STRUCTURAL VALIDATION OF SYSTEM DYNAMICS AND AGENT-BASED SIMULATION MODELS

Hassan Qudrat-Ullah
York University
School of Administrative Studies
4700 Keele St. Toronto, ON, M3J 1P3, Canada
Email: hassanq@yorku.ca

KEYWORDS
Simulation; System Dynamics; Structural Validity

ABSTRACT

Simulation models are becoming increasingly popular in the analysis of important policy issues including global warming, population dynamics, energy systems, and urban planning. The usefulness of these models is predicated on their ability to link observable patterns of behavior of a system to micro-level structures. This paper argues that structural validity of a simulation model -right behavior for the right reasons- is a stringent measure to build confidence in a simulation model regardless of how well the model passes behavior validity tests. That leads to an outline of formal structural validity procedures available but less explored in system dynamics modeling 'repertoire'. An illustration of a set of six tests for structural validity of both system dynamics and agent-based simulation models follows. Finally, some conclusions on the increased appeal for simulation models for policy analysis and design are presented.

1. INTRODUCTION

For the remaining pages follow the general guidelines below: Models have been developed and applied to both operational problems as well as policy issues. However, the need of and the evaluation criteria of model validation differs for each case. For instance, in the case of operational problems, the results of a model can be accepted or rejected by exposing the results to a face validity test (Hermann 1967; Emshoff and Sission 1970). In a face validity test, experts assess how the model and its results are close to the real system. Model solutions can be tested in real world environments: e.g., another service window can be opened in the bank; efficiency of the oil refinery can be enhanced under the recommended actions; or inventory control system can be used improve customer satisfaction (Gass 1983).

In contrast, the majority of policy models such as system dynamics (SD) type model and agent-based models are built for the analysis of policy, exploration of possible future scenarios, and management purposes (Gass 1983; Sterman 1984; Oliva 2003; Scholl 2001). From policy research perspective, modeling resolutions to important issues including global warming, population dynamics, energy systems, and urban planning simply defy a face validity test. Instead, for policy models, the key issue in validation is deciding (i) if the model is acceptable for its intended use, i.e., does the model mimic the real world well enough for its stated purpose (Forrester 1971; Goodall 1972; Forrester and Senge 1980) and (ii) how much confidence to place in model-based inferences about the real system (Barlas 1989, 1994; Curry et al. 1989). In order to assess the theoretical content of a policy model, it is imperative to look at the modeling process itself. Therefore, before we could attempt to illustrate the validation for SD models, it is crucial to examine SD modeling process first.

The appeal of SD models in the analysis of policy and managerial issues is due to their ability to link observable patterns of behavior of a system to micro-level structure and decision making processes. In order words, SD models are causal models (Barlas 1989). The crux of SD modeling process is to identify how structure and decision policies help generate the observable patterns of behavior of a system and then identified structures and decision policies be implemented. Therefore, the identification of the appropriate structure is the first step in establishing validity of a SD model. Once the structural validity of a SD model is sufficiently established, behavior validity - how well the model-generated behavior mimics the observed behavior of the real system - is assessed to achieve the overall validity of the model or to build confidence in the model (Gass 1983; Sterman 1984). In fact the validation process becomes iterative: structural validity-behavior validity-structural validity.

Since structural validity involves stakeholders of the model: modelers, clients, and policy researchers, I argue that structural validity is a stringent measure to build confidence in a SD model regardless of how well the model passes behavior validity tests. The second objective of this paper is to illustrate by the way of examples how some of the tests that already exist in the SD validation "repertoire" can help increase confidence in policy models. It is hoped that policy modelers, as a result of our illustrations, will appreciate the usefulness

Proceedings 19th European Conference on Modelling and Simulation
Yuri Merkuryev, Richard Zobel, Eugène Kerckhoffs © ECMS, 2005
ISBN 1-84233-112-4 (Set) / ISBN 1-84233-113-2 (CD)

of already existing but leas explored tests in validation of policy models.

