International System Dynamics Conference 1998, Québec

Model Building and Validation: Contributions of the Taguchi Method

Markus Schwaninger, University of St. Gallen, St. Gallen, Switzerland
Andreas Hadjis, Technikum Vorarlberg, Dornbirn, Austria

1. The Context.

Model validation is a crucial aspect of any model-based methodology in general and system dynamics (SD) methodology in particular. Most of the literature on SD model building and validation revolves around the notion that a SD simulation model constitutes a **theory** about how a system actually works (Forrester 1967: 116). SD models are claimed to be **causal** ones and as such are used to generate information and insights for diagnosis and policy design, theory testing or simply learning. Therefore, there is a strong similarity between how theories are accepted or refuted in science (a major epistemological and philosophical issue) and how system dynamics models are validated. Barlas and Carpenter (1992) give a detailed account of this issue (Barlas/Carpenter 1990: 152) comparing the two major opposing streams of philosophies of science and convincingly showing that the philosophy of system dynamics model validation is in agreement with the relativistic/holistic philosophy of science.

For the traditional reductionist/logical empiricist philosophy, a valid model is an objective representation of the real system. The model is compared to the empirical facts and can be either correct or false. In this philosophy validity is seen as a matter of formal accuracy, not practical use. In contrast, the more recent relativist/holistic philosophy would see a valid model as one of many ways to describe a real situation, connected to a particular purpose. Models, are not necessarily true or false, but more or less „suitable" for their purpose. In this sense, validation is an evolutionary process of social conversation for building confidence in the usefulness of the model with respect to its purpose (Forrester/Senge 1980).

Although different types of SD models can be distinguished and ascribed to different purposes (i.e. policy design models, theory testing models, „flight simulators", generic models etc.) requiring and justifying probably different validation tests (Barlas 1996: 200), one basic assumption remains the same in all types of models: The modeling effort is essentially an *experiment* with the purpose of generating high quality, reliable information, which is generally used in one of three ways:
1. **General understanding** (diagnosing how the feedback structure contributes to dynamic behavior, identifying dominant structures, defining new concepts, organizing and communicating ideas and hypotheses, generally learning.)
2. **Policy design** (selecting objectives, evaluating strategies under different perspectives, finding and analyzing sensitive parameters, optimization of structure against objectives etc.).
3. **Implementation** (examination of long-term patterns of different configurations of the system modeled, extraction of operating plans and instructions etc.).

In the context of experimentation, validity essentially means the extent to which we really observe from an experiment what we say we observe. Validity of the results of an experimental study is decisively determined by the validity of the model constructed and used. Thus one has to make sure that the model constructed is a solid platform for experimentation, by accounting for the many facets of validity. The overall purpose of validity testing is then to establish validity of model structure and behavior and that is a necessary condition for all three types of uses of models. Most SD models are "causal-descriptive" and not purely "correlational" (data-driven). In this case, applying merely "output-validation" tests (such as statistical significance tests) is not enough to establish validity; since these models are statements about how the real system works, it is more important to establish validity of model structure and to explain how this structure generates the reproduced behavior.

Four types of validity are usually distinguished: statistical conclusion, internal, construct and external (Green/Albaum 1988: 208). All types of validity are more or less required for all three uses above. Hence the overwhelming difficulty of practitioners to select suitable tests out of a plethora offered in the literature, particularly when they are confronted with large, complex models.

A necessary condition for inferring causality is that there be covariation between the independent and the dependent variables. *Statistical conclusion validity* hinges on whether the independent variable X and the presumed dependent variable Y are indeed related and to which extent. After it has been determined that variables co-vary, the question arises whether they are causally related. This is the essence of *internal validity*. A given experiment is internally valid when the observed effect is due solely to experimental treatments and not due to some extraneous variable. The third type of validity, *construct validity*, is essentially a measurement issue. It revolves around the extent to which generalizations can be made about higher-order constructs from research operations and it is applicable to cause and effect relationships. In this sense it is a special aspect of *external validity*. External validity, however, refers more broadly to the generalizability of a relationship, beyond the circumstances under which it is observed and establishes how good is an experiment in terms of extent of allowed conclusions to be made to and across populations, persons, settings, times etc.

Directly related to this issue is the concept of dominant structure which plays a major role in the communication of model based insights. The concepts of generalizability, transferrability and inference value (of generic structures for example) are grounded in the notion of "dominant structure" (Richardson 1986) and behavior resulting from it. Ascertaining a dominant structure is a complex problem particularly in large nonlinear models, where it may be argued that it is impossible to realize a complete analysis using traditional approaches, involving repeated simulations until a hypothesis about the most influential structure is formulated and then a simpler model containing that structure is constructed and tested. The marginal contribution of the dominant feedback loop is then viewed in terms of graphs over time, eigenvalues or frequency response (Richardson 1986). Unfortunately, most of the available simulation software cannot automatically report eigenvalues and frequency response information. As this has discouraged modelers in the past to undertake formalized rigorous testing, the need for an efficient method for tracking sensitive parameters having strong influences on model behavior arises.

The problem we are confronted with is one of making "visible" the relationships between structure, noise (external-internal) and prediction of the future system state (behavior). For example, in a corporate model dealing with market dynamics, if the aim is to maximize market share (a specific objective for a possible use of such a model as a planning tool), the modeler needs to know how different values of "design parameters" create variances in the model output for market share. The aim is always to gain confidence that the model produces the right behavior for the right reasons. In this context, ascertaining the sensitivity of model behavior to parameter changes is a critical step in model building, simulation, and, in particular validation. In addition, knowledge about impact differentials between action variables ("policy" or "strategy" variables) can contribute substantially to the robustness of a policy or a strategy.

