Model validation in System Dynamics

Yaman Barlas
Bogazici University
Department of Industrial Engineering
80815 Bebek, Istanbul
Turkey
E-mail: Ybarlas@trboun.bitnet; Fax: 90 212 265 1800

Abstract

Model validation constitutes an important step in system dynamics methodology. Validation is a prolonged and complicated process, involving both formal/quantitative tools and informal/qualitative ones. This paper first provides a summary of the philosophical issues involved in model validation. We then focus on the formal aspects of validation, and present a taxonomy of various aspects and steps of formal model validation. We offer a flowchart that describes the logical sequence in which various validation activities must be carried out. We give examples of specific validity tests used in the three major categories of model validation:

Structural tests, structure-oriented behavior tests and behavior pattern tests. Finally, we focus specifically on the logic of behavior pattern validation and illustrate it on a multi-step validation procedure. Currently, we are in the process of implementing this multi-step procedure on micro-computers, embedded in a friendly user-interface.

Model Validation in System Dynamics

Abstract

Model validation constitutes an important step in system dynamics methodology. Validation is a prolonged and complicated process, involving both formal/quantitative tools and informal/qualitative ones. This paper first provides a summary of the philosophical issues involved in model validation. We then focus on the formal aspects of validation, and present a taxonomy of various aspects and steps of formal model validation. We offer a flowchart that describes the logical sequence in which various validation activities must be carried out. We give examples of specific validity tests used in the three major categories of model validation: *Structural tests, structure-oriented behavior tests* and *behavior pattern tests.* Finally, we focus specifically on the logic of behavior pattern validation and illustrate it on a multi-step validation procedure. Currently, we are in the process of implementing this multi-step procedure on micro-computers, embedded in a friendly user-interface.

Introduction

Model validation is an important aspect of any model-based methodology in general, and system dynamics in particular. Validity of the results of a given study are crucially dependent on the validity of the model. Model validation may be defined as: "establishing confidence in the usefulness of a model with respect to its purpose." This "confidence building" process is a gradual one, dispersed throughout the methodology, starting with model conceptualization, and continuing even after implementation of policy recommendations. (See Forrester, 1961 and Forrester and Senge 1980, for excellent treatments of the general nature of model validity). Although model validation does take place in every stage of modeling methodology, it is safe to state that a majority of "formal" validation activities take place right after the model construction has been completed and before policy design simulations. In simulation literature, this set of activities is formally called "validation". And there is a good practical reason for doing so; by focusing on those aspects of a model validation that can be reasonably separated from the rest of the modeling activities, it becomes possible to carry out a structured and rigorous discussion of model validation. Thus, on the one hand we acknowledge that some degree of validation takes place in every step of modeling, and on the other hand, in order to make a rigorous discussion possible, we focus our efforts on those validation activities that can be separated from the rest. This set of validation activities is called "formal" validation in this article.

The article is about the major aspects of formal model validation. It starts with a brief summary of the philosophical aspects of model validity. It then provides a flowchart that describes the general logic of formal model validation. Finally, the article briefly mentions several specific tests that can be used in the various steps of validation.

Philosophical Aspects of Model Validity

In some fundamental ways, the issue of validity of system dynamics models has strong ties with philosophy of science issues. This is due to the fact that system dynamics models claim to be *causal* ones. A system dynamics model is refuted if a critic can show that a model equation conflicts with a known causality, even if the output behavior of the model matches the observed problem behavior. In system dynamics "validity" means validity of the internal structure of the model, not its output behavior. (This principle is known as "right behavior for the right reasons"). It can be said that a valid system dynamics model embodies a *theory* about how a system actually works in some respect. Therefore, there has to be a strong connection between how theories are justified in sciences (a major philosophy of science question) and how system dynamics model are validates. Barlas and Carpenter (1990) discuss this issue in detail. Here is a very brief summary: There are two opposing philosophies of science. The traditional reductionit/logical empiricist philosophy would see a valid model as an objective representation of a real system. The model can be either "correct" or "incorrect". Once the model confronts the empirical facts, its truth or falsehood would be automatically revealed. In this philosophy, validity is seen as a matter of accuracy, rather than usefulness. (Barlas and Carpenter 1990). The more recent relativist/holistic philosophy, in contrast would see a valid model as one of many possible ways of describing a real situation. "No particular representation is superior to others in any absolute sense, although one could prove to be more effective. No model can claim absolute objectivity, for every model carries in it the modeler's worldview. Models are not true or false, but lie on a continuum of usefulness." (Barlas and Carpenter 1990). The authors, by citing major system dynamics articles, show that the philosophy of system dynamics model validations is in agreement with the relativist/holistic philosophy of science. (Forrester 1961, chapter 13; Forrester and Senge 1980). Accordingly, model validation can not be entirely objective, quantitative and formal. Since validity means "usefulness with respect to a purpose," model validation has to have informal, subjective and qualitative components.