For the discussion of this paper, model refers to a SD type simulation model. However, there exist strong similarities between SD and agent-based modeling approaches: (i) both are unique in modeling nonlinear, complex systems such as urban planning systems, (ii) both assume that micro-structures of a system are responsible for its behavior, and (iii) both aim at discovering leverage points in complex systems, modelers of agent-based models seek them in rules and agents, while SD modelers do so in the feedback structure of a system (Scholl 2001). Therefore, arguments made and the validity procedures illustrated in this article should equally benefit agent-based modeling community. This paper is organized as follows: In § 2, an argument that structural validity is a stringent measure to build confidence in SD type models is established. Structural validity procedures are described in § 3. § 4 provide an illustration of structural validity tests. Conclusions are presented in § 5.

2. STUCTURAL VALIDITY AS A STRINGENT MEASURE FOR A MODEL VALIDATION

In general, validation of SD models draws on two fundamental assumption of SD modeling process: (1) SD models are built to fulfill a purpose, and (2) structure of the model drives its behavior (Forrester 1961). SD modeling process begins with 'conceptualization' of the policy issue and produces a 'quantitative computer simulation model' for policy assessment and design. The purpose of the model informs the construction of both qualitative and quantitative model.

Since its inception, SD has linked the validation of a model with its "purpose". As Forrester emphatically sates that the validity of model should be judged by its suitability for a particular purpose and validity, as an abstract concept divorced from purpose, has no useful use (Forrester 1961). This view of model validation is widely shared by other modelers and policy scientists (Barlas and Carpenter 1990; Holling 1978; Overton 1977). Forrester and Senge (1980) stress that a model is built for a purpose and its validity is determined by the extent to which it satisfies that purpose.

Although SD modeling process is iterative in nature, essence of a SD type model lies in how well the problem has been conceptualized and causal relationships are identified or the qualitative model is constructed. It is the qualitative modeling stage that takes the temporal precedence over the quantitative modeling stage of any SD modeling endeavor: you have to have a conceptual model ready before any effort to realize a computer simulation model could ensue. At the qualitative modeling stage, focus is on (i) having appropriate representation of the problem, and (ii) identifying the causal relationships between the elements of the conceptual model. If problem is either misrepresented or the causal relationships in the model are faulty, model generated data or model's recommendations would simply be misleading. Or in Balras's words, you will get "right behavior for the wrong reasons" Therefore, structural validity: "right behavior for the right reasons" becomes the core of the SD modeling validation process (Barlas 1989).

Moreover, model validation depends on the cultural context and background of the model builders and model users. It depends on whether one is an "observer" (e.g., an academic researcher) or an "operator" (e.g., a decision maker who must act without waiting for data of further analysis (Greenberger et al. 1976). Nevertheless, involvement of stakeholders in the modeling process results in the increased credibility of the model (Kleindorfer at al. 1998). Again it is the conceptual model building stage of SD modeling process where the involvement of stakeholders is prominent: e.g., model assumptions and model boundary: what to model and what not to model is decided based on clients' needs and model builders' approach to modeling. Thus, the conceptual modeling stage allows realize the expertise of the relevant stakeholders and hence increase the likelihood of the acceptance of the model-based recommendations (Coyle and Exelby 2000). Consequently, structural validity that assesses the validity of the conceptual model becomes a stringent measure to build confidence in a SD model. It must be emphasized here that in no way I am discounting the usefulness of behavioral validity of a SD model. Instead, I want to highlight the significance of structural validity, often less explored in SD model validation endeavors.

3. STRUCTURAL VALIDITY PROCEDURES

Identification of the appropriate structure, responsible for the 'right' behavior, is a multidimensional process: problem representation, logical structures, and mathematical and causal relationships. Forrester and Senge (1980) discussed some of the tests used for structural validation of a SD model:

Boundary adequacy: Whether the important concepts and structures for addressing the policy issue are endogenous to the model?

Structure verification: Whether the model structure is consistent with relevant descriptive knowledge of the system being modeled?

