ELSEVIER

Ecological Modelling 90 (1996) 229–244

ECOLOGICAL MODELLING

Testing ecological models: the meaning of validation

Edward J. Rykiel, Jr. *

Center for Biosystems Modelling, Industrial Engineering Department, Texas A&M University, College Station, TX 77843-3131, USA

Received 7 March 1995; accepted 25 August 1995

Abstract

The ecological literature reveals considerable confusion about the meaning of validation in the context of simulation models. The confusion arises as much from semantic and philosophical considerations as from the selection of validation procedures. Validation is not a procedure for testing scientific theory or for certifying the 'truth' of current scientific understanding, nor is it a required activity of every modelling project. Validation means that a model is acceptable for its intended use because it meets specified performance requirements.

Before validation is undertaken, (1) the purpose of the model, (2) the performance criteria, and (3) the model context must be specified. The validation process can be decomposed into several components: (1) operation, (2) theory, and (3) data. Important concepts needed to understand the model evaluation process are verification, calibration, validation, credibility, and qualification. These terms are defined in a limited technical sense applicable to the evaluation of simulation models, and not as general philosophical concepts. Different tests and standards are applied to the operational, theoretical, and data components. The operational and data components can be validated; the theoretical component cannot.

The most common problem with ecological and environmental models is failure to state what the validation criteria are. Criteria must be explicitly stated because there are no universal standards for selecting what test procedures or criteria to use for validation. A test based on comparison of simulated versus observed data is generally included whenever possible. Because the objective and subjective components of validation are not mutually exclusive, disagreements over the meaning of validation can only be resolved by establishing a convention.

Keywords: Calibration; Credibility; Policy; Qualification; Validation; Verification

1. Introduction

Validation is a thorny issue for both ecological model builders and model users as exemplified by the confusing and often mutually exclusive statements in the literature. For example, model validation is sometimes considered essential (e.g., Gentil and Blake, 1981; Power, 1993), and sometimes considered impossible (e.g., Starfield and Bleloch, 1986; Oreskes et al., 1994). Some authors suggest that models can be validated (Law and Kelton, 1991), while others contend that models can only be invalidated (e.g., Holling, 1978; McCarl, 1984). Validation may be an integral part of the model building process (e.g., Overton, 1977), and also testing to be conducted after the model is built (e.g., Goodall, 1972; Shugart, 1984). Validation may be a technical process of statistical analysis (e.g., Kleijnen, 1987;

* Present address: Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352, USA. Fax: (+1-509) 372-3515; e-mail: ej_rykiel@pnl.gov

SSDI 0304-3800(95)00152-2

Mayer and Butler, 1993) or hypothesis testing (e.g., Jeffers, 1978), and also an exercise in formal logic (e.g., Caswell, 1976). Validation is sometimes even confused with 'truth' (e.g., Reckhow and Chapra, 1983; Swartzman and Kaluzny, 1987). Because of these conflicting ideas, some modellers are prompted to avoid using the terms verification and validation altogether (e.g., Botkin, 1993). Amazingly, despite this confusion, it has now become commonplace for empiricists to demand that models be validated, and for modellers to state that models will be validated by comparison to field data.

Is validation possible, and, if so, are there universal criteria for model validation? To what extent are these conflicts based on scientific versus philosophical differences among ecologists? Can these conflicts be objectively and definitively resolved? These are important questions because modelling, from purely mental models to the construction of physical models, is an essential scientific and engineering activity (Giere, 1991). The purpose of this paper is to explore the underlying causes of these conflicts and to assess the potential for resolving them. Descriptions of specific test procedures for simulation models can be found elsewhere (e.g., Law and Kelton, 1991; Mayer and Butler, 1993; Power, 1993).

The vast majority of simulation models are built to meet practical management, industrial and engineering needs. The majority of ecological models are built for scientific research purposes, but increasingly for forecasting and management purposes. From the ecological research perspective, the validation problem reflects ambiguity about how to certify the operational capability of a model versus how to test its theoretical content. The crux of the matter is deciding (1) if the model is acceptable for its intended use, i.e., whether the model mimics the real world well enough for its stated purpose (Giere, 1991), and, (2) how much confidence to place in inferences about the real system that are based on model results (Curry et al., 1989). The former is validation, the latter is scientific hypothesis testing.

For this discusssion, the term 'model' refers generally to computer simulation models, but many of the points are applicable to mathematical and theoretical models as well. Although the focus is on validation, the concepts of verification, calibration, qualification and credibility must also be considered to place the validation step in the proper context of model building and evaluation.

My arguments are that (a) models can indeed be validated as acceptable for pragmatic purposes, whereas theoretical validity is always provisional, (b) validation can be a useful model evaluation activity regardless of whether the model is declared validated or invalidated, (c) insufficient attention has been given to specifying validation criteria or standards, (d) the fear of producing an invalidated model inhibits the development of ecology, and (e) validation is not an essential activity for evaluating research models, but is important for building model credibility in the user community.

2. A chronological review of validation concepts in ecological literature

The authors cited in this review present more detailed and thoughtful discussion of validation ideas than can be represented in a synopsis. My intent is to provide an overview that I hope fairly characterizes the major points in the research cited.

The subject of validation began to appear explicitly in ecological modeling discussions in the 1960s. For over 20 years, the issue of whether a model can be treated as a scientific hypothesis has been debated without being resolved. Levins (1966) initiated the discussion, "A mathematical model is neither an hypothesis nor a theory. Unlike the scientific hypothesis, a model is not verifiable directly by experiment. For all models are both true and false. ... leave out a lot and are in that sense false, incomplete, inadequate. The validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems." Levins is obviously not using the term validation in a quantitative or technical sense, but rather as a value judgment based on experience with a model.

Goodall (1972) equated validation with testing to determine the degree of agreement between a model and the real system, and suggested that the appropriate question to ask of an ecosystem model is how good its predictions are, not whether it should be accepted or rejected in the sense of hypothesis testing. Although he stated that validation is never absolute, he did not suggest any validation standards.

Goodall suggested that the model data and field data used for comparisons should be statistically independent. The notion of validation by comparison to an independent set of data has subsequently been mentioned by many authors (e.g., Odum, 1983; Shugart, 1984; Jørgensen, 1986; Power, 1993). Goodall also made a distinction between testing the adequacy of a model's predictions for a particular ecosystem and generalization of its applicability to a range of ecosystems. A model might be useful for a particular ecosystem, but not be generalizable. Finally, he pointed out that a single complex ecosystem model was likely to predict some variables well and others not so well, thus complicating the issue of validation.