This article deals with formal aspects of model validation. But this should not imply that formal/quantitative tests can alone determine the validity of a model. Formal/quantitative tests merely provide inputs to the larger validation process, which is gradual, semi-formal and conversational. We must also note that relativist/holistic philosophy does not reject the role of formal/quantitative tests. On the contrary, since this philosophy claims that validity is gradually established as a result of a "conversational" (rather than confrontational) process, collecting, organizing, interpreting and efficiently communicating information on model validity would play a major role in the process.

Another philosophical issue has to do with the role of "statistical significance" testing in model validation. There are at least three reasons why the concept of statistical significance has little relevance in model validation. One reason is that data generated by system dynamics models are often autocorrelated and crosscorrelated. Traditional significance tests are all based on the assumption of independent data. Therefore, applying statistical tests to correlated data requires extensive model simplification, and/or data transformation, frequently a complex problem in itself, sometimes with no satisfactory solution at all. (See Barlas 1985, Forrester and Senge 1980 and Senge 1977). A second problem, also discussed in Forrester (1961, chapter 13 and 1973) and Barlas (1985), has to do with the common practice of arbitrarily fixing the "significance level" (typically at 0.05) and accepting/rejecting hypotheses, depending on the outcome of the test. In standard tests, the significance level is fixed arbitrarily, not a function of the purpose and nature of a given model. This type of binary accept/reject decision in validity testing is very much against the relativist/holistic philosophy of system dynamics described above. Finally, a third technical problem is that in applying statistical tests to validation, the null hypothesis would be of the form "Xm=Xr" (where Xm represents some measure of the model and Xr corresponds to the same measure of the real system). The problem is that this kind of statistical test is strong only if we reject the null hypothesis (which would mean "the model not valid"). If, on the other hand, we fail to reject the null hypothesis (which is our goal in validation), the result is a very weak one. (That is why, in standard statistical tests, the hypothesis that we try to prove is typically placed in the "alternative" hypothesis). There are many other problematic aspects of the concept of "statistical significance" which are beyond the scope of this article.

In the next section, we discuss the overall logic of formal model validation and certain specific tests. The material should be read with the philosophical perspectives outlined above.

Aspects of Formal Model Validation

Perhaps the most important principle of system dynamics model validation is that the ultimate objective is to establish the validity of the *structure* of the model. Accuracy of the model behavior is also evaluated through certain tests, but this is meaningful only if we already have sufficient confidence in the structure of the model. Thus, the general logical order of validation is, first to test the validity of the structure, and then start testing the behavior accuracy, only if the structure of the model is perceived adequate. This logical sequence is depicted in figure 1.

1- Direct Structure Tests.

