VALIDATION TESTING: A CASE STUDY

C. Y. Lin
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

ABSTRACT

Validation testing provides the tool for building confidence in a model. It enables an analyst to verify the correctness and usefulness of a model and to gain better insight into, and understanding of, the system being modeled. Although important, validation testing is sometimes difficult to conduct. This paper presents the author's experiences with using the model validation tests to validate a system dynamics model. The paper describes the tests and applications that were most useful in examining the validity of the model, identifies difficulties that can arise during validation testing and offers suggestions for reducing their impact on the process of model validation.

VALIDATION TESTING: A CASE STUDY

I. INTRODUCTION

Building confidence in a model through validation testing is an essential phase of system dynamics modeling. Validation testing enables the analyst to identify and correct errors which arise from, for instance, performing the complicated procedure of interpretating and integrating mental models of a system into an information-feedback model structure, or estimating data for the required input parameters. Intrinsic to the nature of validation testing is another important benefit -- it provides an opportunity for one to gain better insight into, and understanding of, the system being modeled.

Various model validation tests have been developed by experts in the field of system dynamics.[1] Each of these tests attempts to incrementally support the correctness and

1. In their "Tests for Building Confidence In System Dynamics Models" article, J. Forrester and P. Senge described available tests and discussed how the tests can contribute to model validation.[4]

usefulness of a model. Thus, as J. Randers states, "the model becomes 'better' through repetitive testing and correction of weaknesses."[1]

This paper presents the author's experiences with using the various model validation tests to validate a Telecommunications Network (TN) model. The tests that were used to examine the validity of the model, and difficulties that can arise while performing these tests, are discussed in Section II. A model validation procedure, which evolved from validating the TN model, is presented in Section III. Finally, Section IV briefly summarizes the outcome of TN model validation, and presents the author's point of view on the significance of validation testing.

II. VALIDATION TESTING FOR A TELECOMMUNICATION NETWORK MODEL

As with all system dynamics models, two major aspects of the TN model were examined for validity, those of structure and behavior. This section discusses the individual tests that were employed during the TN model structure and behavior verification.

A. The Model Structure Verification

The structure of the TN model was carefully examined prior to simulation runs. During this examination process, the following two objectives were set and achieved. The first objective was to promote user interest in utilizing the model. The second, and most important, was to verify that the model structure properly represented the structure of the system under study.

To best promote the use of the model, and lessen the likelihood of its abandonment, users were encouraged to participate in its development, particularly during the validation phase. The users represent the major authoritative group from which the model builder must draw system structure and obtain approval of later model recommendations. Thus, user involvement not only provides assistance but also promotes enthusiasm in, and approval of, the model itself.

To meet the objective of verifying model structure, the analyst checked for correspondence: (1) between model parameters and system components, and (2) between feedback loops linking model variables and interactions among system components. This comparison of model constructs to the corresponding parts of the real system provides the means for

fundamentally establishing the usefulness of the model, and is further described as follows.

A.1 The Parameter-Representation Test

Every level of management in a system establishes goals and objectives; a pertinent model is one in which parameters are both representative of system components and relevant to the management interest in achieving these goals and objectives. Parameters within the TN model were evaluated with these requirements in mind.

The parameter representation test is useful in evaluating the utility of a model. However, when applying this test, some complications can arise. One major complication is differing opinions among users on the relevance of a given parameter or on the definition of its units. For example, during the evaluation of TN model parameters, there were several opinions about the best measurement to use when presenting quantities of equipment. In this case, the selection of a particular unit influenced the degree of complication associated with data collection.

Usually, users come to an agreement when a parameter is under dispute. Nevertheless, if an agreement cannot be achieved, the model builder should make a personal judgement so as not to delay the progress of the model validation. This judgement can later be re-evaluated through the analysis of model behavior.

A.2 The Feedback-Loop Analysis

Feedback-loops within the TN model were repeatedly analyzed to ensure their proper representation and pertinence to the purpose for which the model was built. Yet, how does feedback-loop analysis really work in examining the validity of a model? As R. Coyle states:

> "This method works by looking at overall structure of the loop pattern, to eliminate any obvious faults, and then analyzing each loop in detail to see how it interacts with other loops."[2]

R. Coyle also indicates that loop analysis, if done poorly, is time-consuming.[3] The author found that to avoid this, and carry out loop analysis more efficiently, the analyst shold bear in mind the following key issues:

- What is the purpose of the model?
- What goals are managers trying to achieve?

