Credibility, validity and testing of Dynamic Simulation Models

Yaman Barlas
Boğaziçi University
Industrial Engineering Department
34342 Bebek Istanbul, Turkey
ybarlas@boun.edu.tr
http://www.ie.boun.edu.tr/?q=faculty/yaman-barlas
SESDYN Group: http://www.ie.boun.edu.tr/labs/sesdyn/

Conceptual and Philosophical Foundations

- Model Credibility as Related to Two Main Types of Models
  - Statistical Forecasting models (black box)

    Purpose: Assisting decisions by providing forecasts

    Nature: Short term (a few) point forecasts

    Procedure: Curve fitting to given data (Regression, time series anaysis, most econometric models…)

    Scope: Typically one dependent and many independent variables

    Essence of credibility: statistical fit of model output to real data

  - Descriptive/Causal Policy models (transparent)

    Purpose: Assist in policy evaluation, improvement and design

    Nature: Long term trajectory (dynamic behavior) forecasts

    Pocedure: Causal-desriptive modeling of real processes (Models in sciences, simulation, system dynamics, some economic models…)

    Scope: Typically many inter-dependent variables and a few independents

    Essence of credibility: Adequacy of relations in the model (‘structure’)

Philosophical Concepts and Issues (For Causal-Descriptive Models)

- Philosophy of Science (verification*, falsification, and…)

- Logical Empiricism and Absolute Truth
- Verification/falsification as a ‘true-false’ outcome

- Conversational justification & relative truth
- Validity (credibility) as ‘adequacy’, ‘sufficiency’, ‘usefulness…
- Theory testing as a prolonged, gradual process
- Theory Justification as a social process
- ‘Purpose’ and credibility; all stakeholders…

- Role of statistical significance testing

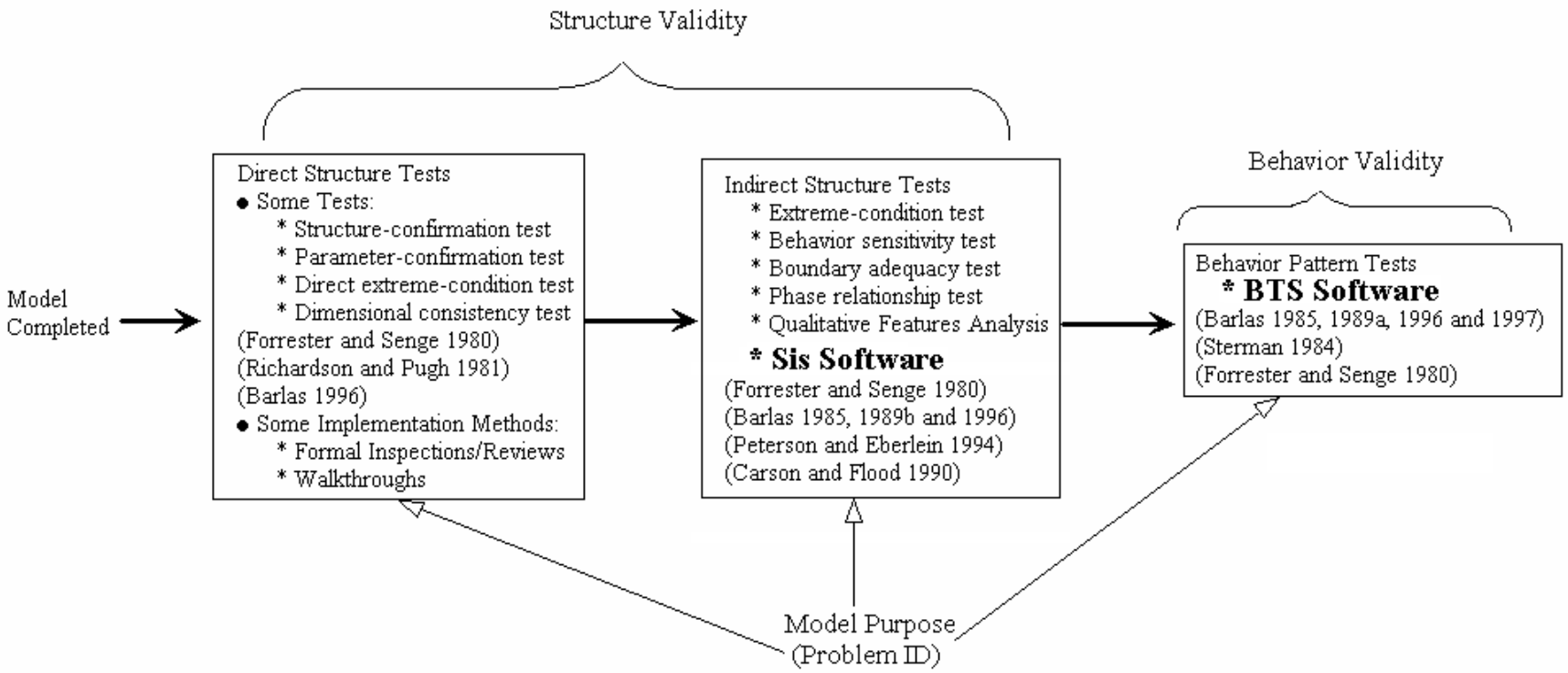
*Terminology confusion/reversal: validation-verification