Observe in figure 1, that I distinguish between two types of structure tests: 1- *Direct Structure Tests* and 2- *Structure-oriented Behavior Tests*. Direct structure tests assess the validity of the model equations individually, by directly comparing them against available knowledge. There is no simulation involved. These tests can in turn be of two kind: *Empirical or theoretical.* Empirical structure test involves comparing the model equations against knowledge from the real system being modeled. Theoretical structure tests involve comparing model equations against generalized knowledge on the system that exists in the literature. Both type of tests are important in direct structure validation. Forrester and Senge (1980) give examples of direct structure tests, such as, *Structure and parameter verification tests, direct extreme-conditions test* and *dimensional consistency test.* Structure verification test means comparing the structure of the model against the structure of the real system (Forrester and Senge 1980). It may also be carried out as a theoretical structure test, by comparing the model structures against knowledge in the literature. Parameter verification test means evaluating the constant parameters against knowledge of the real system, both conceptually and numerically (Forrester and Senge 1980). Direct extreme condition testing involves evaluating the model equations under extreme conditions and assessing the plausibility of the resulting values against knowledge/anticipation of what would happen under similar condition in real life. Unlike normal operating conditions, it is relatively easy to anticipate how a certain structure of the real system would behave under extreme condition. (Forrester and Senge 1980). Finally, Dimensional consistency test entails dimensional analysis of model equations. To be meaningful, we must require that the model pass the test without including any dummy "scaling" parameters that have no meaning in real life. (Forrester and Senge 1980).

2 - Structure-Oriented Behavior Tests

The second general category of structural tests, Structure-oriented behavior tests, assess the validity of the structure indirectly, by applying certain behavior tests on model-generated behavior patterns. (See Barlas 1989b). These are "strong" behavior tests that can help the modeler uncover potential structural flows. Figure 1 includes four such tests: *Extreme-condition (behavior)* test involves assigning extreme values to selected parameters and comparing the model-generated behavior to the observed (or anticipated) behavior of the real system under the same extreme condition. *Behavior sensitivity* test consists of determining those parameters to which the model is sensitive, and asking if the real system would exhibit similar high sensitivity to the corresponding parameter. *Modified-behavior prediction* can be done if it is possible to find data about the behavior of a modified version of the real system. The model passes this test if it can generate similar modified behavior, when simulated with similar modifications. (See Barlas 1989b and Forrester and Senge 1980 for more details).

3 - Behavior Pattern Tests

The two categories of tests discussed above are designed to evaluate the validity of the model structure. As a result of these tests, once we have built enough confidence in the validity of model structure, we can start applying a number of tests designed to measure how accurately the model can reproduce the major behavior patterns observed in the behavior of the real system. It is crucial to note that the emphasis is on *pattern* prediction (periods, frequencies, trends, phase lags, amplitudes...), rather than point prediction. This is a logical result of long-term policy orientation of system dynamics models. (See Barlas 1985, Forrester and Senge 1980). Among the behavior pattern tests are the multiple-test procedure by Barlas (1985, 1989a), an overall summary statistic proposed by Sterman (1984) and several tests discussed in Forrester and Senge (1980).

Figure 1 summarizes the general nature of the three stages of model validation discussed above. Observe that, in figure 1, all three stages are dependent on "model purpose", which is determined in the problem identification step (the very first step) of system dynamics methodology. No validity test can be carried out in the absolute sense, without reference to the specific purpose of the model.

