Robert A. Meyers (Ed.)

Encyclopedia of Complexity and Systems Science

With 4300 Figures and 420 Tables

Volume 9

S

Springer

UNIVERSITÄT ST.GALLEN
HOCHSCHULE FÜR WIRTSCHAFTS-,
RECHTS- UND SOZIALWISSENSCHAFTEN
BIBLIOTHEK

ROBERT A. MEYERS, Ph. D.
Editor-in-Chief
RAMTECH LIMITED
122 Escalle Lane
Larkspur, CA 94939
USA
robert.meyers@ramtechlimited.org

Library of Congress Control Number: 2008933604

ISBN: 978-0-387-75888-6

This publication is available also as:

Electronic publication under ISBN: 978-0-387-30440-3 and
Print and electronic bundle under ISBN: 978-0-387-69572-3

© 2009 SpringerScience+Business Media, LLC.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

springer.com

Printed on acid free paper

SPIN: 11560258 2109letex – 5 4 3 2 1 0

59. Wheat D (2007) The feedback method of teaching macroeconomics: Is it effective? *Syst Dyn Rev* 23(4):391–413
60. Yamaguchi K (2007) Balance of payments and foreign exchange dynamics – S D Macroeconomic Modeling (4). In: Sterman J, Oliva R, Langer R, Rowe J, Yanni J (eds) *Proceedings of the of the Twenty-Fifth International Conference of the System Dynamics Society*. Available at <http://www.systemdynamics.org/>
61. Zellner A (1980) Comment on 'information sources for modeling the national economy' by Jay W Forrester. *J Am Stat Assoc* 75(371):567–569

Books and Reviews

- Alfeld L, Graham A (1976) *Introduction to urban dynamics*. Wright-Allen Press, Cambridge
- Lyneis J (1980) *Corporate planning and policy design: A system dynamics approach*. PA Consulting, Cambridge
- Meadows D, Meadows D (1973) *Toward global equilibrium: Collected papers*. Wright-Allen Press, Cambridge
- Meadows D, Behrens III W, Meadows D, Naill R, Randers J, Zahn E (1974) *Dynamics of growth in a finite world*. Wright-Allen Press, Cambridge
- Meadows D, Robinson J (1985) *The electronic oracle: Computer models and social decisions*. Wiley, New York
- Richardson G (1999) *Feedback thought in social science and systems theory*. Pegasus Communications, Inc., Waltham
- Sterman J (1988) A skeptic's guide to computer models. In: Grant L (ed) *Foresight and National Decisions*. University Press of America, Lanham, pp 133–169, Revised and Reprinted in: Barney G, Kreutzer W, Garrett M (1991) *Managing a Nation*. Westview Press, Boulder, pp 209–230

System Dynamics Modeling: Validation for Quality Assurance

MARKUS SCHWANINGER, STEFAN GROESSER
Institute of Management, University of St. Gallen,
St. Gallen, Switzerland

Article Outline

Glossary
Definition of the Subject
Introduction
Epistemological Foundations
Validation Methods
Validation Process
Synopsis and Outlook
Appendix: Overview of the Tests Described in This Chapter
Bibliography

Glossary

Model/model system A model is a simplified representation of a real system. Models can be descriptive or pre-

scriptive (normative). Their functions can be to enable explanation, anticipation or design. A distinction used in this contribution is between causal and non-causal models, with System Dynamics models being of the former type. The term *model system* is used to stress the systemic character of a model; this serves to identify it as an organized whole of variables and relationships on the one hand, and to distinguish it from the *real system* which is to be modeled, on the other.

Model validity A model's property of adequately reflecting the system modeled. Validity is the primary measure of model quality. It is a matter of degree, not a dichotomized property.

Model purpose The goal for which a model is designed or the function it is intended to fulfill. The model purpose is closely linked to the end-model user or model owner. Model purpose is the criterion for the choice of a model's boundary and design.

Modeling process The process involving phases such as problem articulation, boundary selection, development of a dynamic hypothesis, model formulation, model testing, policy formulation and policy evaluation [28]. The modeling process is followed by model use and implementation, i. e., the realization of actions designed or facilitated by the use of the model.

Validation process Validation is the process by which model validity is enhanced systematically. It consists in gradually building confidence in the usefulness of a model by applying validation tests as outlined in this chapter. In principle, validation pervades all phases of the modeling process, and, in addition, extends into the phases of model use and implementation.

Definition of the Subject

The present chapter addresses the question of building better models. This is crucial for coping with complexity in general, and in particular for the management of dynamic systems. Both the epistemological and the methodological-technological aspects of model validation for the achievement of high-quality models are discussed. The focus is on formal models, i. e. those formulated in a stringent, logical, and mostly mathematical language.

Introduction

The etymological root of *valid* is the Latin word *validus*, which denotes attributes such as strong, powerful and firm. A valid model, then, is well-founded and difficult to reject because it accurately represents the perceived real system which it is supposed to reflect. This system can

be either one that already exists or one that is being constructed, or even anticipated, by a modeler or a group of modelers.

Validation standards in System Dynamics are more rigorous than those of many other methodologies. Let us distinguish between two types of mathematical models, which are fundamentally different: Causal, theory-like models and non-causal, statistical (correlational) models [4]. The former are explanatory, i. e., they embody theory about the functioning of a real system. The latter are descriptive and express observed associations among different elements of a real system. System Dynamics models are causal models.

Non-causal models are tested globally, in that the statistical fit between model and data series from the real system under study is assessed. If the fit is satisfactory, the model is considered to be accurate ("valid", "true"). In contrast, system dynamicists postulate that models be not only right, but right for the right reasons. As the models are made up of causal interdependencies, accuracy is required for each and every variable and relationship. The following principle applies: if only one component of the model is shown to be wrong, the whole model is rejected even if the overall model output fits the data [4]. This strict standard is conducive to high-quality modeling practice.

A model is an abstract version of a perceived reality. Simulation is a way of experimenting with mathematical models to gain insights and to employ these to improve the real system under study. It is often said that System Dynamics models should portray problems or issues, not systems. This statement must be interpreted in the sense that one should not try to set the boundaries of the model too widely, but rather give the model a focus by concentrating on an object in accordance with the specific purpose of the model. In a narrower definition, even an issue or problem can be conceived of as a "system", i. e., "a portion of the world sufficiently well defined to be the subject of study" [21]. Validity then consists in a stringent correspondence between model system and real system.

We will treat the issue of model validation as a means of assuring high-quality models. We interject that validity is not the only criterion of model quality, other criteria including parsimony, ease-of-use, practicality, importance, etc. [22].

In the following, the epistemological foundations of model validity are reviewed (Sect. "Epistemological Foundations"). Then, an overview of the methods for assuring model validity is given (Sect. "Validation Methods"). Further, the survey includes an overview of the validation process (Sect. "Validation Process") and our final conclusions (Sect. "Synopsis and Outlook").