Caswell (1976) addressed the fundamental duality of model as scientific theory versus model as engineering practice by differentiating the theoretical component from the operational component of a model. He distinguished two general purposes for which models are constructed: understanding (which he equated with theoretical models) and prediction. The point that evaluation of a model depends on its purpose has also been stated repeatedly by scientists and engineers alike (e.g., Mankin et al., 1977; Rykiel, 1984; Hoover and Perry, 1989; Mayer and Butler, 1993). Caswell applied the term validation to predictive models and the term corroboration to theoretical models. He concluded that scientific models can be corroborated or refuted (falsified) in the sense of scientific hypothesis testing, while predictive models can be validated or invalidated in the sense of engineering performance testing. The same model can be judged on both grounds, and thus a model might simultaneously be declared predictively validated and scientifically refuted. Later, in defending theoretical ecology, he asserted that neither theoretical models nor empirical data are ever validated (Caswell, 1988).

Mankin et al. (1977) suggested that the objectives of model-building may be achieved without validating the model. The idea that validation is not a required activity is a significant point that has been generally ignored. Mankin et al. set as a criterion for a valid model that all model behavior must correspond to some real ecosystem behavior. They defined a useful model as one that correctly predicts some but not all behavior. On the basis of this distinction, they proposed that models should be judged on their usefulness rather than their validity because no ecosystem model could meet their criterion for validity. In essence, they shifted the question of validation to a question of usefulness.

Overton (1977) viewed validation as an integral part of the modelling process, and stated that it is a misconception that models are first built, then validated. He pointed out that modelling is an iterative process that has no definitive end without an explicit specification of requirements that the model is to meet. The specified objectives have been met when the validation criteria are met and the modelling activity ceases when the model achieves the validation standards set by the specifications. After the model is constructed and validated, it can be used to answer the questions for which it was designed, and sometimes other questions as well. In Overton's view, validation is strongly related to hypothesis testing. In his view, hypotheses cannot be proven true, and thus models cannot be validated in an absolute sense of proving that a particular form is correct, or even best. Nevertheless, he distinguished two regions of 'validation space', objective validation over the region of prescribed behavior and theoretical or extended validation over the prediction region. These regions roughly correspond to testing the model against data available for model development and testing model predictions against independent data (Goodall, 1972).

Holling (1978) pronounced it a fable that the purpose of validation is to establish the truth of the model. He expressed the view that models are hypotheses, which can only be falsified, and in this regard his opinion is opposite that of Levins (1966). He therefore asserted that invalidation to establish the limits of model credibility is the proper view of model testing. Thus, the issue is changed from establishing truth (which is evidently scientifically impossible) to increasing a model's credibility, "Provisional acceptance of any model implies not certainty, but rather a sufficient degree of belief to justify further action. ... In practice, the problem is one of model invalidation – of setting the model at risk so as to suggest the limits of its credibility". In essence, the more validation tests a model passes the more confident we become in its predictions. He considered the demands for 'valid' models unsound because models can only be invalidated. Although a model might pass many tests, the very next one might require rejection just as in testing a scientific hypothesis.

Shugart (1984) also took the position that models are complex hypotheses, and that evaluation of a model's performance is the hypothesis-testing step in the scientific method. He defined verification as "procedures, in which a model is tested to determine whether it can be made consistent with some set of observations" (region one of the Overton (1977) validation space; Goodall, 1972). He then defined validation as "procedures, in which a model is tested on its agreement with a set of observations that are independent of those observations used to structure the model and estimate its parameters". (region two of the Overton (1977) validation space; Goodall, 1972).

Botkin (1993) expressed the concern that usage of the terms verification and validation was not consistent with their logical meanings. He interpreted verification as evidence that confirms the accuracy or truth of something as contrasted with certifying that a model is a correct computational implementation. He interpreted validation as the logical procedure of drawing a valid conclusion from the premises of an argument, and considered validation by comparison to independent data as a contradictory meaning. He stated his intention to avoid using the terms verification and validation because of these conflicting meanings.

Few common threads run through this decade-long discussion. The only area of general agreement is that the purpose of a model ought to be clearly stated. Obviously, the issue of model as scientific hypothesis versus engineering practice has not been resolved by the debates thus far. Ecological models commonly combine the aims of theory and practice, and the same model may be tested for both general purposes.

3. Definitions

3.1. Verification

Since 'verify' and 'validate' are synonyms in ordinary language, we must assign special meanings to distinguish them for modelling purposes. Fishman and Kiviat (1968) are often cited as first defining these terms in the context of simulation modelling and differentiating between verification and validation. *Verification is a demonstration that the modeling formalism is correct.*

There are two types of verification errors: mechanical and logical. The first amounts to debugging a computer program and in mathematical models showing that the mathematics is mechanically correct. A more subtle and difficult verification problem is showing that the program logic is correct. Some program errors only appear under circumstances that do not routinely occur, and may not have been anticipated. Verification is a technical matter that relates to how faithfully and accurately ideas are translated into computer code or mathematical formalisms. For large models, it is extremely difficult to verify that the model is entirely accurate and error-free under all circumstances, and that modifications to existing code have only the intended effect. Models are thus generally verified for the normal circumstances in which they are expected to be applied, and such verification is presumed inapplicable if the model is run outside these circumstances.

It is important to distinguish verification logic which relates to program operation from conceptual model logic which refers to the ecological logic used in structuring the model. Hoover and Perry (1989) state, "The computer model is verified by showing that the computer program is a correct implementation of the logical model. Verifying the computer model is quite different from showing the computer model is a valid representation of the real system and a verified model does not guarantee a valid model". In their technical sense, a valid model is one whose scientific or conceptual content is acceptable for its purpose.

3.2. Calibration

Models typically have parameters and constants that need to be given values to produce numerical results. Ideally, these factors have a clear ecological basis from which they can be calculated. The process of determining these values is parameter estimation. *Calibration is the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set.* Model calibration is, in essence, the step of making a model as consistent as possible with the data set from which parameters are estimated (the verification process of Shugart, 1984). Calibration procedures can be used to estimate parameter values that are otherwise unknown.

3.3. Validation

As with verification, validation is better understood as a process that results in an explicit statement about the behavior of a model. *Validation is a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model* (e.g., Sargent, 1984; Curry et al., 1989). This demonstration indicates that the model is acceptable for use, not that it embodies any absolute truth, nor even that it is the best model available. For operational validation, the demonstration involves a comparison of simulated data with data obtained by observation and measurement of the real system (e.g., Parton et al., 1987; Mayer and Butler, 1993). Such a test cannot demonstrate the logical validity of the model's scientific content (Oreskes et al., 1994).

