuities and the qualitative sensitivity of output to parameter values. One should be aware, however, that aspects of model behaviour that seem counter-intuitive at first sometimes turn out to be realistic.

If the behaviour of a model matches qualitatively that of the real world system, a quantitative comparison and an evaluation of the predictive success of the model should be made. At this stage, statistical tools can be useful. But even when sufficient and accurate data are available, a model cannot be proved to be correct. Sometimes model behaviour can be falsified, and thus one or more model components may be shown to be in error; a model as a whole cannot be proved to be incorrect. 'Calibration' of a model, the adjustment of some parameters such that model behaviour matches one set of real-world data, is a very restricted form of evaluation. A useful form of behavioural analysis is an extension of sensitivity analysis. It is done by increasing or decreasing one parameter value over a broad range, and comparing direction and shape of the output with the known or expected direction.

Large system simulation models have been developed by various groups. The evaluation of such models is difficult because many detailed observations are needed before a critical overall test can be made. It was found that if such observations are available before the final tests of a model are performed, some of the information is often, unintentionally, used for 'tuning' of some parameter values. It is almost impossible to avoid this and it should therefore be realised that the basic inputs of the model are then not independent of the real-world system with which the model is compared. When these many detailed observations are not available, evaluation must take place at the level of subsystems.

## PUBLICATION

Most workshop participants experienced difficulties in publishing simulation models: a thorough presentation is too long according to most journal criteria, but a short publication does no justice to the model, the research effort and the modeller, and does not satisfy the interested reader. Papers on well-developed models still need to be long because concepts and methods of simulation are not yet standardised and one can seldom refer to descriptions of principles in the literature. However, a convergence of thoughts about simulation is developing, particularly on the value of the state variable approach as a basic element of the simulation of continuous systems. Some of the technical problems are now well worked out. Standardisation is therefore expected to develop in the next decade.

Large and thorough communications need an expanded introduction to avoid the reader becoming lost in a mass of detail. A 'spiral approach' to publishing may be appropriate in such cases, describing the model several times at increasing depths. Such increasingly involved treatments, with a full listing of the simulation program, may be in one volume and require great skill in writing. It also seems to be a good practice to publish a first and second round of the spiral in different scientific journals and to write a full report as an internal publication or a monograph, for interested colleagues and for oneself, for reference. All publications should then refer to each other and preferably deal with one version of the model. Unfortunately, this is difficult for large models, since their development usually continues while the publications are being written.

It is in the nature of models with a high scientific value to identify areas that are poorly understood. Some of the new questions will usually be answered as a part of the modelling effort, but the solution of others demands more than the modelling team can do within its limits of personnel, time and facilities. Passing on such questions in publications to others is a legitimate and valuable way of suggesting research topics to a broader audience.

Comparing the behaviour of a real-world system in a certain situation with its simulated behaviour in the analogous situation usually requires many data of the time course of state variables, of rates of processes and of boundary conditions. Such a complete description of experiments is not readily accepted by most journals, but may be published in internal reports or bulletins. These should then be referenced through the normal research papers.

Sources of increasing concern are errors in models and in their published descriptions. The most fundamental and difficult errors are conceptual mistakes. Apart from these, even carefully screened simulation programs often contain simple technical errors, such as key-punching errors, incorrectly dimensionalised parameters and deleted variables in expressions, or deleted equations. Some of these may be detected by use of an appropriate program compiler. Others appear when the model is used to simulate quite different experiments, and when someone else

studies the simulation program. FORTRAN cross-reference facilities may be more useful than is often recognised, while the development of a program to check the dimensions of all model variables could be a useful addition to evaluation techniques. No guarantee, however, can be given that a model is free of errors. Through vigorous evaluation, modellers should eliminate as many technical errors as possible before releasing the model. But users should always be on their guard for the presence of errors. It was suggested that it is helpful to keep track of and to date corrections of published simulation programs, although this may demand more administrative work than many modellers are prepared to do.

## APPENDIX: PARTICIPANTS OF THE WORKSHOP

Dr G. Ågren,  
Barrskoglandskapet ekologi,  
College of Agriculture,  
Box 7008,  
S-750 07 Uppsala,  
Sweden.