System Dynamics Modeling: Validation for Quality Assurance, Figure 1

The Validation Cube – A frame of reference showing three dimensions of the validation topic

The substance of this article will be made more palpable by means of the following frame of reference. We call it the Validation Cube. The diagram in Fig. 1 shows three dimensions of the validation topic:

- *Orders of Reflection:* We distinguish between an *epistemological* and a *methodological* layer. These define the objects of the next two Sects. "Epistemological Foundations" and "Validation Methods".
- *Domains of Validation:* The three domains, *context*, *structure* and *behavior* refer to the groups of validation methods as described in Sect. "Validation Methods".
- *Degrees of Resolution:* We address the different granularities of models. *Micro* refers to the smallest building blocks of models (e. g., variables or small sets of variables), *meso* to modules which constitute a model, and *macro* to the model as a whole.

Epistemological Foundations

Epistemology is the theory that enquires into the nature and grounds of knowledge: "What can we know and how do we know it?" [13]. These questions are of utmost importance when dealing with models and their validity, because a method of validation is only as good as its epistemological basis.

We can only briefly refer to the antecedents of the epistemological perspective inherent in the idea of model validation as commonly held today in the community of sys-

tem dynamicists. One could go back to Socrates who, in Plato's *Republic* (fourth century BC), addressed the problematic relationship between reality, image and knowledge. One could also refer to John Locke (seventeenth century), the first British empiricist who maintained that ideas could come only from experience, while admitting that our knowledge about external objects is uncertain. We will address the philosophical movements of the nineteenth and twentieth centuries, which are direct sources of the epistemology which is important for model validation. The reader may kindly excuse us for certain massive simplifications that we are obliged to make.

What will be said here about theories applies equally to formal models. In System Dynamics, models either embody theories or they are considered essential components of theories. In addition, processes of modeling and theory-building are of the same nature; a model, like any theory, is built and improved in a dialectic of propositions and refutations [22].

Positivism and Critique

Positivism is a scientific doctrine founded by Auguste Comte (nineteenth century) which raises the *positive* to the principle of all scientific knowledge. "Positive", in this context, is not meant to be the opposite of negative, but the given, factual, or indubitably existent. The positive is associated with features such as being real, useful, certain, and precise. Positivism confines science to the observable and manipulable, drawing on the mathematical, empirical orientation of the natural sciences as its paragon. The objectivist claim of positivism is that things exist independently of the mind and that truths are detached from human values and beliefs. This stance calls for models that approximate an objective reality.

A younger development in this vein is the school of logical positivism, also logical empiricism (with Schlick, Neurath, Hempel, etc.), which concentrates on the problem of meaning and has developed the verifiability principle: Something is meaningful only if verifiable empirically, i. e., ultimately by observation through the senses. To verify here means to show to be true [13]. For the logical positivists, the method of verification is the essence of theory-building. Tests of theories hinge on their confirmation by facts. In System Dynamics, testing models on real-world data is a core component of validation.

Positivism has been criticized for being reductionist, i. e., for its tendency to reduce concepts to simpler or empirically more accessible ones, and to conceive of learning as an accumulation of particular details. The critique has also asserted that there is no theory-independent identifi-

cation of facts, and therefore different theories cannot be tested by means of the same data [6]. Another objection maintains that social facts are not merely given, but produced by human action, and that they are subject to interpretation [23]. These arguments introduce the principle of relativity, which is of crucial importance for the field of model validation: A model is a subjective construction by an observer.

Pragmatism – A Challenge to Positivism

Pragmatism, which arose in the second half of the nineteenth century, emphasizes action and the practical consequences of thinking. Its founder, Charles Sanders Peirce, was interested in the effects that the meaning of scientific concepts could have on human experience and action. He defined truth as "the opinion which is fated to be ultimately agreed to by all who investigate" [13], whereby truth is linked to consensual validation. For pragmatists, truth is in what works (Ferdinand Schiller) or satisfies us (John Dewey), and what we find believable and consistent: "The true' ... is only the expedient in the way of our thinking", and "truth is *made* ... in the course of experience." (see p. 581 and p. 583 in [11]).

Pragmatism is often erroneously disdained for supposedly being a crass variety of utilitarianism and embodying a crude instrumentalist rationality. A more accurate view considers the fact that pragmatists are not satisfied with a mere ascertainment of truth; instead they ask: "If an idea or assumption is true, does this make a concrete difference to the life of people? How can this truth be actualized?" In other words, pragmatism does not crudely equate truth and utility. It rather postulates that those truths which are useful to people ought to be put into practice [23].

Pragmatism introduces the criteria of confidence and usefulness, which are more operational as guides to the evaluation of experiments than is the notion of an absolute truth, which is unattainable in the realm of human affairs. At the same time, pragmatism triggers a crucial insight for the context of model-building: The validity of a model depends not only on the absolute quality of that model but also hinges on its suitability with respect to a purpose [7]. In the context of model validation, then, truth is a relative property; more exactly, a *truth* holds for a limited domain only.

More Challenges to Positivism

We discuss three more challenges to positivism in the twentieth century. First, Thomas Kuhn's theory of scientific revolutions [12]: Kuhn shows, by means of historical cases, that in the sphere of science, generally accepted

ways of looking at the world ("paradigms") change over time through fundamental shifts. Therefore, the activities of a scientist are largely shaped by the dominant scientific worldview. Second, Willard Van Orman Quine and Wilfrid Sellars argue that knowledge creation and theory-building is a holistic, conversational process, as opposed to the reductionist and confrontational views [4].

Both of these movements contribute to our understanding of how real systems are to be modeled and validated: as organized wholes, and consciously with respect to the values and beliefs underlying a given modeling process. This approach adheres to the spirit of models themselves, by means of which the behavior of whole systems can be simulated and tested on their inherent assumptions.

A third challenge is presented by the interpretive streams of epistemology (for an overview, see [9]). Among them, a main force which expands the possibilities of scientific methodologies is the strand of hermeneutics. Derived from the Greek *hermeneuein* – to interpret or to explain – the term *hermeneutics* stands for a school, mainly associated with Hans-Georg Gadamer, which pursues the ideal of a human science of understanding. The emphasis is on interpretation in an interplay between a subject-matter and the interpreter's position. This emphasis introduces the subjective into scientific methodology. Hermeneutics denies both that a single "objective true interpretation" can transcend all individual viewpoints, and that humans are forever confined within their own ken [13]. This epistemology offers a necessary complement to a scientific stance, which exclusively hinges on "hard", quantitative methods in order supposedly to achieve absolute objectivity. The implication of hermeneutics for model validation is that it recognizes the pertinence of subjective judgment. In this connection, interpretive discourses play a crucial role in group model-building and validation. Such discourses lead beyond the subjective, entailing the creation of inter-subjective, shared realities. We will return to this factor in Sect. "Validation Process".

Critical Rationalism

Critical rationalism is a philosophical position founded by Karl R. Popper [19,20]. It grew out of positivism but rejected its verificationist stance. Critical rationalism posits that, in the social domain, theories can never be definitely proved, but can only reach greater or lesser levels of truth. Scientific proofs are confined to the realm of the formal sciences, namely logic and mathematics.