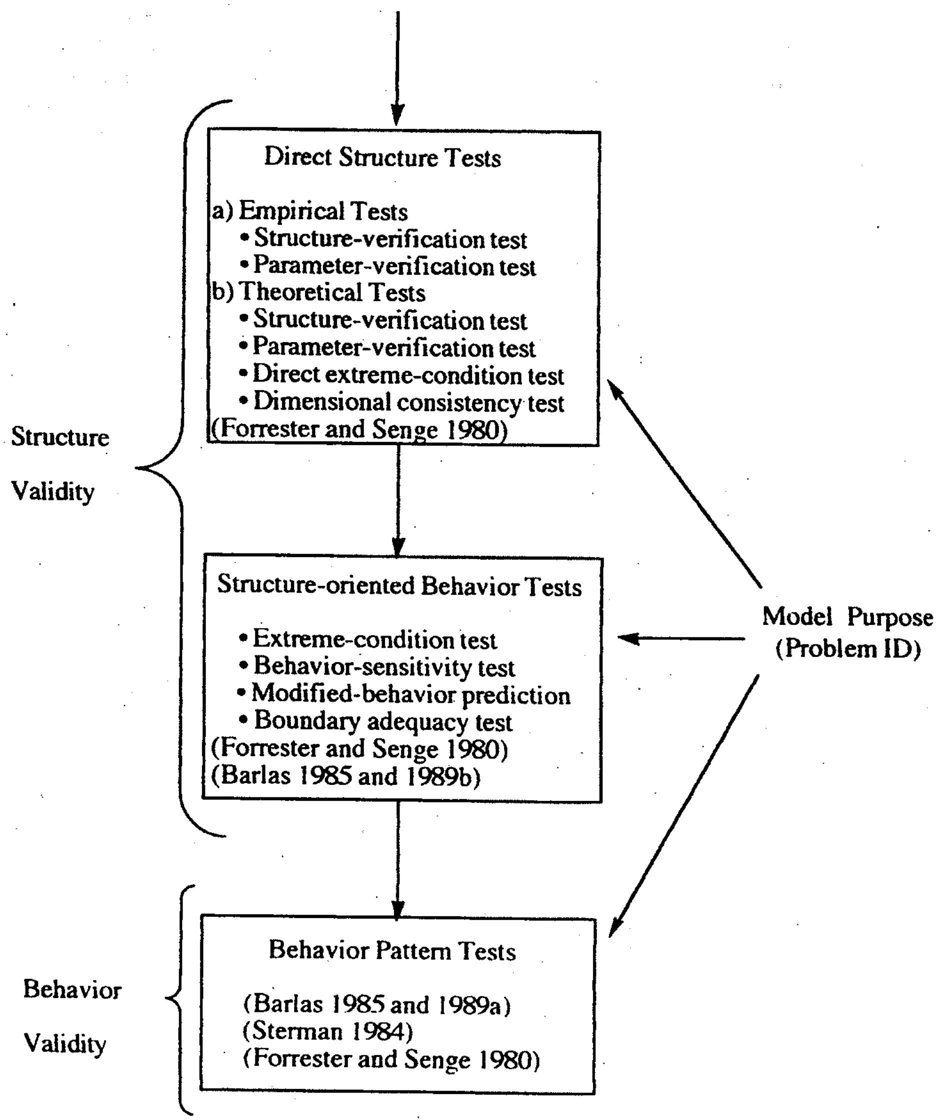

Figure 1. Overall Nature and Selected Tests of Formal Model Validation

Figure 2 outlines the logical sequence of formal steps involved in model validation. The tests are carried out in some logical sequence, and it makes sense to move to the next step, only if we were able to establish sufficient confidence in the current step. Otherwise, we return to work on necessary model revisions (*structural revisions,* not ad-hoc parameter changes) and then come back to the same step. Once the model has been through all the structural tests, we can start assessing the pattern prediction ability of the model by applying a series of behavior tests. In this final step, the emphasis is on pattern prediction "accuracy", and is essentially done for communication and implementation purposes.

Behavior pattern tests must also be carried out in some logical order. Figure 3 illustrates this logic. There are two fundamentally different types of behavior patterns that call for two different types of behavior tests. If the problem involves a transient, highly non-stationary behavior (such as a truncated S-shaped growth, or a single boom-then-bust pattern) then it i s practically impossible to apply any standard statistical tool. The problem is of *no* statistical nature to start with. The best approach in this case is to compare graphical/visual measures of the most typical behavior characteristics (such as the amplitude of a peak, time between two peaks, number of inflection points). There are no general statistical tools that can be offered in this case. If, on the other hand, the problem involves a long-term steady-state simulation, then, it is possible to apply certain standard statistical measures. Figure 3 includes the multi-step behavior validation procedure developed by