Consistent with the relativist/holistic philosophy and bearing in mind that information generated by experiments must have both *statistical conclusion* and *internal validity* (established through correlation techniques) and, more importantly, *construct and external validity* (i.e. generalizibility and inference value), we have followed the testing environment appearing below (figure 1).
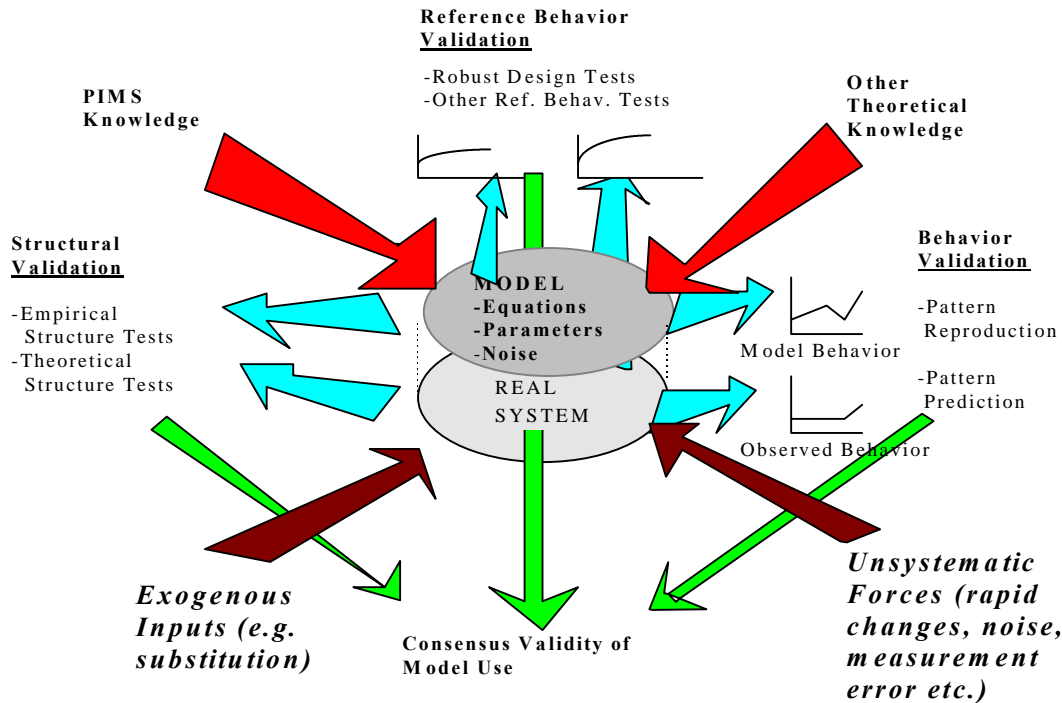
Figure 1. Proposed testing environment. (after: Hadjis 1997)

Note that we propose the parameter design tests[1] right at the beginning of the formal testing procedure and in the category of reference behavior tests although, strictly speaking, they could be positioned in the sub-category of direct structural validation (as proposed by Barlas 1996 and Forrester/Senge 1980) . The major purpose is to help the modeler out of possible conceptual confusion and offer him/her a solid platform to stand on for carrying out the complex testing tasks ahead, in an efficient and economical way. By first investigating the structurally sensitive areas and parameter tolerances (i.e. the limits of "*incubating inertia*", Beer 1985), the modeler creates the conceptual conditions in his mind for carrying out the subsequent direct and behavior oriented structural tests in a focused manner, gradually increasing confidence in his model. This is in line with the philosophy of gradual consensual model validation and constitutes an important step of *complexity management*.

2. The Issue.

In modeling "physical phenomena" the assumption is that the "inertial forces"(Forrester 1967: 133), i.e. the physical laws ruling the parameters, variables and their relationships are essentially much stronger than the noisy elements, i.e. the unsystematic forces affecting the real system. This may not always be the case in reality, particularly when complex systems are modeled, in which parameters and variables are in continual interaction which may lead to surprise behavior. The question arises, when and how this surprise behavior will occur.

Therefore, we need a methodology of sensitivity analysis capable of investigating the „space" of parameter values, but also the sensitivity of parameter interaction. With such a methodology, a first idea about the extent of noise influences on the model-output behavior can be obtained, and „suspect" structural areas isolated for deeper consideration as possible sources of distortion.

---

[1] For a full description of the Taguchi methods for efficient sensitivity testing, see: Phadke 1989.

In SD simulation models, a distinction is normally made between design parameters and noise factors. Parameters are constants, coefficients or exponents, the values of which are determined by the modeler. A noise factor can be called *exogenous* because it is not generated by the model and varies randomly. In models of "physical phenomena" we may know their values with high accuracy and thus exclude the parameter from testing, but, as far as models of social systems are concerned, in many cases these parameters are not known. Hence they constitute a major source of uncertainty and a major cause of „unexpected" behavior.

The literature on sensitivity analysis for system dynamics is not very large (Clemson 1995: 31). Traditional methods using the „one variable at a time" (holding everything else constant) approach and simple random sampling (or similar) processes, may be efficient enough for small models with very few parameters. For models with a larger number of parameters however, traditional methods lose their efficiency very fast, because one normally runs out of resources and patience before all parameters can be exhaustively tested. In the case of 9 parameters (as in the example in section 3), each of which is to be tested for a low , a medium and a high value, a full factorial experimental approach would require 3^9 or 19'683 trials. Furthermore, traditional methods using the univariate approach cannot identify parameter interaction sensitivities, which in SD models is very likely to appear (Forrester/Senge 1980: 213).