Validation demonstrates that a model meets some specified performance standard under specified conditions. It is often overlooked that the 'specified conditions' include all implicit and explicit assumptions about the real system the model represents as well as the environmental context. That is, a model is declared validated within a specific context which is an integral part of the certification. If the context changes, the model must be re-validated; however, that does not invalidate the model for the context in which it was originally validated. Validation is a yes or no proposition in the sense that a model does or does not meet the specified validation criteria. These criteria may include requirements for statistical properties (e.g., goodness-of-fit) of the data generated by the model, and thus are not necessarily deterministic.

Ambiguous situations can arise when the model meets some but not all of the criteria. The criteria may need to be prioritized, and the model may be validated with respect to these priorities. Because modelling is an iterative process, validation criteria may evolve along with the model. This is more typically the case with scientific research models than with engineering models.

3.4. Credibility

A credible model is one in which a user has sufficient confidence to base scientific and management decisions (Holling, 1978; Sargent, 1984). *Credibility is a sufficient degree of belief in the validity of a model to justify its use for research and decision making*. Credibility is relative to the particular context of the model. The credibility of a model is therefore related to the amount of knowledge available, the purpose of the model, and the consequences of any decisions based on it. A model with high operational credibility does not necessarily have high conceptual credibility. Credibility is a subjective qualitative judgment, and cannot be quantified in any absolute sense.

3.5. Qualification

Qualification is aimed at discovering the domain over which a validated model may properly be used, i.e., whether the model is acceptable for use in a more general context, and amounts to revalidating a model for new cases. As revalidation tests are passed, the domain of the model's applicability increases. When the model fails a revalidation test, its domain of applicability is qualified, i.e., restricted to those situations where it has been validated. The implication of qualification is that the model remains useful for those situations for which it has been validated irrespective of its inability to pass other revalidation tests. The generality of a model can be inferred only inductively from repeated qualification testing.

These definitions do not give any specific metrics by which to judge models precisely because no validation standards have been established for ecological models. Ecologists are most familiar with statistical tests and tend to make judgments on the basis of them. However, a variety of metrics are available and the pros and cons are under discussion (e.g., Mayer and Butler, 1993).

4. Validation concepts

The validation process can encompass a large number of tests. In light of the history of validation

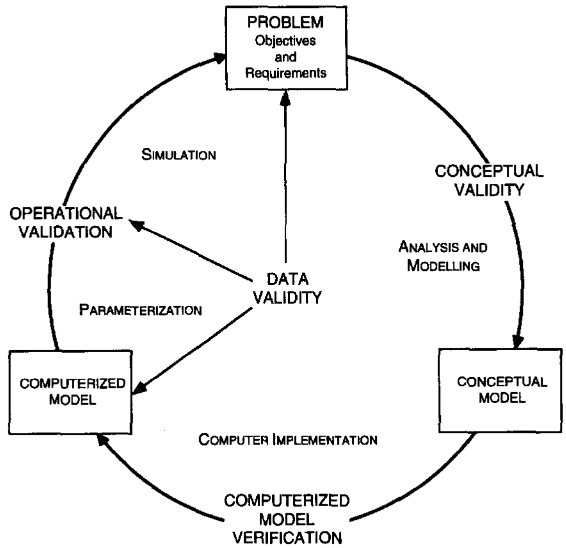

Fig. 1. A diagrammatic representation of the modelling cycle and the position of operational, conceptual, and data validation processes (modified from Sargent, 1984).

ideas for ecological models, it seems sufficient to consider a minimal partition of validation activities. Sargent (1984) described a simplified version of the modelling process that identified three principal areas where validity checking may be needed (Fig. 1). In his view, a model is validated only when it meets the design criteria for operational, conceptual, and data validity. Validation does not require that the model apply to more than one case unless that criterion is asserted as a validation requirement. That is, generality is no more required of a simulation model than it is of a regression equation. Testing procedures are used during model construction, at model completion, and for model qualification.

4.1. Operational validation (whole model validation)

Operational validation is a test protocol to demonstrate that model output meets the performance standards required for the model's purpose. This is the Caswell (1976) purpose of prediction and includes the Rykiel (1984) purpose of projection (qualitative correspondence with event dynamics). This process is a pragmatic approach to validation because it is concerned primarily with how well the model mimics the system regardless of the mechanisms built into the model. Statistical tests of comparisons between simulated and real data are widely used to evaluate model behavior (Mayer and Butler, 1993; Power, 1993). If the output of a model corresponds with observed data, then the model is an adequate representation of the system. However, such a correspondence does not guarantee that the scientific basis of the model and its internal structure correspond to the actual processes or cause–effect relationships operating in the real system. Although it may seem paradoxical, in general, it is not true that good predictions can only be obtained from a model that is mechanistically correct. Operational validation is the engineering side of the validation process that seeks to demonstrate model capability and continues the model building process until a prescribed level of performance is attained. Failure to meet an operational validation test can also reveal underlying conceptual problems (e.g., Fleming and Shoemaker, 1992).

4.2. Conceptual validity

Conceptual validity means that the theories and assumptions underlying the conceptual model are correct, or at least justifiable, and that the model representation of the problem or system, its structure, logic, mathematical, and causal relationships, are reasonable for the model's intended use. Conceptual validity depends on providing a scientifically acceptable explanation of the cause–effect relationships included in the model. Alternatively, justification is given for using simplifications of known processes, and for conjectured relationships for poorly known processes and mechanisms. Such justification may extend to providing a rationale for leaving out processes known to be involved in ecosystem behavior, and for using representations known to be ecologically false, i.e., for using a conceptually invalid model. Because the model is an abstraction of the real system, many components and processes must be left out for the purposes of a particular model. Conceptual validity, considered as acceptable justification of the scientific content of the model, does not guarantee that the model will make accurate predictions.

4.3. Data validation

Data are not an infallible standard for judging model performance (Fagerstrom, 1987). Rather the model and the data are two moving targets that we try to overlay one upon the other. Validation of data certifies that the data meet a specified standard (quality assurance/quality control). We cannot assume that data accurately represent the real system and therefore constitute the best test of the model. The validity not only of the data but also of its interpretation must also be demonstrated. Holling (1978) cited an instance where a model was re-analyzed for two months when it was finally determined that the data had been misinterpreted and the model was not at fault. The relative inaccuracy and imprecision of ecological data also places limits on model testability. Obviously, computer simulation models cannot be expected to provide results that are more accurate and precise than the data that are available. Conversely, it can be argued that the model may be a better representation of reality than data that are limited by our technological abilities for measurement and subjectively biased by our perceptions of the system.