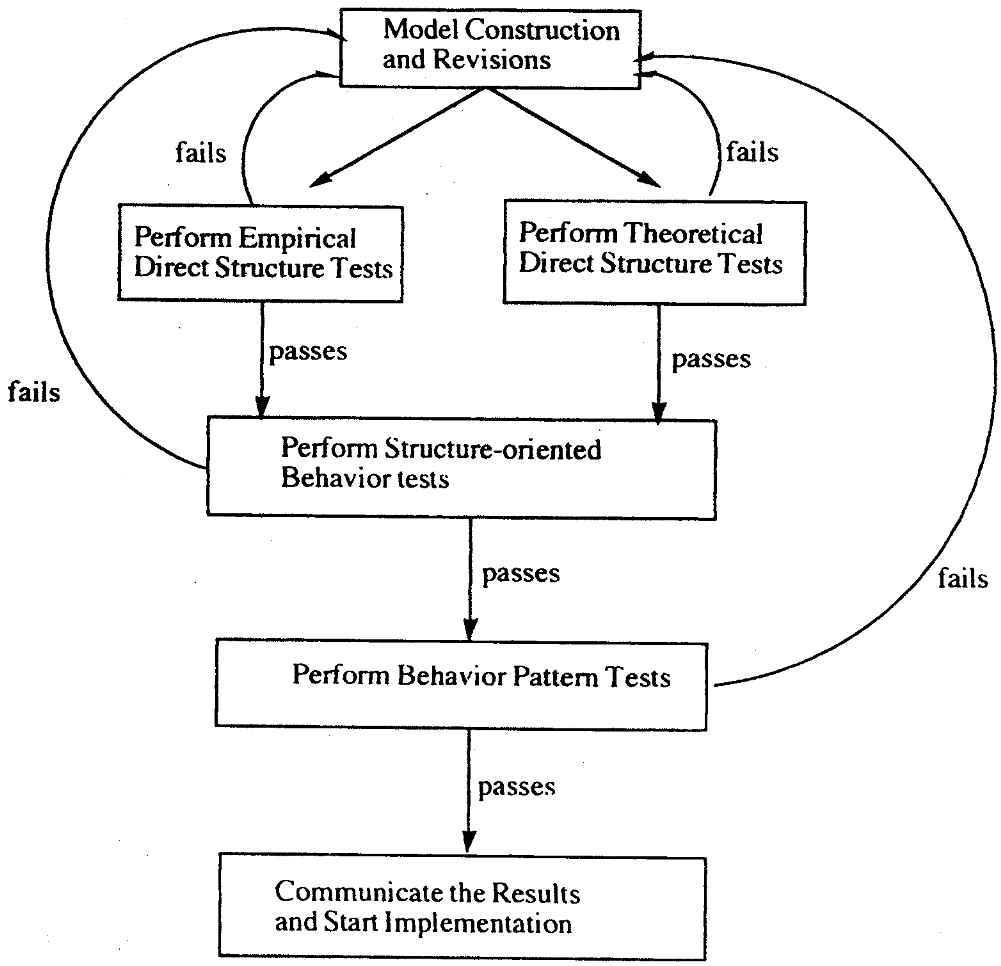

Figure 2. Logical Sequence of Formal Steps of Model Validation

Barlas (1985, 1989a and 1990). Formulas and more detailed information are provided in the appendix. Note that, if a model is judged to fail the behavior tests, we return once again to work on "model revisions". But the difference is that, in this case, since confidence in model structure must have been already established, model revisions involve parameter changes, rather than structural revisions.

Conclusions

Model validation is an important aspect of system dynamics methodology. System dynamics methodology has often been criticized for its lack of formal validation tools. The purpose of this paper is to provide an overview of model validity and to analyze the formal aspects of system dynamics validation. We first summarize the philosophical issues involved in validation. Then, in analyzing the major aspects of system dynamics validation, we provide flowcharts that describe the formal logic involved. We also proved a formal, multi-step , behavior pattern validation procedure. Currently, we are in the process of implementing this multi-step validation procedure a user-friendly computer program on microcomputers. Our future plan is to develop a general computerized "validation environment" following the logic of the flowcharts provided in this article.