An efficient solution appears if we recognize the similarity of the problem at hand (investigation of signal-noise relationship), with the problem of product quality improvements in industrial engineering contexts. The nature of the modeler`s problem is almost identical to the design-engineer's challenge in product development, to reach the best product quality. The modeler`s „product" is the model and the „quality" he has to reach is the model's degree of validity with respect to its purpose (Barlas 1996: 203). The design-engineer spends a great deal of his time in order to win information about the effects of different development parameters on product performance, in various usage contexts. Engineers have effectively used the methods of Robust Design (Taguchi 1978) to generate reliable information for their design decisions. They have developed a very efficient and economical way of experimental design that drastically reduces the number of required experiments.

The fundamental principle of „Robust Design" methods is product improvement through minimization of the effects of various „causes of variation". This is achieved by an optimization of the product and process development procedure, so that the sensitivity of product performance to the variation of design parameters stays at a minimum. The engineers call this process „*parameter design*" (Phadke: 1989: 42-49), a term which makes the similarity to the modeler`s problem quite obvious.

The three most important steps in product development are *concept design* (selection of product technology for example), *parameter design* (selection of optimal levels for the steering variables to achieve at least a minimum robustness) and *tolerance design* (definition of values for the tolerance factors). There are almost identical concerns of the modeler who has to find out which parameters show high impact on his „product's" quality, i.e. the validity of his model with respect to its purpose and usefulness. Using this engineering metaphor makes theoretically solid structural tests like the *robust design tests* accessible to SD methodology, with the purpose of obtaining insights on "parameter design". The respective insights are in turn applied to further structure and structure-oriented testing. Therewith, knowledge about aspects of "dominant structure" (Richardson 1986) can be gained. This is not only helpful to direct attention of model builders to "essential" parameters. It is also of great value to ascertain areas of concern for further testing and priorities for parameter optimization.

3. An Application
In a research project funded by the University of St. Gallen we attempted to leverage the complementarities of SD modeling and simulation capabilities on one hand and of multivariate models

validated with data from the PIMS/SPI[2] database on the other. This appeared to be a promising endeavor for testing the PIMS logic (i.e. "the laws of the market" (Buzzell/Gale 1987: 12-46)), which emerged from a different path, i.e. the multivariate analysis of large volumes of data. At the same time this could entail a desirable improvement of planning methodologies: Solid theoretical-conceptual models, in which the logical levels of management with their pertinent criteria of organizational fitness and hence steering sequences are clearly represented and validated empirically[3], could be turned into formalized dynamic environments for organizational planning and learning. We have taken the following steps to combine the strengths of both methodologies:

1. Construction of a generic corporate model on the basis of the PIMS models, however additionally containing feedback, delays and non-linearity.
2. Validation of the SD model with data from the PIMS database and
3. Comparison of the relative performance of the SD and PIMS models, particularly with respect to behavior reproduction.

Management issues and relevant theories would determine model and module boundaries and the scope of each module of the model, as well as interconnections and aggregation-disaggregation of variables within modules. The model should be capable of addressing multiple dimensions of strategy, including different market dynamics and product categories. The investigation of modern approaches such as stake- or shareholder value as well as considerations of normative management reflecting different organizational cultural types and strategic predispositions[4] was also to be made possible.

The final conceptual layout for the model is shown in figure 2.
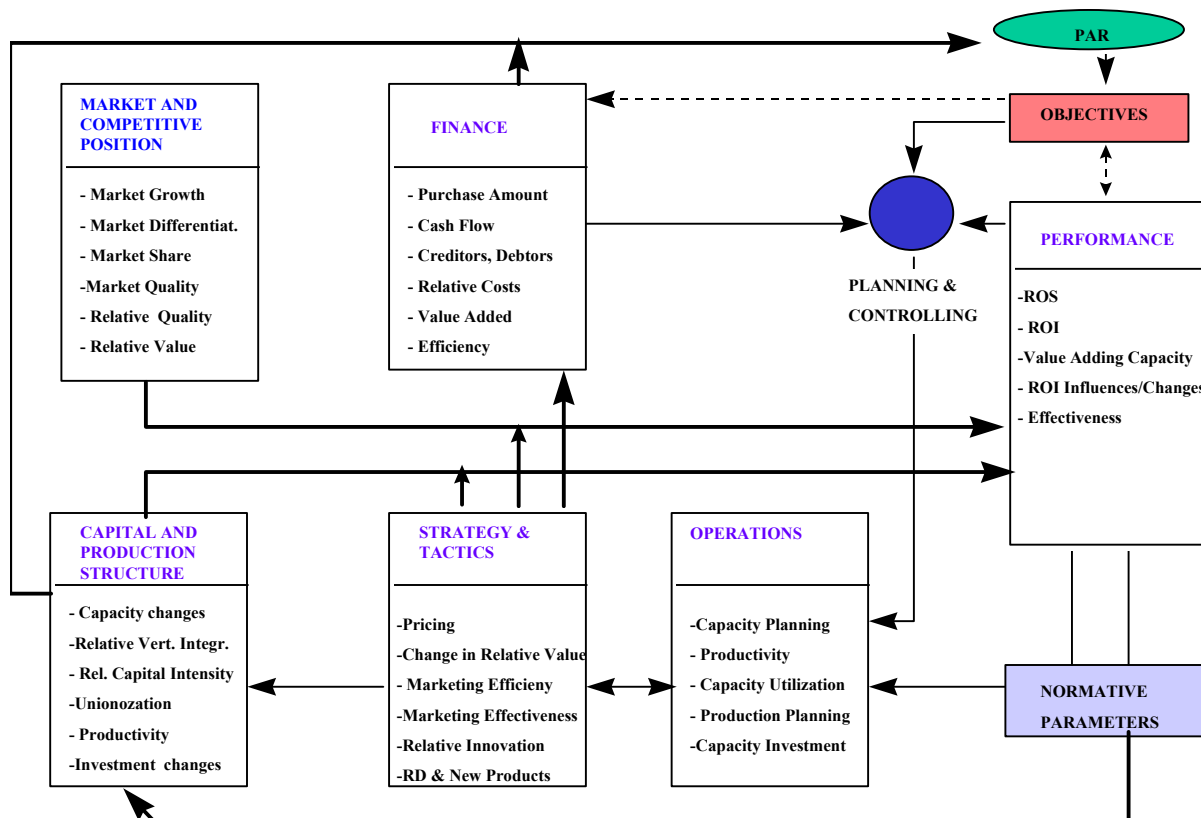

