Evaluation of models

H. van Keulen

Introduction

In the increasing stream of publications on the use of computer modelling and simulation as tools for ecological research, relatively little attention is paid to the evaluation of the models presented.
Models originating from the technical sciences are in general based on detailed knowledge of the theory of the underlying processes, whose mathematical description is exact. Hence such models hardly require experimental verification to prove their validity.
In the biological sciences and certainly in ecology, we are, however, dealing with dynamic systems that are not man-made and in many areas our understanding of the basic principles is fragmentary if present at all. Models of biological systems are therefore often not more than a subjective expression of our opinion about its structure and behaviour. Complex models, when properly formulated, do represent a consistent argument based on these opinions; but that is still no guarantee of their validity.
It should be recognized, of course, that the validity of a model is primarily determined by its purpose. *The* model of a system does not exist, as there may be several models of one system, all perfectly valid, but aiming at different goals. A model of an aeroplane, developed for the purpose of flight control gives satisfactory results without taking into account detailed aerodynamics. However, when a model is built to decide on the design of the machine, aerodynamic laws cannot be neglected.
In general a model, like any theory, aims at summarizing and predicting. Thorough proof must be given that existing historical data can be satisfactorily explained by the model before sufficient confidence can be placed in the predictive results. Verification of the model is therefore an important part of the simulation. Results of carefully designed experiments with computer models based on the purpose of the model, should be tested at all stages with the results obtained from experiments with the real system.

Evaluation at different levels

a. *Postulates*

Whenever the modeling approach is used to investigate a problem, a number of decisions must be taken. The first and most basic problem is the choice of the postulates on which the model is to be based. This involves the decision on the boundaries of the system to be studied, determining which processes are included in the model and which are introduced as forcing functions. The choice is not always obvious. For example, if one is interested in the dry matter production of a maize crop, the macro-weather may be considered as an external variable which is not affected by the standing vegetation. When, however, the influence of a pollutant from a nearby chemical plant on the yield is the main interest, aerodynamic differences caused by the presence of vegetative surfaces may be of decisive influence on the effect of the macro-weather on the fate of the pollutant. The main criterion must be the purpose of the simulation, which should not be too ambitious to keep the model verifiable. The model should be designed in such a way that it yields the kind, the quantity and the quality of data necessary to draw conclusions relevant to its purpose. Hence the system should be chosen in such a way, that the inputs and outputs at the boundaries can be measured. Also for the decision on the distinction between different subsystems each subsystem must be defined such, that it may be subject to isolated experimentation, with measurable inputs and outputs. In whatever way the postulates are chosen or implicitly included in the model by intuitive incorporation or omission of certain processes or interactions, at some point in the evaluation phase we must return to them and check how adequate they are and in which way they influence the results. Although this may seem obvious, the spectacular impact of the results of Meadows' (1972) world model showed that the implied postulates were not explicitly recognized.

b. *Processes incorporated in the model*

Once the decision about the postulates and the boundaries is taken, a set of mathematical equations, each one describing a relevant physical, physiological or ecological process or part of a process, is combined to form the model. All mathematical relations must be subject to evaluation. In general the processes are studied under controlled conditions to establish the relation between external or internal state variables and the dependent rates. Often the technique to obtain maximum information from such experiments is the application of stepwise changes in state variables and recording the dynamics of the response. Such experiments serve as validation tests for independent submodels. Such submodels may then be used in full in the final model when the dynamics of the processes are of interest or the results of the submodel are entered through analytical expressions or tabulated functions. This implies a hierarchical approach to modelling, which can help to make complexity manageable. An alternative to the use of dynamic submodels to obtain quantitative mathematical relations, is the determination of a number of equilibrium situations, which may be described by an analytical expression, like the photosynthesis-light response curve of individual leaves in a crop growth model. Such relations can only be applied, however, if instantaneous adaptation to changing conditions may be assumed, i.e. when the effects of time-lags can be neglected.