As Popper demonstrates, all theories are provisional. As a consequence, the main criterion for the assessment of a theory's truth status is *falsification* [19]. A theory holds as

long as it is not refuted. Consequently, any theory can be upheld as long as it passes the test of falsification. In other words, the fertile approaches to science are not those of corroboration, but the falsificationist efforts to test if theories can be upheld. In the context of modeling this means that validation must undertake attempts to falsify a model, thereby testing its robustness.

Even Popper's theory of science is not unchallenged. For example, Kuhn has made the point that its principles are applicable only to normal science, which operates incrementally within a given paradigm, but not to anomalous science, which uncovers unsuspected phenomena in periods of scientific revolution [12]. This observation has an implication for model validation: Alternative and even multiple model designs should be assessed for their ability to account for fundamental change.

On the Meaning of Validity and Validation

One of the predominant convictions about science is the obsessive idea that proofs are the touchstone of the validity of both theories and models. We follow a different rationale, reverting to the philosophy of science as embodied in critical rationalism.

Popper's refutationist concept (as opposed to a verificationist concept) of theory-testing implies both an evolutionist perspective and an empiricist stance. The evolutionist perspective is primary because it welcomes the challenges posed to a theory, since these attempts at falsification lead to an evolutionary process: successful falsification efforts result in revisions and improvements of the theory. Correspondingly, empiricism is paramount in the social sciences, because the main source for the refutation of a theory is empirical evidence. However, falsification can also be grounded in logical arguments where empirical evidence cannot be obtained. In this sense, a structuralist approach as used in System Dynamics validation transcends the bounds of logical empiricism.

As a consequence of the evolutionist perspective, there is no such thing as absolute validity. Validity is always imperfect, but it can be improved over time. The empiricist aspect of theory-building implies that theories must be validated by means of empirical data. However, logical assay, estimation and judgment are complementary to this empiricist component (see below).

A validation process is about gradually building confidence in the model under study [2]. This is both analytical and synthetic. It is directed at the model as a whole as much as it is at the components of the model. The touchstone of validity is less whether the model is right or wrong: as Sterman states, "... all models are wrong." [28].

Some models, however, fulfill the purpose ascribed to them, i. e., they are useful. Models are inherently incomplete; they cannot claim to be true in an absolute sense, but only to be relatively true [4]. In this sense, *validation* is a *goal-oriented* activity and *validity* a *relative* concept.

Finally, the validation process often involves several people because the necessary knowledge is distributed. In these cases, the dialectics of propositions and refutations, as well as the interaction of different subjective viewpoints, and consensus-building, are integral. Validation processes, then, are semiformal, discursive social procedures with a holistic as opposed to a fragmentary orientation [ibidem].

On Objectivity

If subjective views and judgments are as prominent as alleged above, does objectivity play a role at all? Operational philosophy shows a way out of this dilemma: Rapoport defines objectivity as "invariance with respect to different observers." [21]. Popper has a similar stance in proposing that general statements must be formulated in a way that they can be criticized and, where applicable, falsified [20]. This concept of objectivity is a challenge to model validation: When defining concepts and functions, one must first of all strive for falsifiable statements. In principle, formal models meet this criterion: each variable and every function or relationship can be challenged. And they must be challenged, so that their robustness can be tested. The duty, then, is in finding the invariances that are inter-subjectively accepted as the best approximations to truth. Frequently this is best achieved in group model-building processes [30]. Finally, truth is something we search for but do not possess [20], i. e., even an accepted model cannot guarantee truth with final certainty.

Validation Methods

A considerable set of qualitative and quantitative tests has been developed for the enhancement of model validity. The state-of-the-art has been documented in seminal publications [2,4,7,8,14,17,28]. Our purpose here is to present and exemplify the different tests to encourage and help those who strive to develop high-quality System Dynamics models.

In the following, an overview of the types of tests developed for System Dynamics models is given, without any claim to completeness. Most of these tests have been documented extensively in [2,7,8,28]. The descriptions of the tests adhere closely to the specifications of these authors (mainly Forrester and Senge). In addition, we have developed a new category for tests that concentrate on the con-

text in which the model is to be developed. High-quality models can be created only if the relevant context is taken into consideration. To facilitate orientation, we have attached an overview of all described tests in the Appendix.

In this section we describe three groups of tests: those related to model-related context, tests of model structure and tests of model behavior. Many of the tests described in the following can be utilized for explanatory analysis which aims at an understanding of the problematic behavior of the issue under study. Others are suitable for normative ends, in analyzes targeted on improvements of system performance with regard to a specified objective of the reference system. Also known as policy tests, or policy analyses, these "tests of policy implications differ from other tests in their explicit focus on comparing changes in a model and in the corresponding reality. Policy ... tests attempt to verify that response of a real system to a policy change would correspond to the response predicted by the model" [8]. Policy testing can show the risk involved in adopting the model for policy making.

Tests About the Model-Related Context

These tests deal with aspects related to the situation in which the model is to be developed and embedded. They imply metalevel decisions which have to be taken in the first place, before engaging in model-building. Applied ex-post-facto, i. e., after modeling, they allow for assessing the utility of the modeling endeavor as such.

Issue Identification Test. The *raison d'être* of a System Dynamics model is its ability to adequately address an issue and to enhance stakeholders' understanding, an ability which may lead to policy insights and system improvements. The issue identification test examines whether or not the identified issue or problem is indeed meaningful. Has the "right" problem been identified? Does the problem statement address the origins of an issue or only superficial *symptoms*? Whenever complex issues are addressed by a model, different perspectives (e. g. professional, economic, political) must be integrated for accurate problem identification and modeling. This is not a "one-shot-only" test; it must be applied recurrently during the modeling procedure. By reflecting regularly on the correctness of the identified issue, the modeler can increase the likelihood of capturing the origins of suboptimal system behavior.

Adequacy of Methodology Test. Simulation models respond to the limitations of humans' mental ability to comprehend complex, dynamic feedback systems [27]. The adequacy of methodology test scrutinizes whether the

System Dynamics methodology is best-suited for dealing with the issue under study. One needs to clearly ascertain if that issue is characterized by dynamic complexity, feedback mechanisms, nonlinear interdependency of structural elements and delays between causes and effects. One needs to ask also if the issue under study could be better addressed by another methodology. For example, in a case where the question is to understand the difference in numerical outcomes between two configurations of a production system, it lets one determine whether discrete event simulation would fulfill this requirement more accurately than System Dynamics.

System Configuration Test. This test asks the fundamental question about whether the structural configuration chosen can be accepted. It challenges the assumption that the model represents the actual working of the system under study. The applicability of a different design would be suggested by its ability to capture new conditions, such as different system configurations, phenomena or rules of the game. Even revolutionary changes might be considered. Such an outlook may require a totally new model, or an alternative model designed from a different vantage point. This would at least feasibly approximate the need to take paradigmatic change into account.