Figure 2. Architecture of the Model

[2] PIMS stands for Profit Impact of Market Strategy. This acronym refers to a database and quantitative simulation methodology developed by SPI-Strategic Planning Institute, Cambridge, Massachusetts.

[3] See Schwaninger (1989: 169-203), which has a basis for the St. Galler Management Framework.

[4] See for example Kotter /Heskett 1992.

The modular architecture used, which matches the PIMS strategy paradigm, offers structural discipline for the purpose of achieving a valid reduction of the number of variables, as some 18 variables, measured by PIMS, explain more than 80% of ROI variances (statistical-internal validity). Additionally the modeler can benefit from aggregation and dissaggregation possibilities, which enable selective focusing on dominant issues and better selection (through partial sensitivity analysis) of parameters and steering levers (investigation of external validity-generalizability). In this manner, the comparison between model behavior structures and generic or observed ones (such as the PIMS findings) could proceed with partial simulation, reinforcing a gradual confidence building and making the complex formal validation procedure a lot easier. This construction, while still drawing on the PIMS framework, allows the inclusion and investigation of many approaches developed outside PIMS, in a rigorous manner. The arrows show most of the interactions between modules and the boxes most of the types of issues that the module can handle.

The module selected for demonstration of the application has the structure shown in the diagram in figure 3.
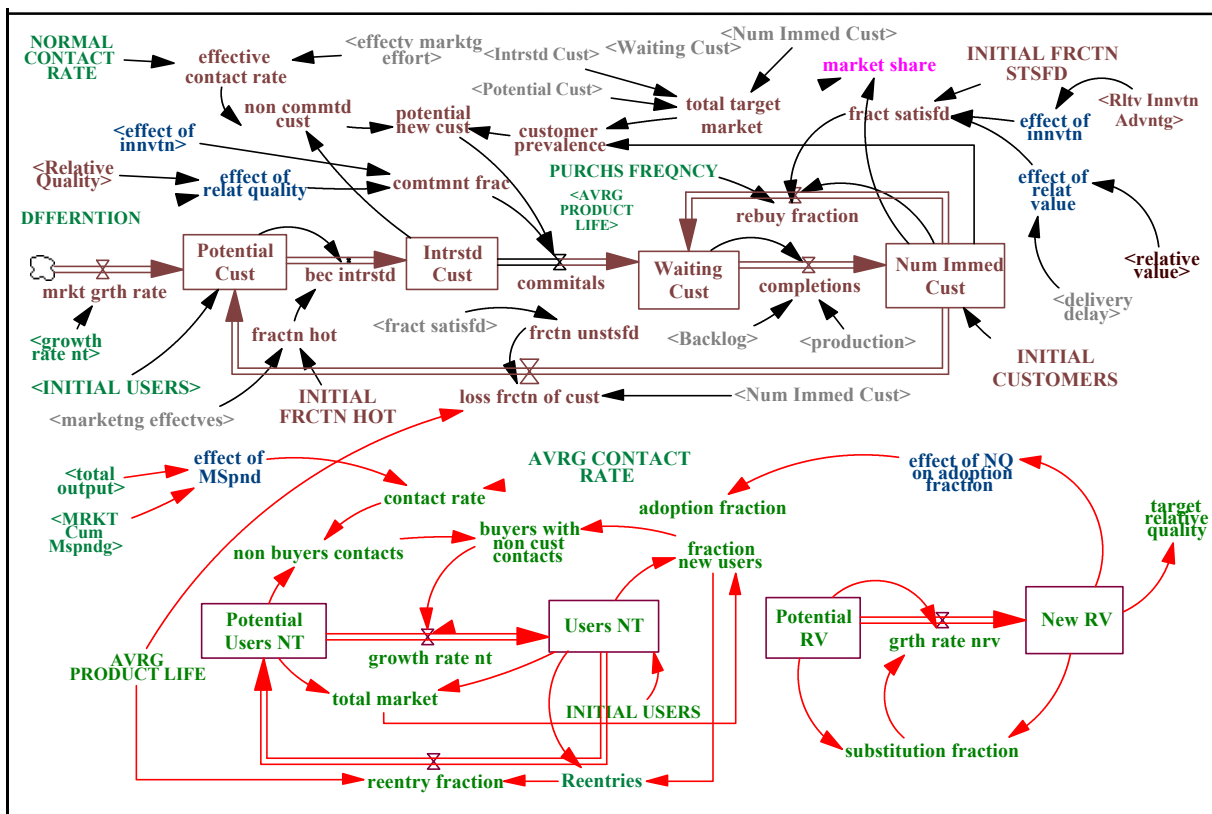

Figure 3: Diagram of the market module of the SD model

The Taguchi method is based on 18 standardized orthogonal arrays. In many cases, one of these arrays can be used directly or can be modified to fit a specific project. A given matrix can be used with fewer than the maximum number of parameters (e.g. fewer than 63 as established in the largest standardized orthogonal array). In the case where there are noise factors in the model, a noise matrix is constructed with trials along the top and noise factors down the side. Then the parameter matrix is crossed with the noise matrix to get a set of experiments providing information about all parameters and all noise conditions.