- What questions/strategies are users interested in?
- What components influence the essential output variables with which users are concerned?

If the answers to these questions have been clearly defined, the analyst may use them as guidelines while conducting feedback-loop analysis.

An example of feedback-loop analysis is demonstrated in Figure 1 in which a highly simplified causal-loop diagram of one sector within the TN model is illustrated. The diagram depicts the network requirements, generated by the customers, which drive the amount of operational equipment required to provide service. The amount of operational equipment further determines the needed level of maintenance and operations (M&O) personnel. However, the actual level of M&O staffing may not meet the needed level due to, for example, outside perturbations in the labor market. The resulting discrepancy then influences labor effectiveness. Finally, the impact of labor effectiveness on system performance will in turn influence the network requirements growth rate.

The feedback-loop structure described above was revised after applying loop analysis. It was determined that an additional factor influences labor effectiveness -- labor

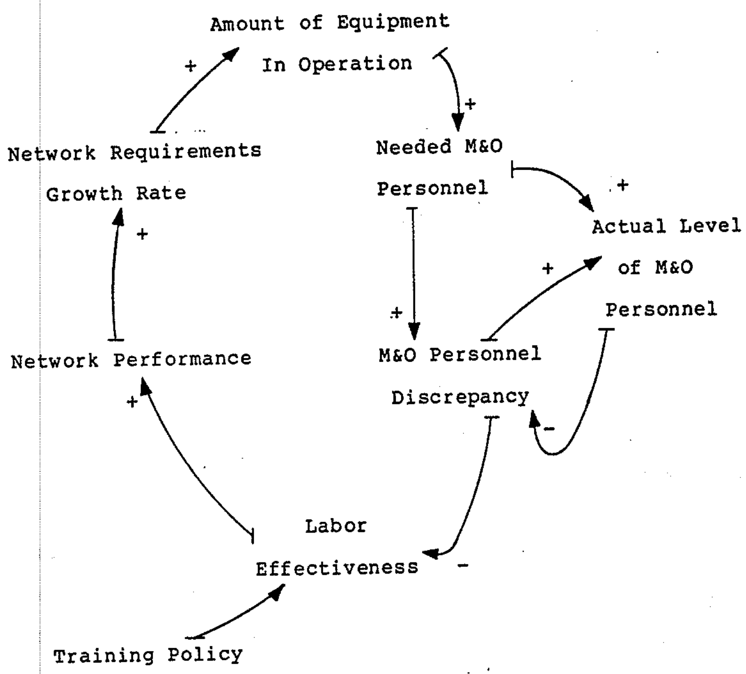

FIGURE 1: Simplified Causal-Loop Diagram of One Sector Within The TN Model

---

Legend: $A \xrightarrow{+} B$: if A increases then B increases

$A \xrightarrow{-} B$: if A increases then B decreases

training. This connection was important because it provided managers with the ability to assess the impact of various training policies on labor effectiveness, which impacts network performance and hence network requirements.

B. The Model Behavior Verification

Model behavior tests are an extension of model structure tests. Behavior tests evaluate adequacy of model structure through analysis of behavior generated by the structure.[4] The model behavior test serves as a tool for identifying faults, if any, in parameter values, mathematical equations, and the model structure itself.

Following are examples of the behavior tests employed for the examination of the TN model behavior. Once again, users play a significant role during this validation phase. They possess, either intuitively or statistically, the knowledge of the dynamic behavior exhibited, or likely to be exhibited, in the real system.

B.1 The Behavior-Anomaly Test

Very often the anomalous behavior produced by a model, especially during the early stages of verification, is due to errors which exist in parameter values or mathematical equations.[2] For instance, the TN model once generated a performance index with values greater than one. However, the range of valid measurements in the real system were between zero and one. Once aware of the abnormality, the analyst quickly found that one of the modifiers for calculating the performance variable needed to be inversed.

The behavior-anomaly analysis is very helpful in identifying the aforementioned type of error; it is often easier to diagnose these problems through behavior runs than it is through code checks. Thus, this analysis is performed so as to eliminate data estimation and/or computational errors. As these errors do not really relate to the model structure itself, this test helps the analyst to avoid the possibility of confusion during further analysis of the model behavior vis-a-vis structure.