System Improvement Test. The purpose of modeling is to understand a part of reality and to resolve an issue. The system improvement test can be performed only after the modeling project (an ex-post-facto test), once the insights derived from the model have already been implemented in the real system. This test reestablishes the connection between the abstract mathematical model and the real system. The system improvement test helps to evaluate whether or not model development was successful. In operational terms, any improvements of the real system under study must be compared with explicit objectives. In practice, the test might assess the impact of the modeling process or the model use either on the mental models of decision makers or on changes in organization structures. In principle, assessing the impact of a modeling endeavor is very difficult (one preliminary example is provided by Snabe and Grössler [25]).

Tests of Model Structure

Tests of model structure refer to the “nuts and bolts” of System Dynamics modeling, i.e., to the formal concepts and interrelationships which represent the real system. Model structure tests aim to increase confidence in the structure of the created theory about the behavior mode of

interest. The model structure can be assessed by means of either direct or indirect inspection. Tests of model structure assess whether the logic of the model is attuned to the corresponding structure in the real world. They do not yet compare the model behavior with time series data from the real system.

Direct Structure Tests Direct structure tests assess whether or not the model structure conforms to relevant descriptive knowledge about the real system or class of systems under study. By means of direct comparison, they qualitatively assess any disparities between the original system structure and the model structure.

Structure Examination Test. Examination in this case means comparison in the sense just outlined. Qualitative or quantitative information about the real system structure can be obtained either empirically or theoretically. Empirically based tests include reviews of model assumptions about system elements and their interdependencies, e.g., reviews made by highly knowledgeable experts of the real system. Theory-based tests compare the model structure with theoretical knowledge from literature about the type of system being studied. Thereby, a preference for theoretical knowledge specific to the modeled situation over more abstract and general knowledge is usually the case.

To pass the structure examination test, a model must not contradict either the evidence or knowledge about the structure of the real system. This test ensures that the model contains only those structural elements and interconnections that are most likely extant in the real system. In this context, formal inspections of the model's equations, reviews of the syntax for the stock and flow diagram, and walkthroughs along the causal loop diagrams and their embodied causal explanations may be indicated. The experienced reader might recommend the use of statistical tests to identify and validate model structure. As Forrester and Senge [8] indicate, a long-standing discussion exists about the application of inferential statistical tests for structure examination. After a series of experiments, Forrester and Senge conclude “that conventional statistical tests of model structure are not sufficient grounds for rejecting the causal hypotheses in a system dynamics model.” [8]. In the future, however, new statistical approaches might enrich the testing procedures.

Parameter Examination Test. A parameter is a quantity that characterizes a system and is held constant in a case under study, but may be varied in different cases (e.g., energy consumption per capita per day). The aim of parameter examination is to evaluate a model's parameters

against evidence or knowledge about the real system. The test can utilize both empirical and theoretical information. Furthermore, the test can be conceptual or numerical. The conceptual parameter examination test is about construct validity; it identifies elements in the real system that correspond to the parameters of the model. Conceptual correspondence means that the parameters match elements of the real system's structure. Numerical parameter examination checks to see if the quantities of the conceptually confirmed parameters are estimated accurately. Techniques for the estimation of parameters are described in [9].

Direct Extreme Condition Test. Extreme conditions do not often occur in reality; they are exceptions. The validity of a model's equations under extreme conditions is evaluated by assessing the plausibility of the results generated by the model equations against the knowledge about what would happen under a similar condition in reality. Direct extreme condition testing is a mental process and does not involve computer simulation. Ideally, it is applied to each equation separately. It consists of assigning extreme values to the input variables of each equation. The values of the output variables are then interpreted in terms of what would happen in the real system under these extreme conditions. For example, if a population is zero, then neither births, deaths, nor consumption of resources can occur.

Boundary Adequacy Structure Test. Boundary adequacy is given if the model contains the relevant structural relationships that are necessary and sufficient to satisfy a model's purpose. Consequently, the boundary adequacy test inquires whether the chosen level of aggregation is appropriate and if the model includes all relevant aspects of structure. It should ensure that the model contains the concepts that are important for addressing the problem endogenously. For instance, if parameters are likely to change over time, they should be endogenized [8]. The pertinent validation question is: "Should this parameter be endogenized or not?" That question must be decided in view of the model's purpose.

The boundary adequacy test can be applied in three ways: as a structural test, as a behavioral test, and as a policy test. The names are correspondingly: boundary adequacy structure test, boundary adequacy behavior test, and boundary adequacy policy test.

As a test of model structure, the boundary adequacy test involves developing a convincing hypothesis relating the proposed model structure to the particular issue addressed by the model. The boundary adequacy behavior/policy test (explained in Subsect. "Indirect Structure Tests") continues this line of thinking.

Dimensional Consistency Test. This test checks the dimensional consistency of measurement units of the expressions on both sides of an equation. The test is performed only at the equation level. When all tests of the individual equations are passed, a large system of dimensionally consistent equations results. This test is passed only if consistency is achieved without the use of parameters that have no meaning in respect to the real world. The dimensional consistency test is a powerful test to establish the internal validity of a model.

Indirect Structure Tests Indirect structure tests assess the validity of the model structure indirectly by examining model-generated outcome behaviors. These tests require computer simulation. The comparative activities in these tests are based on logical plausibility considerations which in turn are based on the mental models of the analyst. Comparisons of model generated data and time series about the real system are not yet involved. The tests can be applied to different degrees of model completeness, i. e., to the smallest "atomic" model components, to sub-models, as well as to the entire model.

Indirect Extreme Condition Test. For this test, the modeler assigns extreme values to selected model parameters and compares the generated model behavior to the observed or expected behavior of the real system under the same extreme conditions. This test is the logical continuation of the direct extreme condition test, i. e., many of the extreme conditions mentally developed in the previous stage can now be deployed to evaluate the simulated behavioral consequences. This test can be used for the explanatory analysis phase of modeling, but also for the normative phase of policy development. In the first instance, indirect extreme conditions are used to develop a structure that can reproduce the system behavior of interest and guard against developments impossible in reality. In the latter instance, the introduction of policies aims to improve the system's performance. The indirect extreme policy test introduces extreme policies to the model and compares the simulated consequences to what would be the most likely outcome of the real system if the same extreme policies would have been implemented.

Behavior Sensitivity Test. Sensitivity analysis assesses changes of model outcome behavior given a systematic variation of input parameters. This test reveals those parameters to which the model behavior is highly sensitive, and asks if the real system would exhibit a similar sensitivity to changes in the corresponding parameters. "The behavior sensitivity test examines whether or not plausi-

ble shifts in model parameters can cause a model to fail behavior tests previously passed. To the extent that such alternative parameter values are not found, confidence in the model is enhanced." [8]. A model can be numerically sensitive, i. e., the numerical values of variables change significantly, but the behavioral patterns are conserved. It can also exhibit behavioral sensitivity, i. e., the modes of model behavior change remarkably based on systematic parameter variations (Barlas [3] defines several distinct patterns of model behavior).