In some simulation models, there are no noise factors, and the tests required are specified by the parameter matrix alone. Parameters (constants) must correspond *conceptually* and *numerically* to real

life. Most of our model's equations are derived from PIMS descriptions of „real life". As a result, the model's structure is rather deterministic than „noisy", a fact shown by the absence of oscillations on the major output-performance variables. Thus we confined ourselves to the parameter matrices after Taguchi. Nevertheless, equations that have random
components received our special attention. Applying the method separately (when possible) on the individual modules of the model, significantly facilitates the process.

The steps for using the Taguchi methods in sensitivity testing are given below. In addition to these steps, a reference text that provides Taguchi matrices (e.g. Phadke 1989) is also required.

1.) Define characteristic measures of model behavior: The problem is to select those curves that best fits the purpose of the model and collectively provide the relevant measure of model output. In our case the curve of market share fulfills this criterion nicely. The curve of ROI or ROS could also be used, particularly for modules directly responsible for the generation of these outcomes.

2.) Identify the noise factors: Noise factors are the (mostly exogenous) ones, varying randomly or unpredictably. In a model like the one constructed (descriptive-deterministic), there are no exogenous factors, particularly in the case when the model is adjusted to represent a specific firm in a certain market situation, at the beginning of a simulation. Within the model and module boundaries some of the different *effects* will be „noises" and some will be functions the data points of which are largely known. Those effects which correspond to measured PIMS-graphs can be excluded from analysis, for example the effect of innovation in the market evolution module. On the contrary, effects not already well known should be included for analysis, particularly in the phase of policy analysis.

The effect of "marketing effectiveness" on "fraction hot" (those interested and actively seeking solutions) in the market module would definitely be a candidate, if we were exploring control strategies to maximize market share. In that case, we would clearly have to deal with all noise factors surrounding the relevant structure. However, at this stage of testing, our concern is to proceed at sensitivity analysis in order to determine parameter sensitivities and parameter interaction sensitivities, with the purpose of detecting sensitive structural areas and eliminating those very sensitive parameters which distort the model's behavior. Thus we do not consider noise factors.

3.) Identify parameters and their values: This step requires listing all constants, coefficients or exponents whose values are uncertain. The experiments must cover the whole bandwidth of possible parameter values. Another important issue is the numerical adjustment of the parameters to specific product categories and the corresponding observed real live parameters. For example, the normal contact rate or purchase frequency is obviously quite different for consumer goods (also between consumer durables and consumables) and for industrial products. PIMS-findings help decisively in selecting plausible minimum, medium and maximum values for these parameters. In our case, this step means listing parameters and effects in each module.

4.) Design matrix experiment and define data analysis procedure: This is simply to select a parameter matrix and a noise matrix that best corresponds to the case at hand. Selecting matrices depends upon the number of parameters and noise factors, the number of levels (values) used with each parameter and the expected interactions among them.

5.) Conduct matrix experiments: This is a simulation run with the parameter and noise conditions defined in the matrices. The results (also called the objective function) are listed in the last column (J) of the matrix (see example below, and data in figure 4).

6.) Calculate the response table: Here we take the average of the objective function values for each value level of each parameter. These average results are then inserted in cell a, level 1 etc. The procedure is repeated for all levels and cells to complete the *response table* (example: figure 5). Finally,

delta is the maximal difference between means for the value levels in each column. A large delta indicates that the model is sensitive to the respective parameter.

7.) Prepare the response table among parameters: The additional columns in the chosen matrix, those we originally did not use, allow us to assess the degree of interaction among parameters (example: figure 6). This step is possible only because earlier, before selecting a matrix, suspected interactions were specified and the suitable matrix supporting this analysis was chosen (see Phadke 1989: 121).

8.) Confirm the results: Taguchi methods assume that the experimenter has identified all strong interactions among parameters. The overall results could be nonsensical if some strong interactions were not considered. The main effects response table (step 6) allows us to predict the parameter combinations that produce the highest and lowest output measures. For instance, the lowest values should be generated by setting all parameters' values at level 1 (lowest). Similarly, the highest output measure (market share in our example) should be produced by setting all parameters at level 2 (if two value levels were chosen), or level 3 (highest) if three value levels were chosen. If the predictions are not confirmed, necessarily we must reconsider the interactions and run the experiments again with a new design, based on the new interactions.

In our case, knowledge of PIMS findings on different market situations and product categories could significantly reduce the amount of possible parameter interactions, and therewith decisively reinforce this verification process. This has not excluded, of course, the comparison of various parameter combinations (e.g. all set low or high) with other empirically researched results about the output measure (market share).