Parameter verification: Whether the parameters in the model are consistent with relevant descriptive and numerical knowledge of the system?

Dimensional consistency: Whether each equation in the model dimensionally corresponds to the real system?

Extreme conditions: Whether the model exhibits a logical behavior when selected parameters are assigned extreme values?

Barlas (1989) has demonstrated that behavior sensitivity test, originally suggested by Forrester and Senge (1980) as a behavior validity test, can detect major structural flaws of the model despite the fact that model can generate highly accurate behavior patterns. He termed it as a *structurally-oriented behavior test*: Whether the real system would exhibit a similar high sensitivity to those parameters to which model behavior displays high sensitivity.

4. AN ILLUSTRATION OF STRUCTURAL VALIDITY TESTS

All the tests listed in §3 have been applied to evaluate the structural validity of a system dynamics model MDESRAP: a model for understanding the dynamics of electricity supply, resources and pollution (Qudrat-Ullah and Davidsen 2001). These tests by no means are exhaustive but constitute the core of battery of tests for the structural validity of SD type simulation models. The purpose of the model is to assess the impact of investment incentives on electricity-generating technology mix and emissions level, over the long term (the simulations runs from 1980 to 2030). MDESRAP is a dynamic general disequilibrium representation of Pakistan's electricity supply sector, excluding nuclear generation. An illustration of the applicability of structural validity tests to MDESRAP, one-by-one, follows. Although MDESRAP is not an urban planning model *per se*, structural validity tests being demonstrated here are applicable to any simulation model build to support policy decision making in complex dynamic systems with uncertain data including urban planning systems.

***B*oundary Adequacy**

Consistent with the purpose of MDESRAP, all the major aggregates: electricity demand, investment, capital, resource, production, environment, and costs and pricing are generated endogenously. Only one variable, GDP is exogenous variable. The historical GDP of Pakistan is represented annually from 1980 to 2000 and linear extrapolation is used for the remaining years.

***S*tructure Verification**

The structural verification is of fundamental importance in the overall validation process. For the structural verification of MDESRAP, a two-pronged approach was applied. First, during the construction of the model, we utilized (i) the specific case-Pakistan's data (or available knowledge about the real system), and (ii) the sub-models/ structures of the existing models of the domain, as given in Table 1. The causal relationships developed in the model, which were based on the available knowledge about the real system, provided a sort of 'empirical' structural validation. The adopted sub-models of the existing models of the domain served as a 'theoretical' structural validation (Forrester and Senge, 1980).

Table 1: Adopted Structures in MDESRAP

Structures/ Concepts	Remarks
Investment incentive dynamics (Dyner and Bun, 1997)	Causal structure was adopted
Substitution mechanism between electricity and oil (Davidsen, 1989)	Structural formulation was adopted
Production capital structure (Moxnes, 1990)	Structural formulation was adopted
Gross margin (Sterman, 1980)	Structural formulation was adopted

***P*arameter Verification**

The values assigned to the parameters of MDESRAP are sourced from the existing knowledge and numerical data form case-Pakistan's data. For illustration purpose, Table 2 lists some of the parameters, their values and the source.

Table 2: Some Parameters of MDESRAP and Their Assigned Values

Parameters in the Model	Assigned Values	Source
Time to Adjust Investments	2 (years)	(PEY, 1990; PEY, 1991; PEY, 1997)
Average Physical Life of Capital (oil)	30 (years)	
Average Physical Life of Capital (hydro)	40 (years)	
Target Limit for CO_2 Emission	20.20 M tons	
Construction Delay for Power Plant (oil)	4 (years)	
Construction Delay for Power Plant (hydro)	6 (years)	
Fuel Efficiency	0.4 (%)	
Safety Margin for Resource Inventory	0.5 (year)	
Operating Cost (oil)	0.57 ($/MWh)	
Operating Cost (hydro)	0.22 ($/MWh)	

***D*imensional Consistency**

Dimensional consistency test requires that each mathematical equation in the model be tested if the measurement units of all the variables and constants involved are dimensionally consistent: in (apples) = out (apples). For instance, the following equation represents one of the equations of MDESRAP. This equation describes that share of each competing electricity generating technologies (EnergyTechShare) in the new capital investments being made is dependent on two factors: (i) the coefficient for the distribution of **α** and (ii) the cost of electricity generating technology (CostOfElectTech).