As the test for indirect extreme conditions, the behavior sensitivity test can also be deployed to assess policy sensitivity. It can reveal the degree of robustness of model behavior and hence indicate to what degree model-based policy recommendations might be influenced by uncertainty in parameter values. If the same policies would be recommended regardless of parameter changes over a plausible range, risk in using the model would be lower than if two plausible sets of parameters lead to distinct policy recommendations.

Integration Error Test. Integration error is the deviation between the analytical solution of differential equations and the numerical solution of difference equations. This test ascertains whether the model behavior is sensitive to changes in either the applied integration method or the chosen integration interval (often referred to as simulation time step). Euler's method is the simplest numerical technique for solving ordinary differential and difference equations. For models that require more precise integration processes, the more elaborated Runge-Kutta integration methods can produce more accurate results, but they require more computational resources.

Boundary Adequacy Behavior Test/Boundary Adequacy Policy Test. The logic for testing boundary adequacy has already been developed under the aspect of direct structure testing in the preceding section. The indirect structure version of this test asks whether model behavior would change significantly if the boundary were extended or reduced; i. e., the test involves conceptualizing additional structure or canceling unnecessary structure with regard to the purpose of the study. As one example of expanding the model boundary, this version of the test allows one to detail the treatment of model assumptions considered as unrealistically simple but still important for the model's purpose. On the other hand, simplifying the model is also a way to reduce the model boundary. The loop-knockout analysis is a useful method to implement this two-sided test. Knockout analysis checks behavior changes induced by the connection and disconnection of a portion of the

model structure, and helps the modeler to evaluate the usefulness of those changes with respect to the model's purpose.

The other version of this test is the boundary adequacy policy test. It examines whether policy recommendations would change significantly if the boundary were extended (or restricted): That is, what would happen if the boundary assumptions were relaxed (or confined)?

Loop Dominance Test. Loop dominance analysis studies the internal mechanisms of a dynamic model and their temporal, relative contribution to the outcome behavior of the model. The relative contribution of a mechanism is a complex quantitative statement that explains the fraction of the analyzed behavior mode caused by the mechanism considered in ► System Dynamics, Analytical Methods for Structural Dominance Analysis in. The analysis reveals the relative strengths of the feedback loops in the model. The loop dominance test compares these results with the modeler's or client's assumption about which are the dominant feedback loops in the real system. Since the results are analytical statements, interpretation and comparison with the real system requires profound knowledge about the system under study.

Loop dominance analysis reveals insights about a model on a different level of analysis than the other validation tests discussed so far: It works not on the level of individual concepts or behaviors of variables but on the level of causal structure, and compares the temporal significance of the different structures to each other. The use of this test for model validation is a novelty. If the relative loop dominances of the model map the relative loop dominances of the real system, confidence in the model is enhanced. If the relative loop dominances of the real system are not known, it is still possible to evaluate whether or not the loop dominance logic in the model is reasonable.

Tests of Model Behavior

Tests of model behavior are empirical and compare simulation outcomes with data from the real system under study. On that basis, inferences about the adequacy of the model can be made. The empirical data can either be historical or refer to reasonable expectations about possible future developments.

Behavior Reproduction Tests The family of behavior reproduction tests examines how well model-generated behavior matches the observed historical behavior of the real system. As a principle, models should be tested against

data not only from periods of stability but also from unstable phases. Policies should not be designed or tested on the premise of normality, but rather should be validated with a view toward robustness and adaptiveness.

Symptom Generation Test. This test indicates whether or not a model produces the symptom of difficulty that motivated the construction of the model. To pass the symptom generation test is a prerequisite for considering policy changes, because “unless one can show how internal policies and structures cause the symptoms, one is in a poor position to alter those causes” [8].

Summary statistics, which measure and enable the interpretation of quantitative deviations, provide the means to operationalize the symptom generation test.

One known example is Theil inequality statistics, which measures the mean square-error (MSE) between the model-generated behavior and the historical time series data. It breaks down the deviation into three sources of error: Bias (U_m), unequal variation (U_s), and unequal co-variation (U_c) [26].

An example taken from Schwaninger and Groesser [22] illustrates the interpretation of the error sources.

This example from an industrial firm concerns the design of a model that replicates the observed, historical product life-cycle pattern with high accuracy (Fig. 2). “Product Revenue” is the main variable of interest and specifies the symptom (growth phase followed by rapid decay). The mean square-error for revenues is 0.35. The individual components of the inequality statistics are: $U_m = 0.01$, $U_s = 0.01$, $U_c = 0.98$. The break down of the statistics shows that the major part of the error is in

the U_c component, while the other two sources of error are small. This signifies that the point-by-point values of the simulated and historical data do not match, even though the model captures the dominant trend and the average values in the historical data. Such a situation indicates that a major part of the error is probably unsystematic, and therefore the model should not be rejected for failing to match the noise component of the data. The residuals of the historic and simulated time series show no significant trend. This strengthens the assessment that the model comprises a structure that captures the fundamental dynamics of the issue under consideration.

Frequency Generation and Phase Relationship Tests. These tests focus on the frequencies of time series and phase relationships between variables. An example is the pattern of investment cycles in an industry. These tests are superior to point-by-point comparisons between model-generated and observed behavior (cf. [7]).

Frequency refers to periodicities of fluctuation in a time series. Phase relationship is the relationship between the time series of at least two variables. In principle, three phase relations are possible: Preceding, simultaneous, and successive. The frequency generation test evaluates whether or not the periodicity of a variable is in accordance with the real system. The phase relationship test assesses the phase shifts of at least two variables by comparing their trajectories.

If the phase shift between the selected simulation variables contradicts the phase shift between the same variables as observed or expected in the real system, a structural flaw in the model might be diagnosed. The test can

System Dynamics Modeling: Validation for Quality Assurance, Figure 2

An example comparison of historical and simulated time series for product revenues. The explained variance is close to 100% ($R^2 = 0.9967$)

uncover failures in the model, but offers only little guidance as to where the erroneous part of the model might be. The autocorrelation function test is one way to operationalize the frequency generation test [1]. The function test consists in comparing the autocorrelation functions of the observed and the model-generated behavior outputs, and can detect if significant errors between them exist.

Modified Behavior Test. Modified behavior can arise from a modified model structure or changes in parameter values. This test concerns changes in the model structure. It can be performed if data about the behavior of a structurally modified version of the real system are available. "The model passes this test if it can generate similar modified behavior, when simulated with structural modifications that reflect the structure of the "modified" real system" [2]. The applicability of this test is rather limited since it requires specific data about the modified real system which must be similar in kind to the original real system. Only under this condition can additional insights into the suitability of the original model structure be obtained. If the modified real system deviates strongly from the original real system, the test does not result in any additional insights, because no stringent conclusions about the validity of the original system can be derived from a model that is dissimilar in its structure.

Multiple Modes Test. A mode is a pattern of observed behavior. The multiple mode test considers whether a model is able to generate more than one mode of observed behavior, for instance, if a model about the production sector of an economy generates distinct patterns of fluctuations for the short-term (production, employment, inventories, and prices) and for the long term (investment, capital stock) [15]. "A model able to generate two distinct periodicities of fluctuation observed in a real system provides the possibility for studying possible interaction of the modes and how policies differentially affect each mode" [8].