4.4. Validation procedures

Sargent (1984) discussed some of the tests used for model validation. These tests are described briefly to provide a feeling for simulation industry validation concepts and a general sense of the variety of possible validation tests. This list is representative, not exhaustive (e.g., McCarl, 1984). Validation tests include both qualitative and quantitative measures of system performance, and project-specific tests are common. In commercial and government systems development projects, independent verification and validation are often required (Lewis, 1992). This technique has not been pursued in ecological modelling because there is no institutional or scientific infrastructure for concurrent independent verification and validation of another ecologist's simulation model. In addition, ecological simulation models often develop in an ad hoc fashion rather than as highly structured software development projects. In contrast to industry and government, no structured reporting requirement for the verification and validation processes exists in ecological modelling (or scientific modelling in general), although proposals for reporting guidelines have begun to appear (e.g., Tsang, 1991; Bart, 1995).

4.4.1. Face validity

Knowledgeable people are asked if the model and its behavior are reasonable. This test suggests whether the model logic and input–output relationships appear reasonable 'on the face of it' given the model's purpose. Some ecological models have high face validity by virtue of their longevity and wide spread use.

4.4.2. Turing tests

Knowledgeable individuals are asked if they can discriminate between system and model outputs. Appeals to the reader to observe how closely the simulated and actual data match in a graphical display (e.g., a time series graphic) are essentially unsupervised Turing tests combined with visualization. The subjective elements of this test need to be carefully considered (Mayer and Butler, 1993).

4.4.3. Visualization techniques

Time series plots, state space phase plots and other visual displays form the basis for comparisons between system and model. Most often, validation is determined subjectively by a statement that extols the visual goodness of fit.

4.4.4. Comparison to other models

The output of one model can be compared to that of another model. In some cases, such as global climate models, this may be the principal means of evaluation (Cess et al., 1990). Comparisons of ecological models are just beginning to occur on an ad hoc basis (Ågren et al., 1991).

4.4.5. Internal validity

A test data set (initial conditions, parameter values, and input data for driving variables) can be shown to produce a consistent output each time the model is run. This test is particularly applicable to stochastic models.

4.4.6. Event validity

A comparison between the model and system is made of the occurrence, timing and magnitude of simulated and actual events. Event validity may also be interpreted as qualitative validation in which the model is tested for its ability to reproduce the proper relationships among model variables and their dynamic behavior rather than to accurately reproduce their quantitative values.

4.4.7. Historical data validation

When historical data exist, part of the data is used to build the model and part is used to test if the model behaves as the system does. This procedure is also referred to as data-splitting (Power, 1993).

4.4.8. Extreme-condition tests

The model structure and output should be plausible for extreme or unlikely combinations of factors in the system. This test reveals if behavior outside of normal operating conditions is bounded in a reasonable manner.

4.4.9. Traces

The behavior of specific variables is traced through the model and through simulations to determine if the behavior is correct and if necessary accuracy is obtained.

4.4.10. Sensitivity analysis

The same relationships that occur in the system should occur in the model. Those parameters that cause significant changes in the model's behavior should be estimated with greatest accuracy. Frequently, there is disparity between parameters to which the system is sensitive and those to which the model is sensitive, but this issue is seldom mentioned.

4.4.11. Multistage validation

Validation methods are applied to critical stages in the model building process: (1) design: develop the model's assumptions based on theory, observations, general knowledge, and intuition; (2) implementation: empirically test the model's assumptions where possible; and (3) operation: compare the input–output relationships of the model and the real system. The three validation steps correspond roughly to conceptual, data, and operational components.

4.4.12. Predictive validation

The model is used to forecast the system behavior and comparisons are made to determine if the system's behavior and the model's predictions are the same. The system data may come from data sets not used in model development or from future observations of the system. The strongest case is when the model output is generated *before* the data are collected. Evaluation of model predictions is often considered as a kind of hypothesis testing.

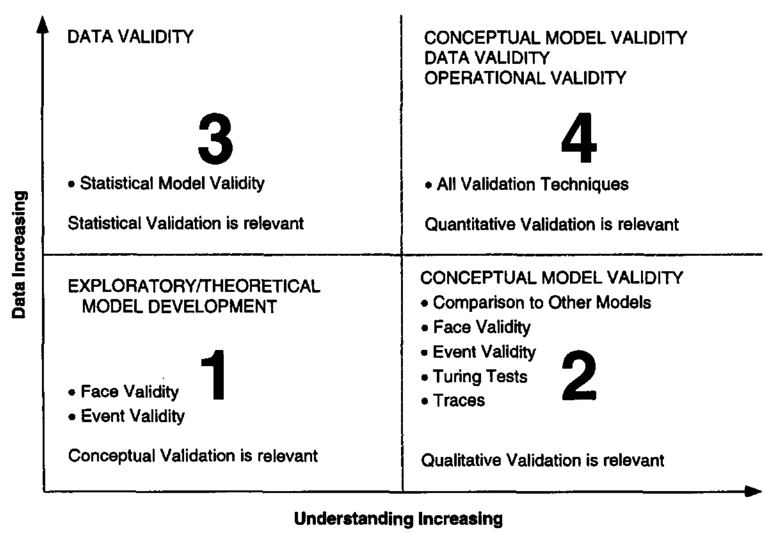

Fig. 2. Classes of modelling problems in relation to available data and understanding. Variation in the amount of data available and the level of understanding of the system influences the types of validation tests that can be conducted (modified from Holling, 1978, and Starfield and Bleloch, 1986).

4.4.13. Statistical validation

Statistical validation includes a variety of tests performed during model calibration and operation. Three cases occur most often: (1) the model produces output that has the same statistical properties as the observations obtained from the real system; (2) the error associated with critical output variables falls within specified or acceptable limits; (3) several models are evaluated statistically to determine which best fits the available data.

The relevance of various validation tests can be conceptualized in terms of the data and understanding available for the system being modelled (Fig. 2).

5. Semantics, logic, and philosophy

The rational underpinnings of validation concepts are seldom considered, yet they are a strong influence on how we view the testing of models (e.g., Oreskes et al., 1994). Arguments about validation cannot be resolved at an operational level until a common ground is established. This common ground is based on three areas: the meanings of the terms, the forms of reasoning, and philosophical perspectives of individual scientists and engineers.