Compound relationships are also obtained by the use of a 'black box' approach. This is done when the underlying processes are not known, and inputs and outputs of a specific component of the model are then connected by a special programming technique, which 'mimics' their measured relation. In this way no information is obtained on the causal relationship between the variables and it is dangerous to use them for predictive purposes, because under different circumstances different reactions may occur. It should, however, be realized that every relation on a level higher than that of atoms and molecules is a 'black box' to some extent, but the use of this technique becomes increasingly dangerous when applied on higher levels. Ultimately it turns the model from a explanatory model into a descriptive one which cannot be used for extrapolation at all.

An example of a reasonable use of this way of working is given by Janssen (1974) in his model of germination of winter annuals, where changes on the biochemical level are 'mimicked'.

All three methods described can be evaluated either by statistical methods or by judging the accuracy of the relations from independent knowledge of the measuring methods. Often, however, no quantitative data are available at all and relations are introduced based on 'intelligent guesses'. This may not be disturbing when it concerns minor details of a model but when important relations are based upon this

principle, model validation becomes a recreative pastime and the investigator should consider going back to experiments with the real system in order to establish the relevant relations. At best, results obtained from such models may then serve as a guideline in designing proper experiments.
An additional problem arises from the parametrization of the functional relationships. In plant production models often the quantitative reactions of plants grown under different conditions show large differences, though the processes are the same. In general, studies at the process level yield most information in the evaluation phase of modelling, especially when simulation is aimed at gaining more insight into the relevance of various factors.

c. *Evaluation of output and model behaviour*
Testing of the whole model may still be done at two levels: gross output of the model, like yield in crop growth models, may be tested, or we may test the internal behaviour of the subsystems, comprising the model. Testing the gross output is in general not very enlightening, especially with crop growth models. On the one hand the experimental data available are subject to sampling errors, which are seldom smaller than 10%. This implies that the error in the measured growth rates is of the order of 20%, so that, when statistical analysis is applied, 'reasonable agreement' is easily obtained. On the other hand, such models contain so many feedback relations that internal compensation may lead to levelling out of deviations caused by the introduction of erroneous relations.
When, however, only gross output data are amenable to testing, as is often the case in models used in ecology, proper evaluation should contain two phases: (Wigan, 1972) *calibration* and *validation*. The calibration procedure is best described by the term curve fitting. One set of data is used to adapt, within reasonable limits, weak or unknown parameters or relations, so as to reach the best overall agreement between simulated and observed results. Even the most simple ecological model, however, contains already such a large number of parameters that such a procedure often requires an unrealistic amount of experimental data.
In the final stage of validation still other sets of completely independent data must be used to show that the model yields proper results under different conditions. Many of the ecological models developed at

present do not permit this full procedure because of lack of data. This implies that all or part of the same data are used in both the calibration and validation phase so that all that can be evaluated is the extent to which the model regenerates its own inputs. Such techniques are widely accepted in econometric sciences and are completely based upon successive application of statistical methods to obtain goodness of fit. This may be called *identification*, rather than validation and it is questionable whether in such cases simulation has any advantage over multiple regression techniques. The most that can be concluded from such models is that historical events under a given set of conditions may be described by the generated set of equations but no insight into the dynamics of the processes is gained.
Even when in the model the processes are described completely on the basis of physical or physiological principles and the validation experiments are carried out by the same team working in the modelling part, it is difficult to completely separate the two. Unintentionally observations from the experiments play a role in the decisions about the relations that enter the model. Therefore data that were available during development of the model give a better comparison with the simulated results than independent data that were collected later on (van Keulen, 1975). This also shows that the risk of circular reasoning is very high when partly empirical or semi-empirical relations obtained in validation experiments are used in the model. Hence when results from simulation and real system conflict with each other, no attempt at parameter adaptation should be made but the individual processes should be re-examined and improved at the weakest points. This is done more directly when the internal behaviour of the subsystems is used for validation. Although this may be a huge task in more complex models it is the only way to develop simulation models that are not only convincing in their summarizing behaviour, but have also predictive value and can be used to extrapolate knowledge from known situations to new areas or circumstances. A good example of this technique is the use of enclosure studies in which the processes of photosynthesis, respiration and transpiration are subject to direct validation (van Keulen & Louwerse, 1974). Comparison of measured and simulated dynamic behaviour of these processes under different conditions may lead to redesigning of the model, which in turn is a guide line for the design of new experiments (de Wit, 1970). Such an intimate relation between modelling and experimentation will generally not lead to the

rapid production of a great number of models, but will certainly increase confidence in the results that are obtained.