Behavior Characteristic Test. Characteristics of a behavior are features of historical data that are clearly distinguishable, e.g., the peculiar shape of an oscillating time series, sharp peaks, long troughs, or such unusual events as an oil crisis. Since System Dynamics modeling is not about point prediction, the behavior characteristic test evaluates whether or not the model can generate the circumstances and behavior leading to the event. The creation of the exact time of the behavior is not part of the test.

Behavior Anticipation Tests System Dynamics models do not strive to forecast future states of system variables.

Nevertheless, given that the fundamental system structure is not subject to rapid and fundamental change, dynamic models might provide insights about the possible range of future behaviors. Hence, behavior anticipation tests are similar to behavior reproduction tests but possess a higher level of uncertainty.

Pattern Anticipation Test. This test examines whether a model generates patterns of future behavior which are assumed to be qualitatively correct. The limits of anticipation reside in the fact that the structure of the system may change over time. The pattern anticipation test entails evaluation of periods, phase relationships, shape, or other characteristics of behavior anticipated by the model. One possibility for implementing this test is to split the historical time series into two data sets and introduce an artificial present time at the end of the first data series. The first set is then used for model development and calibration. The second data series is employed to perform the behavior anticipation test, i.e., to evaluate whether the model is able to anticipate possible future behavior.

This test can also be used for policy considerations, in which case it is called "Changed Behavior Anticipation Test". It determines whether the model correctly anticipates how the behavior of the real system will change if a governing policy is altered.

Event Anticipation Test. In respect to System Dynamics, the anticipation of events does not imply knowing the exact time at which the events occur; it rather means understanding the dynamic nature of events and being able to identify the antecedents leading to them. For instance, the event anticipation test is passed if a model has the ability to anticipate a steep peak in food prices based on the development of the conditioning factors.

Behavior Anomaly Test In constructing and analyzing a System Dynamics model, one strives to make it behave like the real system under study. However, the analyst may detect anomalous features of the model's behavior which conflict with the behavior of the real system. Once the behavioral anomaly is traced to components of the model structure responsible for the anomaly, one often finds flaws in model assumptions. The test for recognizing behavioral anomalies is sporadically applied throughout the modeling process.

Family Member Test A System Dynamics model often represents a family of social systems. Whenever possible, a model should be a general representation of the class of that system to which the particular case belongs. One

should ask if the model can generate the behavior in other instances of the same class. "The family-member test permits a repeat of the other tests of the model in the context of different special cases that fall within the general theory covered by the model. The general theory is embodied in the structure of the model. The special cases are embodied in the parameters. To perform this test, one uses the particular member of the general family for picking parameter values. Then one examines the newly parametrized model in terms of the various model tests to see if the model has withstood transplantation to the special case" [8]. The model should be calibrated so as to be applicable to the widest range of related systems. For the family member test, only the parameter values of the model are subject to alterations; changes in the model structure are part of the modified behavior test, as discussed in the preceding section.

Surprise Behavior Test A surprising model behavior is a behavior that is not expected by the analysts. When such an unexpected behavior appears, the model analysts must first understand the causes of the unexpected behavior within the model. They then compare the behavior and its causes with those of the real system. In many cases, the surprising behavior turns out to be due to a formulation flaw in the model. However, if this procedure leads to the identification of behavior previously unrecognized in the real system, the confidence in the model's usefulness is strongly enhanced. Such a situation may signify a model-based identification of a counter-intuitive behavior in a social system.

Turing Test The Turing test is a qualitative test which uses the intuitive knowledge of system experts to evaluate model behavior. Experts are presented with a shuffled collection of real and simulated output behavior patterns. They are asked if they can distinguish between these two types of patterns. If they are unable to discern which pattern belongs to the real system and which to the simulation output, the Turing test is passed. Similar to the phase relationship test, the Turing test is powerful in its ability to indicate structural flaws, but offers only little guidance for locating them in the model.

Validation Process

The validation process pervades all phases of model-building and reaches even beyond, into the phases of model implementation and use. The diagram in Fig. 3 visualizes the function of validation in the process of model-building.

System Dynamics Modeling: Validation for Quality Assurance, Figure 3

Validation in the context of the System Dynamics modeling procedure

For the purposes of this contribution, validation is placed at the center of the scheme. From there it is dispersed through all steps of the modeling process, Map (high-level model creation), Model (build the formal model), Simulate (explore scenarios, etc.) and Design (articulation of policies). We have limited the differentiation of these steps in order to highlight the structure of the process – a recursive structure drawn as a nested loop line. After the initial identification of issues and the articulation of model purpose, the simplified diagram denotes the four phases, of mapping to modeling to simulation and design. The small loops symbolize micro-processes in which, for example, a model is submitted to validation, e. g., a direct structure test, which may lead to its modification (two small arrows). The larger loops illustrate more comprehensive processes. For example, an indirect structure test of the model is carried out, in which the behavior is tested by means of simulation. Or a policy test by simulation leads to implications for design (large loop), and the design is validated in detail thereafter (small loop).

Now, we should note that the process scheme reminds us of a further aspect which is quite fundamental. If the results of the model's operation, e. g., a "prediction", diverge from the results of a test, then either the model is wrong or the test is inadequate (see p. 168 in [24]). This meta-perspective lets us keep an eye on the adequacy of the tests:

is the logic of the test flawless? Are the data sources in order? (see adequacy of methodology test in Sect. "Validation Methods").

Model-building is a process of knowledge-creation, and model validation is an integral part of it. As the model is validated using the methods described in the former chapter, insights emerge, and a better understanding of the system under study keeps growing. But model-building is also a construction of a reality in the minds of observers [31,32] concerned with an issue. In this procedure, validation is supposed to be a "guarantor" for the realism of the model, a control function for preventing gross aberrations in individual and collective perceptions. Validation should encompass precautions against cognitive limitations and modeler blindness. The set of tests presented above is a system of heuristic devices for enhancing such provisions. A question not yet answered is how these tests should be ordered along the timeline. We have fleshed out three structural principles, which are illustrated in Fig. 4:

1. *Validation is a parallel process:* Validation in all three domains – context, structure and behavior – is carried out in a synchronized fashion, as shown in Fig. 4. Context validation is continuous, while the other two components show alternations.
2. *Parts of the validation process have a sequential structure:* This refers to the alternations between the components of structure and behavior validation. In principle, they occur alternately, with structural validation taking the lead and behavior validation following. After that,

one might revert to structural validation again and so forth.

3. *Validation processes are polyrhythmic:* The length and accentuation of validation activities vary among the three levels. This fact is symbolized by the frequency of the vertical lines in the blocks of the chronogram.