---

2. Of course, anomalous behavior may also be caused by improper model structure. However, in this particular study, the TN model structure was very carefully analyzed prior to initiating behavior-anomaly analysis. The possibility of anomalous behavior caused by an inappropriate structure was therefore reduced.

B.2 The Historical-Reproduction And Future-Projection Tests

A system dynamics model is designed to capture the historical dynamic behavior of a system and project its future trends. Tests of historical-reproduction and future-projection verify whether model generated behavior can meet these known or postulated occurences.

Often, much of the recorded historical data needed to evaluate the behavior of model output variables does not exist, and even if data do exist, they are usually recorded in different forms than the analyst would desire. However, the unavailability of accurate data does not impede the validation process. As a substitute, the user's mental data or historical reference modes provides just as useful a criterion for verifying the ability of a system dynamics model to reproduce the system's historical behavior.

Interestingly, if a model is being developed for the long-range planning purpose, usually its ability to project the system's future behavior can be verified with some significant data. This data can be obtained from the planning office. The planning managers usually develop the organization's plan using specific goals, objectives, milestones, etc. These specifications then provide the basic criteria with which to validate the model's predicted behavior.

As an example, some of the Telecommunications Network activities are planned with a twelve year horizon. Based upon these plans, managers have developed formal charts of anticipated network requirement trends. These trends were then used as a basis for overall comparison with the model-generated projections. Furthermore, the system being modeled is implementing a known amount of new equipment with a target completion date of 1986. Again, this event was used as a specific check-point for TN model behavior verification.

B.3 Extreme-Condition Test

The extreme-condition test was also emphasized in validating the TN model. This test asks what would happen if implausible values were assigned to model parameters. Normally, the analyst would already have the expected answers in mind. Thus, when the behavior generated by the TN model under an extreme condition run did not match the analyst's expectations, an investigtion into the model's feedback-loops, as well as parameter assumptions and mathematical equations, was performed.

B.4 The Sensitivity Test

Much literature supports the importance of sensitivity testing by stressing its ability to reveal errors in model structure and/or parameter assumptions, to provide guidance in data collection, and maybe even to identify sensitive areas of the system. Since the subject addressed in this paper is validation testing, the usefulness of sensitivity testing with regard to input data collection will not be discussed.

The behavior-sensitivity test analyzes the sensitivity of model behavior to plausible changes in parameter values. Normally, the behavior of a system dynamics model would be insensitive to these changes. Why? Forrester points out that it is because the information-feedback structure of the real system has a self-correcting adaptability. Thus, if a model properly represents such a system, it should demonstrate the same characteristic of insensitivity.[5]

When the behavior of a model is shown to be sensitive to change, the analyst must re-evaluate the parameter values and the model structure itself. If both of them are proven to be valid after careful examination, the sensitive areas suggested by the model should be brought to the attention of management. Alternatively, the management should also be informed when the model behavior is proven to be insensitive to plausible changes in parameter values.

As an example of the behavior sensitivity test, the levels of equipment production backlogs of the TN model were monitored while the analyst changed the value of labor availability of new hires to lower than normal but within the plausible range. The generated backlog level suprisingly showed only a slight increase over the normal values. This result was justified after an investigation. It is understood that in the real system, if there was a shortage of staffing, the available staff usually would work harder (such as overtime) so as to make up for the shortage problem. Nevertheless, as the shortage persists or becomes more severe, an obvious increase in the backlog level is illustrated and this can affect other parts of the system.

Although important, however, as A. Ford and P. Gardiner[6] point out, sensitivity testing is not at all a trivial analysis. They identify four primary obstacles which cause the sensitivity analysis to be difficult, and propose a new approach which overcomes three of these four barriers. The remaining obstacle is one in which there are two many model parameters that require testing for complete sensitivity analysis of the model.

It is certainly time-consuming to examine the sensitivity of the model behavior to changes in every individual input parameter. The alternative is to select key parameters. However, in so doing, the possibility of omitting an essential model parameter, which would be best for analyzing the sensitivity of model, may exist. Therefore, when selecting the parameters for examining model sensitivity, the analyst should (1) fully understand the system under study; (2) clearly define the answers to those issues mentioned in feedback-loop analysis; and (3) solicit information from users because they may have already recognized or suspected areas which are sensitive to change.

III. A MODEL VALIDATION PROCEDURE

In conducting the above structural and behavioral tests on the TN model, insights into the overall process were developed. A suggested sequence for the model validation procedure was developed and is summarized and presented in Figure 2. Tests of structure and behavior verification are listed in a suggested priority order which evolved from experiences with the TN model. It is believed this procedure can result in less confusion and greater time-effectiveness in the process of model validation.