Internal evaluation

So far we have been considering the validity of the model as a representation of reality. There are, however, in the validation phase of modelling some other pitfalls that should receive proper attention. Before any comparison with the real world makes sense, the modeller must be sure of the internal consistency of his model.

There is firstly the problem of dimension inconsistency but, although this may create difficulties, the occurrence of such errors generally shows up in the early stages of model development. It would, however, be very helpful if the problem-oriented computer languages contained a dimension check routine. A more serious problem is that of the correct computer implementation: errors during formulation of ideas, during programming, and mistakes introduced during the writing and punching procedures. Especially in more complex models, which may consist of over 1000 statements, such errors are easily made and difficult to detect. Especially errors made during the formation of the ideas, may escape detection because the normal safeguard of independent implementation by more than one person is impossible. The best solution is running the model in limit situations, where its behaviour is known. Such a test does not completely prove that these errors are absent and so far there is no technique to avoid them completely.

Sensitivity analysis

A widely accepted technique in the process of model evaluation, applied specifically in situations where accurate input data are missing, is sensitivity analysis. It is most conveniently defined as a test on the relative influence of changes in input data and parameters on the relevant outputs of the model. This technique may be especially helpful when it must be decided which subsystems should receive most attention in the experimental field. Relations with the strongest impact on the final result must be studied thoroughly, while those which hardly influence the outcome may be introduced as intelligent guesses. There is, however, a dangerous aspect in the technique: the structure and the functional relationships of the model are taken for granted, so that

when conceptional errors are present, the importance of certain relations may not be visible at all.
Research efforts may then be directed into the wrong field and important parameters may be completely neglected. It is therefore necessary to evaluate sensitivity analysis in the light of all assumptions that were made during the development of the model. In many cases it is more significant to study the sensitivity of the model's results to different postulates, than to different parameter values.

Conclusions

As is clear from the foregoing, proper validation of simulation models is an extremely difficult and time-consuming procedure. It is, however, an essential procedure, as this phase of the modelling process must prove the validity of the opinions on which the model is based. It will also lead to the design of relevant experiments and thus to increasing understanding of the system in which we are interested. One may, however, put the question how useful even thoroughly validated models of ecosystems are for predictive purposes. When a perfect simulation model is to be used for predictive purposes, it is still necessary to initialize it properly to obtain the desired answers. The determination of the initial state of such a system is, however, likely to disturb it to such an extent that completely different behaviour is the result. Hence if each ecosystem is unique, as is often stated, we will never be able to find experimental data to test the results of our model. This may lead to the conclusion that only systems which show a repetitive behaviour are amenable to simulation. This is generally the case with systems that are controlled by a negative feedback.
And that is hardly an encouraging thought at a time where ecologists claim or are asked for qualified opinions about explosive situations.

References

Janssen, J. M. G., 1974. Simulation of germination of winter annuals in relation to microclimate and microdistribution. Oecologia (Berl) 14: 197-228.

Keulen, H. van & W. Louwerse, 1974. Simulation models for plant production. In: Proceedings of the WMO symposium 'On agrometeorology of the wheat crop', Braunschweig (in press).

Keulen, H. van, 1975. Simulation of water use and herbage growth in arid regions. Pudoc, Wageningen. (Simulation Monographs).
Meadows, D. L., 1972. The limits to growth. Universe Books, New York.
Wigan, M. R., 1972. The fitting, calibration and validation of simulation models. Simulation 18 (5): 188-192.
Wit, C. T. de, 1970. Dynamic concepts in biology. In: Prediction and measurement of photosynthetic productivity. Proceedings of the IBP/PP Technical Meeting, Třeboň. Pudoc, Wageningen.