A further important factor affecting the validation process is the degree of resolution: micro, meso or macro (as visualized in Fig. 1). The focus of validation is primarily on micro-objects, the smallest building blocks of a model, for example, a stock or a subsystem containing a stock with its flows. One could call them metaphorically *atoms* or *molecules*. Each building-block should be validated individually, before it is integrated into the overall model structure. The reason is that at this atomic level disfunctionalities or errors of thinking are discovered immediately, while at higher levels of resolution the identification of structural flaws is more difficult and cumbersome. The same holds for the relation between modules (meso) and the whole model (macro). Before adding a module, it should be validated in itself. This way, errors at the level of the whole system can be minimized and, it is very important to add, counterintuitive behavior of the model can be understood with more ease.

Until now we have examined what occurs in a validation process and how the process is structured. Finally, we raise the issue of who the actors are and why. In this context, we will concentrate on group processes in model validation.

System Dynamics Modeling: Validation for Quality Assurance, Figure 4
The interplay of validation activities

Different observers associate diverse contents with a system, and they might even conceive the system distinctly, as far as its boundaries, goals and structures are concerned. They might also succumb to erroneous inferences and therefore adhere to defective propositions. Consequently, error-correcting devices are needed. A powerful mechanism for this purpose is the practice of model-building and validation in groups. We have already referred to that concept in respect to several of the methods discussed in Sect. "Validation Methods", and now we will briefly expand on it.

Group Model-building (GMB) is a methodology to facilitate team learning with the help of System Dynamics [30]. The methodology consists of a set of methods and instruments as well as heuristic principles. These are meant to facilitate the elicitation of knowledge, the negotiation of meanings, the creation of a shared understanding of a problem in a team, as well as the joint construction and validation of models. The process of GMB is essentially a dialog in which different interpretations of the real system under study are exposed, transformed, aligned and translated into the concepts and relationships which make up the model system. This is mainly a matter of structural validation, of qualitative mapping and the elaboration of the formal model.

Given its transdisciplinary approach, GMB enables an integration of different perspectives into one shared image of the system-in-focus. GMB is an important provision for attaining higher model quality: it can broaden the available knowledge base, inhibit errors and show itself to be a cohesive force in the quest for consensual model validation. The opportunity for validation inheres in the broad knowledge base normally available in a modeling group. Much of this knowledge can be leveraged for validation purposes. Most validation tests are carried out in coordination with model-building activities. Often the tests become a task to be accomplished between workshops. However, the members of the model-building group can, in principle, be made available for knowledge input into and monitoring of validation activities.

A functioning GMB process requires a number of necessary elements [18]: commitment of key players (e.g., attendance of workshops), impartial facilitation, on-the-spot modeling at conversational pace, with continuous display of the developing model as well as an interactive and iterative group process.

Let us not forget that there are many situations in which one single person is in charge of building and validating a model. In these cases the modeler must constantly challenge his or her own position. Normally, it is preferred that one should also call for external judgment in reviews,

walkthroughs and the like. The same holds for knowledge supply. One-person modelers can find a lot of material in the media, libraries, the internet, etc., but it is also usually beneficial to find experienced persons from whom to elicit relevant knowledge, or even persons who join the modeling and validation venture.

Synopsis and Outlook

Models should be relevant for coping with the complexity of the real world. At the same time, the methods by which they are constructed must be rigorous; otherwise the quality of the model suffers. Rigor and relevance are not entirely dichotomous, but given resource constraints they are in competition to a certain extent. Lack of rigor in building a model is often worse than limitations to the model's relevance. One may say, *cum grano salis*: incomplete validation entails complete irrelevance. Modelers must find a way to ensure both rigor and relevance, as both are necessary conditions for achieving the model purpose. Neither alone is sufficient, but one may assume that, taken together, rigor and relevance are sufficient conditions. The relative importance of these two dimensions of model building may vary over time as a function of the model quality achieved. At the beginning, relevance might be more important, while at high levels of model accomplishment rigor might become prevalent.

Investing in high model quality is indeed both worthwhile and imperative. It is impressive to register the fact that model validation has achieved higher levels of rigor not only in the academic field but also in the world of affairs: According to Coyle and Exelby, the need for orientating decisions about "real-world" affairs has also fueled strong efforts among commercial modelers and consultants for ensuring model validity [5].

We have discussed two essential aspects of model validation, the epistemological foundations and methodological procedures for ensuring model validity. The main conclusion we have reached on epistemology is that crude positivism has been superseded by newer philosophical orientations that provide guidance for an adequate concept of validation in System Dynamics. Validation has been defined as a rich and well-defined process by which the confidence in a model is gradually enhanced. Validity, then, is always a matter of degree, never an absolute property.

Well-defined here is not meant in the sense of a rigid algorithm, but as the rigorous application of a battery of validation methods which we have described in some detail. We have included a number of new validation tests by which modelers' understanding of the relevant con-

text can be scrutinized. These additional tests are rightly supposed to prevent wrong methodological choices. They should also trigger innovative approaches to the issues under study and foster the ability to think in terms of contingencies. Finally, they should liberate modelers from tunnel vision and open avenues to creativity. The imperative here is to cultivate a "sense of the possible" (Robert Musil's *Möglichkeitssinn*) and a skepticism against the supposedly impossible (see also [29]).

Simulation based on formal dynamic models is likely to become ever more important for both private and public organizations. It will continue to support managers at all levels in decision-making and policy design. The more that models are relied upon, the greater the importance of their high quality. Therefore, model validation is one of the big issues lying ahead in System Dynamics modeling.

Appendix: Overview of the Tests Described in This Chapter

1. Tests of the Model-Related Context

- 1.1 Issue Identification Test
- 1.2 Adequacy of Methodology Test
- 1.3 System Configuration Test
- 1.4 System Improvement Test

2. Tests of Model Structure

- 2.1 Direct Structure Tests
 - 2.1.1 Structure Examination Test
 - 2.1.2 Parameter Examination Test
 - 2.1.3 Direct Extreme Condition Test
 - 2.1.4 Boundary Adequacy Structure Test
 - 2.1.5 Dimensional Consistency Test
- 2.2 Indirect Structure Tests
 - 2.2.1 Indirect Extreme Condition Test
 - 2.2.2 Behavior Sensitivity Test
 - 2.2.3 Integration Error Test
 - 2.2.4 Boundary Adequacy Behavior Test/Boundary Adequacy Policy Test
 - 2.2.5 Loop Dominance Test

3. Tests of Model Behavior

- 3.1 Behavior Reproduction Tests
 - 3.1.1 Symptom Generation Test
 - 3.1.2 Frequency Generation and Phase Relationship Test
 - 3.1.3 Modified Behavior Test
 - 3.1.4 Multiple Modes Test
 - 3.1.5 Behavior Characteristic Test
- 3.2 Behavior Anticipation Tests
 - 3.2.1 Pattern Anticipation Test
 - 3.2.2 Event Anticipation Test
- 3.3 Behavior Anomaly Test