5.1. Semantics

The meanings of the words verification and validation have themselves been a problem (Oreskes et al., 1994). For example, the conclusion that validation is impossible is most often based on the argument that truth is certain and we as scientists have no certain knowledge. Therefore, we can never establish that a model is 'true' (Reckhow and Chapra, 1983; Starfield and Bleloch, 1986; Swartzman and Kaluzny, 1987). In effect, validation is equated with certainty rather than a degree of belief (Holling, 1978). Inevitably, the meanings of truth and validity must come into play. There are a number of definitions of 'truth' depending on the context (Merriam-Webster, 1975). Definitions useful for modelling purposes are: (1) the quality or property of being in accord with fact or reality; (2) a judgment, proposition, or idea that is true or accepted as true; and also, (3) fidelity to an original or standard. For model validation, these definitions correspond respectively to: (1) consistent with available data, (2) in accord with current knowledge and beliefs, and (3) in conformance with design criteria. To the extent that truth is an issue, validation means establishing the truth of a model in these latter senses. Scientific truth is relative to what is known and believed to be true at the time a model is constructed.

The terms confirmation and corroboration have been proposed as alternatives (Swartzman and Kaluzny, 1987; Oreskes et al., 1994). However, 'validate' means to support or corroborate on a sound or authoritative basis. For a synonym, the dictionary points to confirm and for an antonym to invalidate. The entry for confirm indicates that the terms authenticate, confirm, corroborate, substantiate, validate, and verify are synonyms with the shared meaning element "to attest the truth or validity of something". Therefore, at the level of ordinary language, neither 'confirmation' (Oreskes et al., 1994) nor 'corroboration' has a greater claim on provisional acceptance than does validation. Unless we make some technical distinctions, these terms all mean essentially the same thing. There is no compelling semantic reason to reject the term validation on grounds that scientific truth cannot be established with certainty, or that validation implies a strictly logical process of deduction.

Valid means well grounded or justifiable: being at once relevant and meaningful; having a conclusion correctly derived from premises; and also, appropriate to the end in view. In other words, it may be valid to ignore known ecological mechanisms when that is appropriate to the end in view. The synonyms of valid are cogent, convincing, sound, and telling, which have the shared meaning element: having such force as to compel serious attention (i.e., credibility) and usually acceptance. Validation therefore is the process of showing that a model (1) accords with the facts (data) as we know them at the time; and/or, (2) with our judgment of what is true or accepted as true in ecology; and/or, (3) is justifiable and appropriate for our purposes.

5.2. Logic and reasoning

5.2.1. Formal vs. material logic

The process of testing whether a conclusion (inference) drawn from several premises is correct can be expressed by two components: validity and soundness (Jeffrey, 1991). Formal logic specifies how to draw a valid conclusion from the premises of an argument. Validity is guaranteed by the form of the argument (hence form-al logic). If the premises are accepted as true, then the conclusion must follow. The argument is invalid when the premises are true but the conclusion is false. The soundness of reasoning requires in addition that the premises and conclusion correspond to material reality. A logically valid but unsound conclusion can be drawn when one (or more) of the premises are false. This distinction is important because it indicates that the results of a model can be simultaneously formally valid and materially invalid (or unsound). In a modelling context, this situation can arise because (1) we assume but really don't know that the premises are true, or (2) we have insufficient information contained in the model. Logical validity does not necessarily mean that you get the right answer; it only says that the answer you get follows logically from the information you have. The conclusion is unsound if at least one of the premises is false (Jeffrey, 1991). Logical validity is therefore not a sufficient test of a theory or model and should not be equated with model validation.

Neither is a simulation model a formal logical structure like a syllogism, and the logical process of deduction cannot be applied in any strict formal sense. In fact, one of the primary reasons for building a simulation model is that it is impossible to deduce the behavior of a complex set of interacting components on purely logical grounds. The model allows us to see the consequences that we cannot compute in our heads, and is aptly described as an 'assumption analyzer' (Richard Holthausen quoted in Bart (1995)).

5.2.2. *Reasoning by analogy*

An analogy is a statement that two dissimilar things (ideas, objects, processes) are comparable in one or several respects even though they differ greatly in other respects (Toulmin et al., 1984; Giere, 1991). Inferences that can be made about the one can also be applied to the other with respect to their similarities. If the analogy is successful, the things compared may have other features in common that expand the analogy (Leatherdale, 1974). By their nature, all analogies fail at some point because they compare dissimilar things, e.g., computer model with forest ecosystem. Only the particular properties that the model shares with the material system can be tested against our beliefs of reality. As an analogy, a model need only accord with reality or be justified on those specific points that constitute the analogy.

For example, Darwin (1859) drew an analogy between domestic animal breeding and the processes of natality and mortality in wild populations to conclude that natural selection could lead to the generation of new forms and hence evolution of species in the wild. The model (theory) is not falsifiable by any single statistical or logical test. Darwin accumulated an enormous amount of evidence that the analogy was useful and therefore that the observations of changes resulting from domestic animal breeding could be used to draw inferences about wild populations.

Analogical reasoning is not precise. Some will reject as unwarranted an analogy that others will accept. Yet, the human mind is adept at drawing analogies and exploring their consequences. Analogies, both conscious and unconscious, often form the basis for the flashes of insight and inspirations that lead to discoveries and theoretical advances. If we accept a model as an analogy to the real system, then we must accept that analogical reasoning cannot be tested in the same way as formal logical deduction.

5.3. *Philosophies of science*

The third problem encountered in understanding model validation is the philosophical perspective of the individual scientist. I do not intend a review of the philosophy of science, but only to make a few points that seem relevant to the validation issue. Many of the suggestions that validation is impossible rest on the notion that falsification is the critical activity of science. The philosophy of Popper (e.g., in Klemke et al., 1988) is most often cited as the basis for this belief: "... the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability". Popper was seeking a method for distinguishing science from non-science (e.g., Marxist theory of history from Einstein's theory of relativity). Furthermore, Popper rejected the belief that science forms generalizations (theory) by logical induction from particular instances (empirical observations and model simulations), ''Induction, i.e., inference based on many observations, is a myth. ... The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation ... Repeated observations and experiments function in science as tests of our conjectures or hypotheses, i.e., as attempted refutations. ... None of this is altered in the least if we say that induction makes theories only probable rather than certain''.

However, philosophies of science are not only various, but also contradictory (see Klemke et al., 1988). Falsifiability is not accepted as the ultimate (or even sufficient) criterion for scientific testing by all philosophers of science (Thagard, 1988). Neither is Popper's rejection of induction shared by all philosophers of science (e.g., Trusted, 1979) and certainly not by all ecologists (e.g., Mentis, 1988).