Dr D. Bennett,  
CSIRO, Division of Land Resources  
Management,  
Private Bag, PO,  
Wembley, WA, 6014,  
Australia.

Dr B. Curry,  
Ohio Agricultural Research and  
Development Center,  
Wooster, Ohio 44691,  
USA.

Dr F. E. Eckardt,  
Institut for Økologisk Botanik,  
Øster Farimagsgade 2 D,  
1353 København K,  
Denmark.

Dr S. Falk,  
Dept of Plant Physiology,  
University of Göteborg,  
Carl Skottsberg Gata 22,  
S-413 19 Göteborg,  
Sweden.

Ir J. Goudriaan,  
Dept of Theoretical Production  
Ecology,  
De Dreijjen 2,  
Wageningen,  
The Netherlands.

Dr Takeshi Horie,  
National Institute of Agric. Sciences,  
Division of Meteorology,  
Nishigahara, Kita-ku, Tokio 114,  
Japan.

Dr D. Baker,  
USDA-ARS,  
Cotton Production Res. Unit,  
PO Box 5367,  
Mississippi State, Miss. 39762,  
USA.

Dr N. R. Brockington,  
The Grassland Res. Institute,  
Hurley, Nr Maidenhead,  
Berks., SL6 5LR,  
Great Britain.

Dr A. Dovrat,  
School of Agriculture,  
PO Box 12,  
Rehovoth,  
Israel.

Mr P. Edelsten,  
The Grassland Research Institute,  
Hurley, Nr Maidenhead,  
Berks., SL6 5LR,  
Great Britain.

Dr M. J. Frissel,  
Institute for Atomic Sciences in  
Agriculture,  
PO Box 48,  
Wageningen,  
The Netherlands.

H. D. J. van Heemst,  
CABO (Center of Agrobiological  
Res.),  
PO Box 14,  
Wageningen,  
The Netherlands.

Dr G. S. Innis,  
Utah State University,  
Ecology Center,  
Logan, Utah 84322,  
USA.

Dr Toyoki Kozai,  
Dept of Horticultural Engineering,  
College of Horticulture,  
Chiba University,  
Matsudo, Chiba 271,  
Japan.

Prof. R. S. Loomis,  
University of California,  
Dept. of Agronomy and Range Science,  
Davis, California 95616,  
USA.

Dr H. Moldau,  
Astrophysical Observatory of Tartu,  
Toravere,  
Estonia, 202444,  
USSR.

Dr R. Peart,  
Department of Agricultural  
Engineering,  
Purdue University,  
Lafayette, Indiana 47907,  
USA.

Ir R. Rabbinge,  
Dept of Theoretical Production  
Ecology,  
De Dreijen 2,  
Wageningen,  
The Netherlands.

Dr N. G. Seligman,  
Agricultural Research Organization,  
The Volcani Center,  
PO Box 6,  
Bet Dagan,  
Israel.

Prof. C. T. de Wit,  
Dept of Theoretical Production  
Ecology,  
Wageningen,  
The Netherlands.

Dr J. R. Lambert,  
Clemson University,  
Dept of Agricultural Engineering,  
Clemson, S. Carolina 95616,  
USA.

Dr Ph. Miller,  
San Diego State University,  
San Diego, California 92115,  
USA.

Dr I. Noy Meir,  
Hebrew University,  
Department of Botany,  
Jerusalem,  
Israel.

Dr F. W. T. Penning de Vries,  
Dept of Theoretical Production  
Ecology,  
De Dreijen 2,  
Wageningen,  
The Netherlands.

Ir F. Rijsdijk,  
Department of Fytopathology,  
Binnenhaven 9,  
Wageningen,  
The Netherlands.

Dr J. H. M. Thornley,  
Glasshouse Crops Res. Institute,  
Worthing Road,  
Rustington,  
Littlehampton,  
Great Britain.

Dr J. C. Zadoks,  
Department of Fytopathology,  
Binnenhaven 9,  
Wageningen,  
The Netherlands.