3.4 Family Member Test

3.5 Surprise Behavior Test

3.6 Turing Test

Bibliography

Primary Literature

1. Barlas Y (1990) An autocorrelation function test for output validation. *Simulation* 55(1):7–16
2. Barlas Y (1996) Formal aspects of model validity and validation in system dynamics. *Syst Dyn Rev* 12(3):183–210
3. Barlas Y (2006) Model validity and testing in System Dynamics: Two specific tools. Paper presented at the 24th International Conference of the System Dynamics Society, Nijmegen
4. Barlas Y, Carpenter S (1990) Philosophical roots of model validity – two paradigms. *Syst Dyn Rev* 6(2):148–166
5. Coyle G, Exelby D (2000) The validation of commercial system dynamics models. *Syst Dyn Rev* 16(1):27–41
6. Feyerabend P (1993) *Against method*, 3rd edn. Verso, London
7. Forrester JW (1961) *Industrial dynamics*. MIT Press, Cambridge
8. Forrester JW, Senge PM (1980) Test for building confidence in System Dynamics models. In: Legasto AA Jr, Forrester JW, Lyneis JM (eds) *System Dynamics*. North-Holland Publishing Company, Amsterdam, pp 209–228
9. Graham AK (1980) Parameter estimation in system dynamics modeling. *TIMS Studies in the Management Sciences* 14:125–142
10. Heracleous L (2006) *Discourse, interpretation, organization*. Cambridge University Press, Cambridge
11. James W (1987) *Writings 1902–1910*. Library of America, New York
12. Kuhn T (1996) *The structure of scientific revolutions*, 3rd edn. University of Chicago Press, Chicago
13. Lacey AR (1996) *A dictionary of philosophy*, 3rd revised edn. Barnes and Noble, New York
14. Lane DC (1995) The folding star: A comparative reframing and extension of validity concepts in System Dynamics. In: Simada T, Saeed K (eds) *Proceedings of 1995 international System Dynamics conference*, 30 July–4 Aug, vol I. System Dynamics Society, Lincoln, pp 111–130
15. Mass NJ (1975) *Economic cycles: An analysis of underlying causes*. Productivity Press, Cambridge
16. Mattheij RMM, Rienstra SW, Boonkkamp JH MtT (2005) *Partial differential equations: Modeling, analysis, computation*. Society for Industrial and Applied Mathematics (SIAM), Eindhoven
17. Petersen DW, Eberlein RL (1994) Understanding models with vensim. In: Morecroft JDW, Sterman JD (eds) *Modeling for learning organizations*. Productivity Press, Portland, pp 339–358
18. Phillips LD (2007) Decision conferencing. In: Edwards W, Miles RF, von Winterfeldt D (eds) *Advances in decision analysis. From foundations to applications*. Cambridge University Press, Cambridge, pp 375–399
19. Popper KR (1959) *The logic of scientific discovery*. Basic Books, New York (latest edition: 2002, Routledge, London)
20. Popper KR (1972) *Objective knowledge: An evolutionary approach*. Clarendon Press, Oxford
21. Rapoport A (1954) *Operational philosophy. Integrating knowledge and action*. Harper, New York

22. Schwaninger M, Groesser SN (2008) Model-based theory-building with system dynamics. *Syst Res Behav Sci* 25:1–19
23. Seiffert H, Radnitzky G (1994) *Handlexikon der Wissenschaftstheorie*, 2nd edn. DTV Wissenschaft, Munich
24. Smith VL (2008) *Rationality in economics: Constructivist and ecological forms*. Cambridge University Press, Cambridge
25. Snabe B, Grössler A (2006) System dynamics modelling for strategy implementation – case study and issues. *Syst Res Behav Sci* 23(4):467–481
26. Sterman JD (1984) Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica* 10(2):51–66
27. Sterman JD (1989) Misperceptions of Feedback in Dynamic Decision Making. *Organ Behav Human Decis Process* 43(3):301–335
28. Sterman JD (2000) *Business dynamics. Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill, Boston
29. Taleb NN (2007) *The black swan. The impact of the highly improbable*. Random House, New York
30. Vennix JAM (1996) *Group model building: Facilitating team learning using System Dynamics*. Wiley, Chichester
31. von Foerster H (1984) *Observing systems*, 2nd edn. Intersystems Publications, Seaside
32. von Glasersfeld E (1991) Abschied von der Objektivität. In: Watzlawick P, Krieg P (eds) *Das Auge des Betrachters*. Piper, Munich, pp 17–30

Books and Reviews

- Finlay PN (1997) Validity of decision support systems: Towards a validation methodology. *Syst Res Behav Sci* 14(3):169–182
- Forrester JW (1961) *Industrial dynamics*. MIT Press, Cambridge
- Law AM (2007) *Simulation modeling and analysis*, 4th edn. McGraw-Hill, New York
- Legasto AA Jr, Forrester JW, Lyneis JM (eds) (1980) *System Dynamics*. North-Holland, Amsterdam
- Morecroft J (2007) *Strategic modelling and business dynamics: A feedback systems approach*. Wiley, Chichester
- Sargent RG (2004) Validation and verification of simulation models. In: Ingalls RG, Rossetti MD, Smith JS, Peters BA (eds) *Proceedings of the 2004 winter simulation conference*. ACM-Association for Computing Machinery, Washington DC, pp 17–28
- Sterman JD (2001) *Business dynamics. Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill, Boston
- Warren K (2008) *Strategic management dynamics*. Wiley, Chichester

System Dynamics Models of Environment, Energy and Climate Change

ANDREW FORD

School of Earth and Environmental Sciences,
Washington State University, Pullman, Washington, USA

Article Outline

Glossary

Definition of the Subject

Introduction

The Model of Mono Lake

The Model of the Salmon in the Tucannon River

Models of Climate Change

System Dynamics Models of the Carbon Cycle

Lessons from the Regulated Power Industry
in the 1970s

Simulating the Power Industry Response
to a Carbon Market

Conditions for Effective Interdisciplinary Modeling

Future Directions

Bibliography

Glossary

CO₂ Carbon dioxide is the predominant greenhouse gas. Anthropogenic CO₂ emissions are created largely by the combustion of fossil fuels.

CGCM Coupled general circulation model, a climate model which combines the atmospheric and oceanic systems.

GCM General circulation model, a term commonly used to describe climate models maintained at large research centers.

GHG GHG is a greenhouse gas such as CO₂ and methane. These gases contribute to global warming by capturing some of the outgoing infrared radiation before it leaves the atmosphere.

GT Gigaton, a common measure of carbon storage in the global carbon cycle. A GT is a billion metric tons.

IPCC The Intergovernmental Panel on Climate Change was formed in 1988 by the World Meteorological Organization and the United Nations Environmental Program. It reports research on climate change. Their assessments are closely watched because of the requirement for unanimous approval by all participating delegates.

Definition of the Subject

System dynamics is a methodology for studying and managing complex systems which change over time. The method uses computer modeling to focus our attention on the information feedback loops that give rise to the dynamic behavior. Computer simulation is particularly useful when it helps us understand the impact of time delays and nonlinearities in the system. A variety of modeling methods can aid the manager of complex systems. Coyle (p. 2 in [3]) puts the system dynamics approach in perspective when he describes it as that “branch of control theory which deals with socio-economic systems, and that