EnergyTechShare = EXP (**-α**) * CostOfElecTech

Is this equation dimensionally consistent? To answer, we need to know (i) Is the value of **α** based on the real system? and (ii) What is the dimension of the dimension of **α**?

The value of **α** is estimated based on the variation in the fuel costs of electricity generation technology, in Pakistan. We considered all 17 locations of thermal power plants, where the fuel is consumed to generate the electricity. The fuel costs at each of these sites were obtained to estimate the value of α = 0.249 (MWh/$). No if we do the dimensional analysis of the equation above, we can have:

[dimensionless]=[(MWh/$)*($/MWh)]=
[dimensionless]

Thus, not only the value of α is based on the existing knowledge of the real system but also the equation is dimensionally consistent.

Both the extreme conditions test and the structurally-oriented behavior test are explaned in detail in Qudrat-Ullah (2004).

In summary, the structure of MDESRAP was exposed to all these tests for overall structural validity. Based on these evaluations, we have strong confidence in MDESRAP's ability to generate "right behavior for right reasons".

Structural Validation of Agent-based Simulation Models

In agent-based modeling, agents are seen as the generators of emergent behavior in a given space (Holland 1999). In Holland's view, the interactions between the agents are nolinear and the overall behavior of the system cannot be obtained by summing the behaviors of the isolated agents. On the other hand, in SD "feedback" structures are seen as intrinsic in real systems and the generators of the aggregate system behavior (Richardson 1992). Thus, both the modeling approaches aim at discovering leverage points in complex aggregate systems, modelers of agent-based models seek them in rules and agents, while SD modelers do so in the feedback structure of a system (Scholl 2001). In Scholl's words, "At the very least, it will be insightful to compare the aggregate behavior and emergent influence on the environment of agent-based models with the predictions of aggregate-level feedback models regarding the same subject area". Therefore, it is prudent to apply structural validation tests illustrated in the previous section on agent-based models. In fact, only after successful structural validation of models, any meaningful comparison could ensue.

5. CONCLUSION

Although structural validity tests constitute but one of two general types of tests required to build confidence in a SD type simulation model, these tests nevertheless are the core of SD modeling validation process and have temporal precedence over the other type of tests: behavior validity tests. Illustrations provided through the applications of six tests in this paper can help the modelers (and users) in policy domain including urban planning to lend an effective and tangible support to the process of building confidence in a simulation model.

Informed by the 'purpose' and structurally tested simulation models, be it SD type or agent-based type, should result in the increased appeal for simulation models for policy analysis and design. The policy issues exist. The simulation models are being built. Validation need and challenges are being met. Policy analysis simulation modeling community owes no apology to those who would only believe in face validity testing alone.

REFERENCES

Barlas, Y. 1989. "Multiple tests for validations of system dynamics type of simulation models." *European Journal of Operational Research* 42 (1), 59-87.

Barlas, Y. 1994. "Model validation in system dynamics". In *Proceedings of the 1994 International System Dynamics Conference*, E. Wolstenholme and C. Monaghan (Eds.). Sterling, Scotland, 1-10.

Barlas, Y. and Carpenter, S. 1990. "Philosophical roots of model validation: Two paradigms." *System Dynamics Review* 6(2), 148-166.

Coyle, G. and Exelby, D. 2000. "The validation of commercial system dynamics models." *System Dynamics Review* 16(1), 27-41.

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

Davidsen, P. 1989. "A petroleum life cycle model for the United States with endogenous technology, exploration, recovery, and demand." Working Paper # 1910-89. MIT, Boston.