In order to spare the reader time and effort, we give below only the results of the experiments realized with regard to the market module. Similar tests were realized for each module containing parameters with strong influence. The numbers of the following example correspond to the steps explained above:

1. The market share curve was selected to provide an overall measure of model behavior.
2. For this module exogenous variables varying randomly or unpredictably were excluded, at this first stage of sensitivity analysis. If we were exploring control strategies (or steering possibilities) to maximize model output performance we would clearly have to deal with noise factors, coming from the early and late phases of market evolution (substitution and diffusion phenomenon) and mainly affecting the market growth rate. This would imply designing and conducting experiments based on concrete hypotheses derived from „other" than PIMS- theories or „laws" of market evolution. Since these other possibilities are surprisingly few (one example is Marchetti 1985), the model can be easily adjusted to account for them. In this case, plausible scenarios dealing with possible market evolution paths could be reduced to not more than 2 - 3 relevant different reference behaviors for diffusion or substitution, involving investigation of very few parameters (and their respective values). When predictions achieved with this proposed method are cross-checked with forecasts obtained through regression or other methods, then, in case of positive correlations, confidence in model use can be drastically increased. Another source of noise would be the product category, which would largely determine possible values of characteristic parameters. To eliminate this second source, parameters and values were chosen each of the six product categories included in the PIMS database (See Buzzell/Gale, 1987: 127). In this case we took the case of consumer durables. Needless to say that for each product category a model adjustment is necessary for a new set of the same experiments.
3. The following parameters and effects were chosen for an investigation concerning the first module:
A) Normal contact rate: Value levels 50, 100, 150 („visits" to customers per year), representing different „activity levels" of major competitors, hence capturing different intensities of competition. Obviously, measurement of this parameter in the served market of an SBU does not present major difficulties.

B) Average contact rate (Lower chain market evolution): Value levels 100, 150 (number of customer contacts per year). At the market level a value of 50 would be too small to represent the activities of all competitors.

C) Differentiation: Value levels 1.5, 2. According to PIMS definition, differentiation is the number of product characteristics (as defined in a quality index) perceived and used by customers to make up their purchase decision. A level of 1 would mean an undifferentiated market and a level of 2 a highly differentiated one. Since this set of experiments deals with consumer durables, a highly undifferentiated market is rather unlikely (think of stereo equipment, refrigerators or washing machines).

D) Initial fraction hot: Value levels 0.3, 0.5, 0.7. This parameter represents the SBU's initial ability to create interest among potential users of its offering. We considered that if the SBU's offering meets the interest of less than 30% of potential users (or is unknown to more than 70%), then it is rather unlikely to become a considerable player in the served market. Subsequent simulations have proven indeed that with a starting level of 0.1 or 0.15, the company under examination cannot win market share, in spite of intensified marketing efforts.

E) Average product life: Value levels: 1, 2, 4 years, covering the range of most consumer durables in a realistic way.

F) Initial customers: Value levels: 5, 10, 20. The assumption is that the SBU is a manufacturer selling to distributors (values are supported by PIMS-data). Different market conditions, supported by PIMS observations can be easily extracted for model adjustment, in order to represent real market conditions.

G) Initial users: Value levels 50, 100, 200. For a previsible diffusion pattern to take place, the number of initial users of the product should reach a minimum of 10% of potential users (Marchetti 1982). Thus adjustment of this parameter should depict the market conditions under study, at the beginning of simulations. In this case, the parameter is easily measurable under different market conditions.

H) Initial fraction satisfied: Value levels 0.2, 0.5, 0.8. This parameter represents the SBU's initial ability to satisfy its existing customers

I) Purchase amount: Average values of this parameter expressed in thousands of money units per order, are easy to extract from the PIMS database, depicting different ordering patterns observed for different products and distribution realities. In the absence of PIMS data, the range of values for this parameter can easily be measured, given the specific market conditions under study.

All effects were chosen and calibrated to represent PIMS-known graphs. At this stage, no sensitivity simulations were realized. Nevertheless, when discussing output maximization policies and steering possibilities, then all kinds of effects become candidates for context specific investigations:

4.The following matrix was chosen as the smallest that fits the nine parameters and their most important hypothesized interactions. Among different possibilities of parameter interaction, only two were considered relevant, i.e. parameters A x B (normal contact rate) with 3 and 2 levels respectively, and DxH (initial fraction hot x initial fraction satisfied) with 3 levels each. The chosen L 27 matrix appears below with columns A to I referring to the nine parameters specified above, and column J to the outcome variable, market share.

Exp Nr.	A	B	C	D	E	F	G	H	I	J
1	50	50	1	0.3	1	5	50	0.2	0.02	0.1332
2	50	50	1	0.3	2	10	100	0.5	0.04	0.2407
3	50	50	1	0.3	4	20	200	0.8	0.06	0.3788
4	50	100	1.5	0.5	1	5	50	0.5	0.04	0.2496
5	50	100	1.5	0.5	2	10	100	0.8	0.06	0.4246
6	50	100	1.5	0.5	4	20	200	0.2	0.02	0.5055
7	50	150	2	0.7	1	5	50	0.8	0.06	0.4019
8	50	150	2	0.7	2	10	100	0.2	0.02	0.4686
9	50	150	2	0.7	4	20	200	0.5	0.04	0.6844
10	100	50	1.5	0.7	1	10	200	0.2	0.04	0.3324
11	100	50	1.5	0.7	2	20	50	0.5	0.06	0.4055
12	100	50	1.5	0.3	4	5	100	0.8	0.02	0.5288
13	100	100	2	0.3	1	10	200	0.5	0.06	0.1959
14	100	100	2	0.3	2	20	50	0.8	0.02	0.2571
15	100	100	2	0.5	4	5	100	0.2	0.04	0.3241
16	100	150	1	0.5	1	10	200	0.8	0.02	0.3287
17	100	150	1	0.5	2	20	50	0.2	0.04	0.3743
18	100	150	1	0.5	4	5	100	0.5	0.06	0.5582
19	100	50	2	0.5	1	20	100	0.2	0.06	0.2448
20	150	50	2	0.5	2	5	200	0.5	0.02	0.3912
21	150	50	2	0.7	4	10	50	0.8	0.04	0.3954
22	150	100	1	0.7	1	20	100	0.5	0.02	0.3433
23	150	100	1	0.7	2	5	200	0.8	0.04	0.57
24	150	100	1	0.7	4	10	50	0.2	0.06	0.489
25	150	150	1.5	0.3	1	20	100	0.8	0.06	0.2207
26	150	150	1.5	0.3	2	5	200	0.2	0.06	0.2771
27	150	150	1.5	0.3	4	10	50	0.5	0.02	0.367