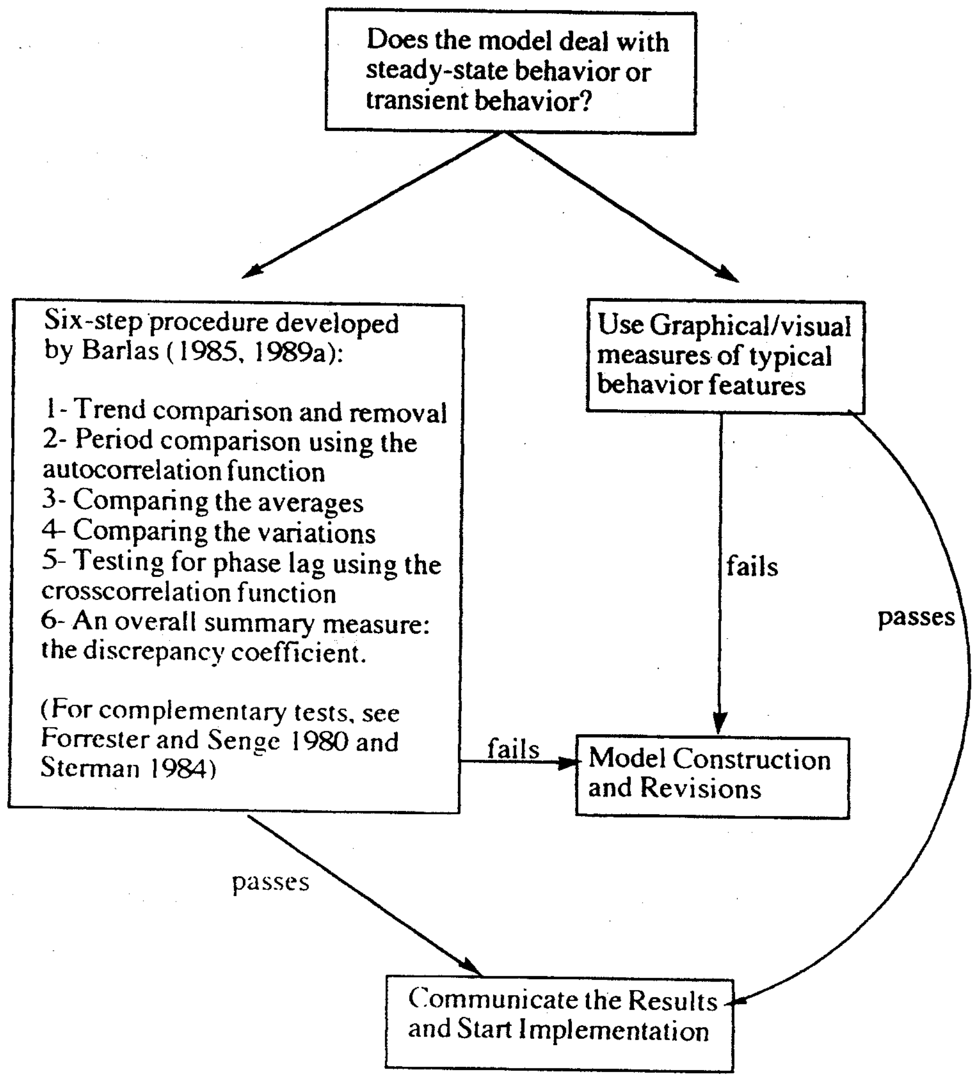

Figure 3. Logical Sequence of Behavior Pattern Validation.

Appendix: The Multi-step Validation Procedure (Barlas 1985, 1989b).

1. Trend comparison and removal : A linear trend can be estimated by $\hat{Y} = b_o + b_1 t$, a quadratic trend by $\hat{Y} = b_o + b_1 t + b_1 t^2$, or an exponential one by $\hat{Y} = b_o e^{bt}$. Then, the trend component can be removed by: $Z_i = Y_i - \hat{Y}_i$

2. Comparing the periods: Autocorrelation function test is able to detect significant errors in the periods. The 'sample autocorrelation function' of a time series X_i is given by

$$Cov(k) = \frac{1}{N}\sum_{i=1}^{N-k}(x_i-\bar{x})(x_{i+k}-\bar{x})$$

for lag k = 0, 1, 2, ...< N. Then, the 'sample autocorrelation function' is obtained by dividing Cov(k) by Cov(0):

$$r(k) = \frac{Cov(k)}{Cov(0)} = \frac{Cov(k)}{Var(X_i)}$$

for lag k = 0, 1, 2, ..., N. We use the following Var(r(k)) provided by O. D. Anderson (1982):