Single-minded focus on falsification is a superficial treatment of a complex subject that includes, for example, much thought about what constitutes scientific explanation and what constitutes a theory (e.g., Suppe, 1989). Theories often contain unobservable components and thus cannot always be falsified directly (Mentis, 1988). For example, competition coefficients in Lotka–Volterra population models are not observable quantities. Practically all simulation models contain unobservable quantities, quantities that can only be observed with significant error, and calculations that cannot be compared to data because no data exist. O'Neill et al. (1989), for example, remark on how few field studies of multiple nutrient interactions exist in the literature (see also Caswell, 1988). The impulse to falsify can result in 'naive falsification'. For example, the fact that the earth's carbon budget does not balance according to current data could naively be taken as a falsification of ecosystem theory that asserts that the earth is an ecologically closed system with respect to matter.

An overemphasis on the philosophical belief that scientific hypotheses cannot be proven but only disproved may contribute to misunderstanding of model validation in another way. In one sense, the falsification argument fails to acknowledge the physical reality of the world. A premise can be proven true by showing that the thing can actually be done or is otherwise objectively verifiable even when scientific understanding is incomplete or even lacking. Such a demonstration is commonly termed 'proof of concept' in engineering jargon. This idea is really the basis for validation testing by comparison of simulated versus observed data.

Finally, as much as we hate to admit it, we do not understand how humans think. Scientific method, hypothesis formation, analysis, and practice, cannot yet be reduced to a series of unequivocal steps which anyone can follow to make new discoveries, enunciate new theories, and validate models. Model evaluation brings to bear all the subjective and objective elements of conscious and unconscious thought and reasoning of which the individual scientist is capable. In the words of Goodman (quoted in Thagard, 1988), ''A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend. The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either''.

6. Discussion

It is increasingly common for modellers to claim that a model has been validated without any reference to validation criteria. Statements such as, ''The model will be validated by comparison with empirical data'', indicate operational validation but are otherwise meaningless until the standards of comparison are specified. Furthermore, operational validation implies nothing about the mechanistic soundness of the model, though it increases the model's credibility. Thus, the ability to predict independent data or future values does not mean that the model is an accurate cause and effect representation of the real system. In addition, a model may accurately simulate the qualitative behavior of the system without quantitative accuracy.

6.1. *Qualification versus invalidation*

When operational validation is the objective, the difference between qualification and invalidation becomes an important distinction. Qualification is testing aimed at determining the domain over which a model is applicable. The model is expected to fail outside this domain (or context). Consequently, a model may be subjected to increasingly stringent tests to determine the conditions under which it fails to be a satisfactory simulator of the real system. When a model fails a validation test, several options are available. (1) The model may be re-calibrated to improve its fit to data by changing parameter values. (2) The model may be modified structurally and conceptually by revising assumptions and by changing the mathematical or logical representation of processes. (3) The application of the model may be restricted to a smaller domain where it is able to pass the validation test or where the particular test is not important. (4) Finally, failure to pass a validation test may be considered to invalidate the model. Invalidation implies that the model cannot pass a comparison test deemed essential for its credibility, acceptance, or usefulness. Notice that the implication is not that the scientific content of the model is erroneous; it may simply be insufficient. The scientific content is embodied in the assumptions, hypotheses and theories on which the model is based and on the system the model is intended to simulate. For complex simulation models, falsification has little or no traditional meaning because the scientific content is too extensive to be proven wrong en masse. Invalidation can only show that a model does not or cannot meet its validation criteria. Parts of the model may well be operationally and conceptually acceptable even though the integrated model is unable to pass particular validation tests. Conversely, the integrated model may be validated even though parts of the model are scientifically invalid.

6.2. Validation and scale

Hierarchy theory indicates that the interpretation of mechanisms varies with the spatial and temporal scales at which a system is observed and modelled. Validation criteria can likewise vary with scale. For example, the level of mechanistic detail that is considered valid and the allowable error tolerances for simulating system behavior are likely to be quite different for a model of the global carbon budget versus a model of leaf carbon balance. Consequently, our notion of what is acceptable depends not only on the objectives of the model but also on the ecological scale at which the model is framed. Changes in scale may permit aggregations that at face value are conceptually erroneous to produce operational results that are acceptable.

6.3. Model validation vs theoretical development

There are many parallels between model development and testing and theory development and testing. A variety of studies suggest that prematurely imposing rigorous testing requirements can result in rejection of correct or at least useful theories (Loehle, 1987). Just as development of ecological theory can be stifled by an overemphasis on hypothesis testing (Fagerstrom, 1987; Mentis, 1988), modelling and the benefits to be gained from it can also be stifled by an overemphasis on model validation.

Neither falsification nor validation are requirements of theoretical and computer model development (Fagerstrom, 1987; Mentis, 1988; Thagard, 1988). The nature of the testing undertaken depends on the model's purpose. Purposes such as exploration, initial development, new perspectives, etc. do not require validation. Theories and models often start off with wrong elements which are discarded as development and understanding improve, and a significant loss of creativity occurs when models are not allowed to mature. Some things that are believed cannot be validated or falsified in any easy, obvious, or immediate way. Caswell (1988) has argued forcefully that theoretical models are useful without any attempt at validation and even when they are refuted.

6.4. Validation and policy

Oreskes et al. (1994) argue that usage of the terms verification and validation imply that the models are 'true' and that this implication is inherently wrong. They call for a neutral language to avoid this implication so policy makers will not be misled by fallible models. It is hard to see how this translation to a neutral language could be accomplished. Consider for example what term ecologists would use to replace 'succession' to conform to the lay understanding of that term while retaining the ecological content. In such cases, the better choice may be to educate the audience.

To the extent that the technical meanings of verification and validation may be misconstrued, modellers themselves should take the lead in asserting the restrictions and limitations of models, and should draw some important lessons from Oreskes et al. (1994). (1) Make clear that verification and validation are used in a technical sense. (2) Carefully specify the context of the model. (3) Use model acceptability (Bart, 1995) and performance indices rather than simple declarations of validity to describe the results of model testing to general audiences. Policies and decisions must be made and models, however imperfect, are needed to assist the process. We have sufficient knowledge to accomplish many useful tasks and build a variety of devices whose very existence demonstrates that pragmatic operational validation is possible despite our philosophical uncertainties.

7. Summary

Validation is just one component of the larger task of model evaluation. Validation describes a test or usually a testing process on which to base an opinion of how well a model performs so that a user can decide whether the model is acceptable for its intended purpose. Invalidation means that a model is unable to perform at the required level. Validation procedures vary from general qualitative tests to highly restrictive quantitative tests. The nature of the tests and their interpretation may depend on one's philosophical beliefs, but the specific procedures are the same.