Dyner, I. and Bun, D. 1997. "A system simulation platform to support energy policy in Columbia". In *System Modeling for Energy Policy*, D. Bunn, and I. Dyner (Eds.). John Wiley, Chichester, 259-271.

Emshoff, J. R. and Sission, R. L. 1970. "Design and use of computer simulation models". Macmillan, New York.

Forrester, J. W. 1971. "The model versus modeling process." *System Dynamics Review* 1(2), 133-134.

Forrester, J. W. and Senge, P. M. 1980. "Tests for building confidence in system dynamics models." *TIME Studies in the Management* Science 14, 209-228.

Gass, S. I. 1983. "Decision-aiding models: Validation, assessment, and related issues for policy analysis." *Operations Research* 31 (4), 603-631.

Goodall, D. W. 1972. "Building and testing ecosystem models". in *Mathematical Models in Ecology*. J. N. J. Jeffers (Ed.). Blackwell, Oxford, 173-194.

Greenberger, M., Crenson, M. A., and Crissey, B. L. 1976. "Models in the Policy Process". Russell Sage Foundation, New York.

Hermann, C. 1967. "Validation problems in games and simulations." *Behavioral Science* 12, 216-230.

Holland, J. H. 1991. "Emergence from Chaos to Order". Addison-Wesley, Reading, Mass.

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

Kleindorfer, G. B., O'Neill, L., and Ganeshan, R. 1998. "Validation in simulations: Various positions in the philosophy of science." *Management Science* 44 (8), 1087-1099.

Moxnes, E. 1990. "Interfuel substitution in OECD-European electricity production." *System Dynamics Review* 6(1), 44-65.

Oliva, R. 2003. "Model calibration as a testing strategy for system dynamics models." *European Journal of Operational Research* 151, 552-568.

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

PEY, 1990."Pakistan Energy Yearbook 1989". Ministry of Petroleum and Natural Resources, Govt. of Pakistan, Islamabad.

PEY, 1991. "Pakistan Energy Yearbook 1989". Ministry of Petroleum and Natural Resources, Govt. of Pakistan, Islamabad.

PEY, 1997. "Pakistan Energy Yearbook 1989". Ministry of Petroleum and Natural Resources, Govt. of Pakistan, Islamabad.

Qudrat-Ullah, H. and Davidsen, P. 2001. "Understanding the dynamics of electricity supply, resources, and pollution: Pakistan's case." *Energy* 26 (6), 595-606.

Richardson, G. P. 1992. "Feedback Thought in Social Science and System Theory". University of Pennsylvania Press, Philadelphia.

Scholl, J. 2001. "Agent-based and system dynamics modeling: A call for cross study and joint research". In *the proceedings of 34th Annual Hawaii International Conference on System Sciences (HICSS-34)*, Vol. 3, Maui, Hawaii.

Sterman, J. D. 1980. "The use of aggregate production functions in disequilibrium models of energy-economy iterations". Working Paper # D-3234. MIT, Cambridge.

Sterman, J. D. 1984. "Appropriate summary statistics for evaluating the historical fit of system dynamics models." *Dynamica* 10 (Winter), 51-66.

AUTHOR BIOGRAPHY

HASSAN QUDRAT-ULLAH was born in Gujrat, Pakistan and went to the Bahauddin Zakarya University, Multan, where he studied Applied Mathematics and obtained his M. Sc. degree in 1988. He worked for several years for the Chashnupp Power Project before moving in 1996 to the University of Bergen, Norway where he completed his M. Phil. degree in System Dynamics. Hassan then joined in 1999 the NUS Business School, National University of Singapore for his PhD degree in Decision Sciences and graduated in 2002. He did post-doctoral fellowship at Carnegie Mellon University, USA in 2002-2003 before joing York University, Canada as Assistant Professor in Management Science in 2003. His research interests include system dynamics modeling and simulation; dynamic decision making; and computer-simulated interactive learning environments. His e-mail address is: hassanq@yorku. ca.