Figure 4. Orthogonal Matrix for Parameter Testing

5. Results of running the matrix simulation experiments are shown in column J. These results are change statistics for the market share curve, calculated from the model behavior under the parameter combinations specified in each experiment row. For example, experiment number 9 having values for normal contact rate 50, average contact rate 150, differentiation 2, initial fraction hot 0.7, average product life 4 years, initial customers 20 (high level), initial users 200, initial fraction satisfied 0.5 and purchase amount 0.04 (million money units) results in a market share of 0.6844, i.e. 68% of market served. This is quite a plausible scenario supported by the PIMS-finding that in highly differentiated, growing markets (differentiation 2, initial users 200), starting from a good position of quality (fraction hot 0.7 and fraction satisfied 0.5) can lead to high market shares in the maturity phase

6. Calculating the response table means taking the average of the objective function (column J) of the matrix (figure 4) values for each level of each parameter. For instance, parameter A level 1 (50) appears in rows 1 to 9. The average of the results of these rows gives 0.3874, which is then inserted into cell a, level 1, of the following response table (figure 5).
.

	A	B	C	D	E	F	G	H	I
Level 1	0.3874	0.3398	0.3795	0.266	0.2731	0.3815	0.3414	0.3498	0.3692
Level 2	0.3672	0.3732	0.3679	0.3858	0.3887	0.3602	0.3726	0.3817	0.3768
Level 3	0.3665	0.3898	0.3737	0.4693	0.4701	0.3793	0.4071	0.3894	0.375
Delta	-0.0209	0.05	0.0058	0.2033	0.2731	0.0021	0.0657	0.0396	0.0058

Figure 5. The Response Table

The procedure is repeated for all cells to complete the response table. Finally, delta is the difference between the means for the highest and lowest levels. A large delta signifies that the model is sensitive to the respective parameter. In our case parameters D and E (deltas 0.2033 and 0.273, respectively)

show a relatively high sensitivity, followed by B, G, H and A with rather small sensitivities. Parameters F and I have the lowest sensitivities, hence they can be ignored for further investigation.

7. The response table for the parameter interactions investigated is given below (figure 6).

	A*B		D*H
Level 1	0.3876		0.2509
Level 2	0.3672		0.3932
Level 3	0.3654		0.5183
Delta	0.0373		0.2674

Figure 6. Parameter Interactions Response Table

Results of this table show that the first interaction (A*B) is rather weak and can be ignored, while the second interaction (D*H) is rather strong. This tells us, that if we run additional sensitivity simulations we may ignore the first interaction and thus choose a smaller matrix with even fewer experiments. For this purpose, averages from ROLA-reports (Reports on Look Alikes[5]) can be used for testing and comparison. This way, the possibility often not available for validation of SD-models due to lack of databases, of directly evaluating parameters from knowledge existing about the operating world is open to the integrative approach followed by this project.

Generally, we may say that the parameter sensitivity tests showed that the model is
- rather insensitive to parameters C (differentiation), F (initial customers) and I (purchase amount.
- moderately sensitive to parameters A, B, G, H.
- most sensitive to values chosen for parameters D, E (initial fraction hot and average product life).

The finding that parameters D and E are highly sensitive and can be useful in one of 3 ways:
1.) Restructure the model to include more detail about the effects of these two parameters.
2.) Collect more data to refine estimates about these parameters.
3.) Remove the parameters and revise the model.

The third possibility was excluded because, as already discussed , the method does not directly detect structural flaws but merely indicates suspect areas. Parameter E represents the average product life which is a valid concept determining repurchase and hence market share changes (win or lose possibilities); therefore it cannot be removed. The same is valid for parameter D, initial fraction hot. In this second case, more data must be collected to refine estimates. That means, in using the model for policy formulation the model must be adjusted to include context specific measurements. An indicator could be a „classical" PIMS index measurement of relative quality (Hadjis 1995: 128-134). The same path was chosen for parameter E (average product life). In this case, collecting more data on the actual product (on physical, service and image attributes) and served market, when using the model on a specific company, would be the proper way.

Obviously, the results of this first test rounds may now synergistically flow into the direct structural tests (see figure 1 above).

4. The Implications
SD methodology must meet three basic requirements in order to achieve the challenging task of formalizing validation processes: First, to provide modelers with an accessible and complete

---

[5] ROLA reports are a PIMS tool, which ascertains the statistical values for outcome variables of subsets of business units characterized by similar structural features (e.g. high differentiation and short life cycle and high marketing intensity).

theoretical-conceptual framework allowing the inclusion of many existing and emerging tests, to account for the multiple facets of validity. Second, to develop innovative tests that can be synergistically applied in a mutually reinforcing way, and third, to incorporate appropriate mathematical and statistical tools, derivied from this framework, in commercially available software.