$$Var(r(k)) = \frac{1}{N(N+2)}\sum_{i=1}^{N-1}(N-i)(r(k-i)+r(k+i)-2r(k)r(i))^2$$

If $r_s(k)$ belongs to the simulation output and $r_A(k)$ to the actual (observed) one, then the null hypothesis is :
$H_o : r_s(1) - r_A(1) = 0,\ r_s(2) - r_A(2) = 0, ...,\ r_s(M) - r_A(M) = 0,$
and the alternative hypothesis is,
$H_1 : r_s(k) - r_A(k) \neq 0$ for at least one k.
Now consider the difference $d_1 = r_s(1) - r_A(1)$. The standard error of d_1 is
$Se(d_1) = \sqrt{Var(r_s(1)) + Var(r_A(1))}$.
Since $d_1 = 0$ under Ho, we construct the interval $\{-2Se(d_1), 2Se(d_1)\}$ and reject H_0 if d_1 falls outside the interval.

3. Comparing the means: 'Percent error in the means' E1 can be defined as:

$$E1 = \frac{|\bar{S} - \bar{A}|}{\bar{A}}$$

where,

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} S_i \qquad \bar{A} = \frac{1}{N}\sum_{i=1}^{N} A_i$$

4. Comparing the variations: 'Percent error in the variations' E2 is defined as:

$$E2 = \frac{|s_S - s_A|}{s_A}$$

where,

$$s_S = \sqrt{\frac{1}{N}\sum (S_i - \bar{S})^2} \qquad s_A = \sqrt{\frac{1}{N}\sum (A_i - \bar{A})^2}$$

5. <u>Testing the phase lag</u>: The cross-correlation function provides an estimate of a potential phase lag. The cross-correlation function between the simulated (S) and the actual (A) time patterns is given by :

$$C_{SA}(k) = \frac{(1/N)\sum_{i=k}^{N} (S_i - \overline{S})(A_{i-k} - \overline{A})}{s_S s_A}$$

for k = 0, 1, 2, ...

$$C_{SA}(k) = \frac{(1/N)\sum_{i=k}^{N} (A_i - \overline{A})(S_{i+k} - \overline{S})}{s_S s_A}$$

for k = 0, -1, -2, ...

6. <u>An overall summary measure</u>: As a final step, only after all validity tests have been passed, the discrepancy coefficient U can be computed as a single summary measure :

$$U = \frac{\sqrt{\sum (S_i - \overline{S} - A_i + \overline{A})^2}}{\sqrt{\sum (A_i - \overline{A})^2} + \sqrt{\sum (S_i - \overline{S})^2}}$$

$$= \frac{\sqrt{\sum (E_i - \overline{E})^2}}{\sqrt{\sum (A_i - \overline{A})^2} + \sqrt{\sum (S_i - \overline{S})^2}}$$

$$= \frac{s_E}{s_A + s_S}$$

References

Barlas, Y. and S. Carpenter. 1990. Philosophical Roots of Model Validation: Two Paradigms. *System Dynamics Review* 6(2): 148-166.

Barlas, Y. 1989a. Multiple Tests for Validation of System Dynamics Type of Simulation Models. *European Journal of Operational Research* 42(1): 59-87.

Barlas, Y. 1989a. Tests of Model Behavior That Can Detect Structural Flaws: Demonstration With Simulation Experiments. In *Computer-Based Management of Complex Systems: International System Dynamics Conference*. Ed. P.M. Milling and E.O.K. Zahn. Berlin: Springer-Verlag.

Barlas, Y. 1985. Validation of System Dynamics Models With a Sequential Procedure Involving Multiple Quantitative Methods. *Unpublished Ph.D Dissertation.* Georgia Institute of Technology. Atlanta, GA.

Forrester J.W. and P.M. Senge. 1980. Tests For Building Confidence in System Dynamics Models. In *System Dynamics*. Ed. Legasto, A.A., J.W. Forrester and J.M. Lyneis. Amsterdam: North-Holland.

Forrester, J.W. 1961. *Industrial Dynamics*. Cambridge, MA: Productivity Press.

Forrester, J.W. 1973. Confidence in Models of Social Behavior with Emphasis on System Dynamics Models. *System Dynamics Group Working Paper.* Sloan School of Management, MIT, Cambridge:MA.

Senge, P.M. 1977. Statistical Estimation of Feedback Models. *Simulation.* 28: 177-184.

Sterman, J.D. 1984. Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models. *Dynamica.* 10(2):51-66.