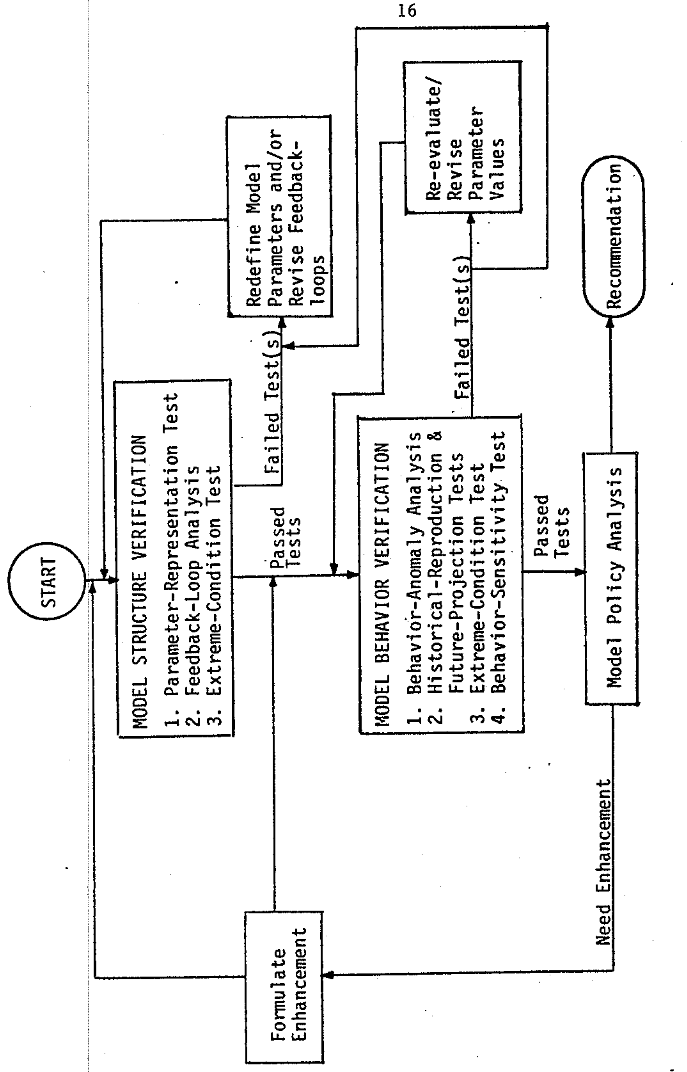

FIGURE 2. A MODEL VALIDATION PROCEDURE

IV. SUMMARY AND CONCLUSIONS

The TN model has successfully passed various structural and behavioral tests. Even though the model is still undergoing policy tests, the users have developed confidence in its utility.

Due to the experiences gained from validating the TN model, the author was able to offer some suggestions in the areas of prioritizing validation tests, and reducing difficulties which can occur during model validation. These suggestions are offered with the intent of supporting an increasingly effective model validation process.

It is clear that the ultimate goal of validation testing is to build confidence in the model, but the process also offers important secondary benefits as well. First, the process of model validation provides the opportunity to look past local concerns and globally explore the system under study. It forces all concerned to examine issues in more depth when differing opinions exist. Second, the validation process may reveal additional management strategies which were not originally known. Both of these secondary benefits occur because the validation process requires open discussion of the actual system as well as of the model itself.

REFERENCES

1. Jorgen Randers, "Guidelines for Model Conceptualization" Jorgen Randers, ed., Elements of the System Dynamics Method (Cambridge, Massachusetts: The MIT Press, 1980), p. 127.

2. R. G. Coyle, Management System Dynamics (London: John Wiley & Sons, 1977), p. 195.

3. Ibid, p. 196.

4. J. W. Forrester and P. M. Senge, "Tests for Building Confidence in System Dynamics Models" (Cambridge, Massachusetts: MIT System Dynamics Group, Alfred P. Sloan School of Management, December, 1978), p. 12.

5. J. W. Forrester, Industrial Dynamics (Cambridge, Massachusetts: The MIT Press, 1964), p. 105.

6. A. Ford and P. Gardiner, "A New Measure of Sensitivity for Social System Simulation Models," IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-9, No. 3, March, 1979.