The Taguchi method for experimental design applied in the testing context as depicted in figure 1 contributes substantially to fulfilling these requirements. From our experience, the method avoids the messy procedures of normal sensitivity analysis by drastically reducing the number of necessary experiments and producing well documented and easily interpretable data. We found that the application of the Taguchi method in the first testing field as proposed here (robust design tests in figure 1 above), creates a focused framework by pinpointing suspect areas and giving correct insights for the application of the structural validity tests, both direct structure tests and transitional structure oriented behavior tests, the purpose of which is to distinguish "true" from "spurious" behavior accuracy. Finding the value ranges (bandwidths) and the variances caused in the model outputs of interest, impact differentials can be assessed and used for parameter optimization, in connection with model use. For example, generating scenarios corresponding to extreme conditions, i.e. with parameter values beyond the observed past, are made possible. One could use this procedure also to create strategic "early warning" systems.

The primary goal of systems dynamic methodology is to understand how the feedback structure of a system contributes to its dynamic behavior. According to Richardson (1986), determining dominant structure involves two sets of choices: how to characterize behavior, and how to define principal structure. Behavior can be represented in terms of graphs over time, eigenvalues or frequency response. Two ways of assessing dominant structure are in terms of marginal contribution of a loop to a given behavior and in terms of a reduced model that exhibits the behavior of the dominant structure. The problem arising is that in nonlinear systems behavior cannot be consistently summarized in terms of eigenvalues or frequency response, as in these systems eigenvalues may change continuously over time. Assessing model reduction and loop contribution involves repeated simulations consuming much time and effort.

The methods proposed here offer a formalized way to focus attention on suspect areas in an economic way. In the sub-category of the theoretical structural validation tests *parameter confirmation, structure confirmation* and *direct extreme conditions* testing are strong tests for assessing validity of structure. Both are greatly facilitated on the basis of insights obtained from using the Taguchi method. In the category of the structure-oriented behavior tests, *indirect extreme condition testing* (as for example with the "reality checks" facility of VENSIM[6]) provides another example of the synergistic application of the methods in an integrated formal testing procedure. Further, Taguchi methods can be run with almost any available SD-software.

Finally, a word of caution: The method outlined assumes that the relation between parameters tested by sensitivity analysis and the respective output measure (in our case market share) is linear. In cases where this assumption must be entirely abandoned, the method might produce false results. Another limitation is that the maximum number of parameters that can be checked by means of the standardized orthogonal arrays available is 63.

References:

Barlas Y.,   Model Validation in System Dynamics, in: Proceedings of International
             System Dynamics Conference, Methodological Issues, pp.1-10, Sterling, Scotland, 1994.

---

[6] Software of Ventana Simulation, 149 Waverley Street, Belmont, MA 02178, USA, Reference Manual Version 1.62.

Barlas Y./
Carpenter S.,   Philosophical Roots of Model Validation: Two Paradigms, in: System
                    Dynamics Review, Vol.6, No. 2, 1990, pp.148-165.

Barlas Y.,   Model Validation in System Dynamics, in: Proceedings of International
                    System Dynamics Conference, Methodological Issues, pp.1-10, Sterling
                    Scotland, 1994.

Beer S.,   Diagnosing The System for Organizations, Chichester: J. Wiley and Sons , 1985.

Buzzell R.D./
Gale B.D.,   The PIMS Principles, New York, London: Free Press/Collier Macmillan,1987.

Clemson B./
Tang Y./
Pyne J./
Unal R.,   Efficient Methods for Sensitivity Analysis, in: System Dynamics Review,
                    Vol.11, No.1, Spring 1995.

Eberlein R./
Qifan W.,   Statistical estimation and system dynamics models, in: Proceedings of the 1985 International
                    Conference of the Systems Dynamics Society, pp.206-222.

Forrester J.,   Industrial Dynamics, Cambridge, MA: MIT Press, 1967.

Forrester J./
Senge P.,   Tests for Building Confidence in System Dynamics Models, in: System Dynamics, A.A. Legasto
                    Jr., ed., TIMS Studies in the Management Sciences, Vol.14, New York: North Holland 1980.

Green P./
Tull D./
Albaum G.,   Research for Marketing Decisions, 5th ed., Englewood Cliffs, NJ: Prentice Hall International
                    1988.

Hadjis A.,   Composite Models in Strategy Development, St.Gallen: HSG Ph.D. Dissertation , No.1641, 1995.

Hadjis A.,   Corporate Models: Integration of PIMS and System Dynamics, in: Research Report, University of
                    St.Gallen, 1997.

Kotter J./
Heskett J.,   Corporate Culture and Performance, New York: Free Press, 1992.

Marchetti, C.,   Die magische Entwicklungskurve, in: Bild der Wissenschaft, Nr. 10, 1982, 115 - 128.
Marchetti C.,   Time Patterns for Technological Choice Options, Laxenburg, Austria: IIASA Working Paper,
                    December 1985.

Phadke S.M.,   Quality Engineering Using Robust Design, Englewood Cliffs, NJ: Prentice Hall, 1989.

Richardson G.,   Dominant Structure, in: System Dynamics Review 2, No. 1, Winter 1986, pp.68-75.

Schwaninger M., Integrale Unternehmensplanung, Frankfurt/New York: Campus, 1989.

Sterman J.D.,   Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models, in:
                    Dynamica 10 (2), pp.51-56, 1984.

Taguchi G.E./
Konishi S.,   Orthogonal Arrays and Linear Graphs, Dearborn, Mich.: American Supplier Institute, 1987.