7.1. Is validation possible?

In the limited technical sense of simulation modelling, validation is certainly possible and often essential for user acceptance. It is common for models to accurately predict the outcome of experiments and operational behavior. Modelling is a fundamental step in engineering design and practice. In a pragmatic sense, a model only needs to be good enough to accomplish the goals of the task to which it is applied.

The context in which a model is developed assumes a conceptual closure of the model system, not the real system, and validation is conducted within this context. There is always the possibility that the context will change or is different than assumed, in which case the model may not apply. In a restricted technical sense, verification and validation are therefore legitimate and useful concepts.

7.2. Does the model require validation?

The purpose of the model may dictate that validation is not a useful activity. Validation may be difficult or impossible because of the lack of data. Exploration of model behavior without validation testing is a legitimate, reportable activity (Caswell, 1988). Model development is a significant scientific contribution in itself without any validation tests being undertaken in addition. When the principal purpose of the model is to describe or systematize knowledge or to develop theory, validation is unnecessary and irrelevant.

Empirical scientists may treat model validation as an obligation of modellers. Reviewers may demand that models be 'validated' as a requirement for publication. However, modellers need not test their own models, nor do they have sole responsibility for validating models. To the extent that the model is a scientific experiment and theoretical development, its testing and validation are within the purview of the scientific community. The task of model development is often so complex in itself that it is legitimate to report models that do not make direct comparisons with field and laboratory data. Theory and models may be proposed even when the author has undertaken no experimentation, or has only suggested what observations could be made. A theory may explain previous observations, but may also suggest other things that ought to be looked at and may not yet have been measured, in which case there may be little or no data to be explained. The notion that models cannot be 'truly validated' is irrelevant and only serves to prolong confusion and argument about the proper role of validation in model testing.

7.3. Purpose, criteria, context

Whenever validation is required, the modeller must specify three things: (1) the purpose of the model, (2) the criteria the model must meet to be declared acceptable for use, and (3) the context in which the model is intended to operate. The latter two items are seldom specified for ecological models. Without them, a model cannot be validated. The context is particularly overlooked. Context embodies all the assumptions, especially those that are unstated and relegated to the system environment of the model. For example, a vegetation model of Mount Saint Helens might not include volcanic eruptions in its context. The model might assume that vegetation dynamics are occurring in a nominal (i.e., non-eruptive) environment, and thus would not be applicable to primary succession following an eruption. Therefore, the model could be validated only for the nominal environment. Ecological models can be so complex that it is not feasible for the modeller to state all assumptions, particularly those that an ecological audience would be expected to know from common scientific knowledge.

Because models are approximations of reality, it is always possible to establish a validation criterion that no model can meet, but such a criterion is not a fair test. On the other hand, the statement that a model has been validated is misleading without stating the purpose of the model, the validation criteria used, and the context to which the claim applies.

7.4. Engineering validation versus scientific validity

Performance testing is fundamentally limited to showing that the model passes the validation tests devised for it. Certification is completed when the model is shown to meet its requirements specification. In the majority of cases to date, this has been understood as some measure of agreement between simulated and observed data and expert judgment. Ecological modellers do not describe in an engineering sense the requirements specifications of the models they build. Because there are no generally agreed validation criteria for ecological models, the best that can be done at present is for the modeller to state explicitly what the validation criteria are and leave it to the user to judge if the criteria are adequate. The most common criteria at present are the "see how well the simulated data matches the observed data" test and the "the model did a reasonable job of simulating..." test in which the reader is asked to agree subjectively that the match is adequate (combinations of face validity, Turing test, and visualization). While validation testing may result in accumulating evidence that the scientific content of a model is correct, such testing cannot logically prove that the mechanisms contained in the model are scientifically complete and correct.

7.5. Need for an operational validation convention

McCarl (1984) emphatically states, "There is not, and never will be, a totally objective and accepted approach to model validation". Once we face up to the fact that concepts like validation have strong subjective elements, it becomes clear that a functional definition requires the establishment of a convention, which is a generally accepted standard (Costanza, 1989; Botkin, 1990; NCASI, 1990). An arbitrary convention that all ecologists are familiar with is the use of a probability level of 0.05 as a test for statistical significance. There is no purely objective basis for this standard.

An appropriate convention may include not only how close to observed data the simulation is, but also how often. An initial proposal is that model outputs fall within the 95% confidence interval 75% of the time for the most important variables in dyamic models. The purpose of the model and the problem-specific level of accuracy required, for example a risk assessment requirement, are additional factors that will influence the construction of a general validation convention and provide the basis for refinements of the convention.

Acknowledgements

This work was supported by NSF Grant BSR-9109240. Mark Fulton, W. Michael Childress, the graduate students in BIEN 674, and many others in informal discussions provided valuable comments and criticisms.

References

Ågren, G.I., McMurtrie, R.E., Parton, W.J., Pastor, J. and Shugart, H.H., 1991. State-of-the-art models of production–decomposition linkages in conifer and grassland ecosystems. Ecol. Appl., 1: 118–138.

Bart, J., 1995. Acceptance criteria for using individual-based models to make management decisions. Ecol. Appl., 5: 411–420.

Botkin, D.B., 1990. Expert Panel Report 1. In: Minimum Standards/Criteria for Assessing Models of Air Pollution Impact on Stand Productivity. NCASI Technical Bulletin No. 593, National Council of the Paper Industry for Air and Stream Improvement, New York, NY, pp. 6–16.

Botkin, D.B., 1993. Forest Dynamics. Oxford University Press, New York, NY, 309 pp.

Caswell, H., 1976. The validation problem. In: B. Patten (Editor), Systems Analysis and Simulation in Ecology, Vol. IV. Academic Press, New York, NY. pp. 313–325.

Caswell, H., 1988. Theory and models in ecology: a different perspective. Ecol. Model., 43: 33–44.

Cess, R.D., Potter, G.L., Blanchet, J.P., Boer, G.J., Del Genio, A.D., Deque, M., Dymnikov, V., Galin, V., Gates, W.L., Ghan, S.J., Kiehl, J.T., Lacis, A.A., Le Treut, H., Li, Z.-X., Liang, X.-Z., McAvaney, B.J., Meleshko, V.P., Mitchell, J.F.B., Morcrette, J.-J., Randall, D.A., Rikus, L., Roeckner, E., Royer, J.F., Schlese, U., Sheinin, D.A., Slingo, A., Sokolov, A.P., Taylor, K.E., Washington, W.M., Wetherald, R.T., Yagai, I. and Zhang, M.-H., 1990. Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. J. Geophys. Res., 95: 16601–16615.

Costanza, R., 1989. Model goodness of fit: a multiple resolution procedure. Ecol. Model., 47: 199–215.

Curry, G.L., Deuermeyer, B.L. and Feldman, R.M., 1989. Discrete Simulation. Holden-Day, Oakland, CA, 297 pp.

Darwin, C.R., 1859. On the Origin of Species. Murray, London.

Fagerstrom, T., 1987. On theory, data and mathematics in ecology. Oikos, 50: 258–261.

Fishman, G.S. and Kiviat, P.J., 1968. The Statistics of Discrete-Event Simulation. Simulation, 10: 185–191.

Fleming, R.A. and Shoemaker, C.A., 1992. Evaluating models for spruce budworm–forest management: comparing output with field data. Ecol. Appl., 2: 460–477.

Gentil, S. and Blake, G., 1981. Validation of complex ecosystems models. Ecol. Model., 14: 21–38.

Giere, R.N., 1991. Understanding Scientific Reasoning. Harcourt Brace Jovanovich College Publishers, New York, 322 pp.

Goodall, D.W., 1972. Building and testing ecosystem models. In: J.N.J. Jeffers (Editor), Mathematical Models in Ecology. Blackwell, Oxford, pp. 173–194.

Holling, C.S., 1978. Adaptive Environmental Assessment and Management. John Wiley & Sons, New York, NY, 377 pp.

Hoover, S.V. and Perry, R.F., 1989. Simulation. Addison-Wesley, Reading, MA, 696 pp.

Jeffers, J.N.R., 1978. An Introduction to Systems Analysis: With Ecological Applications. University Park Press, Baltimore, MD, 198 pp.

Jeffrey, R., 1991. Formal Logic. McGraw-Hill, New York, NY, 198 pp.

Jørgensen, S.E., 1986. Fundamentals of Ecological Modelling. Elsevier, Amsterdam, 389 pp.

Kleijnen, J.P.C., 1987. Statistical Tools for Simulation Practitioners. Marcel Dekker, New York, NY, 429 pp.

Klemke, E.D., Hollinger, R. and Kline, A.D. (Editors), 1988. Introductory Readings in the Philosophy of Science. Prometheus Books, Buffalo, NY, 440 pp.

Law, A.M. and Kelton, W.D., 1991. Simulation Modeling and Analysis. McGraw-Hill, New York, NY, 759 pp.

Leatherdale, W.H., 1974. The Role of Analogy, Model and Metaphor in Science. North-Holland, Amsterdam, 276 pp.

Levins, R., 1966. The strategy of model building in population biology. Am. Sci., 54: 421–431.

Lewis, R.O., 1992. Independent Verification and Validation: A Life Cycle Engineering Process for Quality Software. John Wiley & Sons, New York, NY, 356 pp.

Loehle, C., 1987. Hypothesis testing in ecology: psychological aspects and the importance of theory maturation. Q. Rev. Biol., 62: 397–409.

Mankin, J.B., O'Neill, R.V., Shugart, H.H. and Rust, B.W., 1977. The importance of validation in ecosystem analysis. In: G. Innis (Editor), New Directions in the Analysis of Ecological Systems, Part 1. The Society for Computer Simulation, La-Jolla, CA, pp. 63–71.

Mayer, D.G. and Butler, D.G., 1993. Statistical validation. Ecol. Model., 68: 21–32.

McCarl, B.A., 1984. Model validation: an overview with some emphasis on risk models. Rev. Market. Agric. Econ., 52: 153–173.

Mentis, M.T., 1988. Hypothetico-deductive and inductive approaches in ecology. Funct. Ecol., 2: 5–14.

Merriam-Webster, 1975. Webster's New Collegiate Dictionary. G. & C. Merriam, Springfield, MA, 1536 pp.

NCASI, 1990. Minimum Standards/Criteria for Assessing Models of Air Pollution Impact on Stand Productivity, 1990. NCASI Tech. Bull. 593, National Council of the Paper Industry for Air and Stream Improvement, New York, NY.

Odum, H.T., 1983. Systems Ecology. John Wiley & Sons, New York, NY, 644 pp.

O'Neill, R.V., DeAngelis, D.L., Pastor, J.J., Jackson, B.J. and Post, W.M., 1989. Multiple nutrient limitations in ecological models. Ecol. Model., 46: 147–163.

Oreskes, N., Shrader-Frechette, K. and Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263: 641–646.

Overton, S., 1977. A strategy of model construction. In: C. Hall and J. Day (Editors), Ecosystem Modeling in Theory and Practice: An Introduction with Case Histories. John Wiley & Sons, New York. Reprinted 1990, University Press of Colorado. pp. 49–73.

Parton, W.J., Schimel, D.S., Cole, C.V. and Ojima, D.S., 1987. Analysis of factors controlling soil organic matter levels in great plains grasslands. Soil Sci. Soc. Am. J., 51: 1173–1179.

Power, M., 1993. The predictive validation of ecological and environmental models. Ecol. Model., 68: 33–50.

Reckhow, K.H. and Chapra, S.C., 1983. Engineering Approaches for Lake Management, Volume 1: Data Analysis and Empirical Modeling. Butterworth, Boston, MA, 340 pp.

Rykiel, E.J., Jr., 1984. Modeling agroecosystems: lessons form ecology. In: R. Lowrance, B.R. Stinner and G.J. House (Editors), Agricultural Ecosystems: Unifying Concepts. John Wiley & Sons, New York, NY, pp. 157–178.

Sargent, R.G., 1984. A tutorial on verification and validation of simulation models. In: S. Sheppard, U. Pooch and D. Pegden

(Editors), Proceedings of the 1984 Winter Simulation Conference. IEEE 84CH2098-2, pp. 115–122.

Shugart, H.H., 1984. A Theory of Forest Dynamics. Springer-Verlag, New York, NY.

Starfield, A.M. and Bleloch, A.L., 1986. Building Models for Conservation and Wildlife Management. Macmillan, New York, NY, 253 pp.

Suppe, F., 1989. The Semantic Conception of Theories and Scientific Realism. University of Illinois Press, Urbana, IL, 475 pp.

Swartzman, G.L. and Kaluzny, S.P., 1987. Ecological Simulation Primer. Macmillan, New York, NY, 370 pp.

Thagard, P., 1988. Computational Philosophy of Science. MIT Press, Cambridge, MA, 240 pp.

Toulmin, S., Rieke, R. and Janik, A., 1984. An Introduction to Reasoning. Macmillan, New York, NY, 435 pp.

Trusted, J., 1979. The Logic of Scientific Inference. Macmillan, London, 145 pp.

Tsang, C.-F., 1991. The modeling process and model validation. Ground Water, 29: 825–831.