Two aspects of model credibility
(For Causal-Descriptive Models)

- Structure Credibility
  - Primary role
    (‘Explanation’ in Philosophy of science)
    (Validation/evaluation in descriptive modeling fields)

- Behavior (output/predictive) Credibility
  - The ‘problem of induction’ in Philosophy of Science
  - Role in causal-descriptive modeling fields
    (‘right behavior for the right reasons’)
  - Ex ante versus ex post prediction

Overall Nature and Selected Tests of Dynamic Model Testing (Evaluation)



Structure Credibility Testing

- (*Verification*)

- Direct Structure Tests
  - Direct comparison of model structures with real ones
  - Crucial, yet highly qualitative and informal
  - Distributed through the entire modeling methodology

- Indirect Structure Tests (Structure-oriented behavior)
  - Extreme condition and other ‘special’ test simulations
  - Crucial, and also partly quantitative and formal

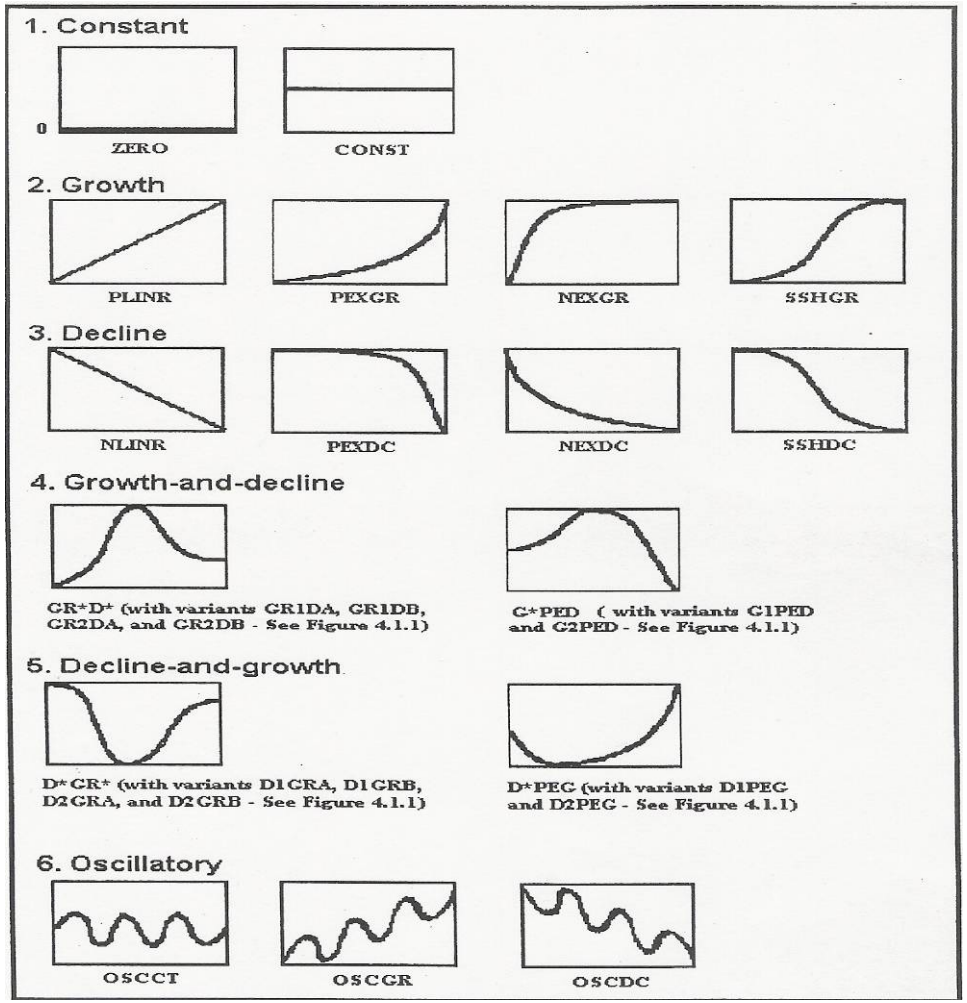
Validity (Quality) ‘Built-in’ vs. ‘Tested’ (Inspected)

- Problem ID and purpose
- Resolution, aggregation, time unit and horizon
- Verification (consistency) tests first. (Does the simulation model accurately represent my conceptual model? Does it do what I intend to do?)
- All variables & parameters with explainable meanings
- All equations with explainable meanings
- Units and unit consistency
- Established equations and structures (in literature) must be used
- Must start with a SMALL model (does NOT mean SIMPLE!)
- Embellish gradually, step by step, one structure at a time, by partial tests
- Models ‘Large enough but not larger than necessary’! (parsimony)
- Good documentation crucial
- And try to end with a SMALL model! (A generic, presentation version of the full model –establishing credibility ultimately means convincing people)

Indirect Structure Testing Software: SiS

- Based on automated dynamic pattern recognition
- Extreme condition pattern testing
- Also in parameter calibration and policy design

Indirect Structure Testing Software (SiS)



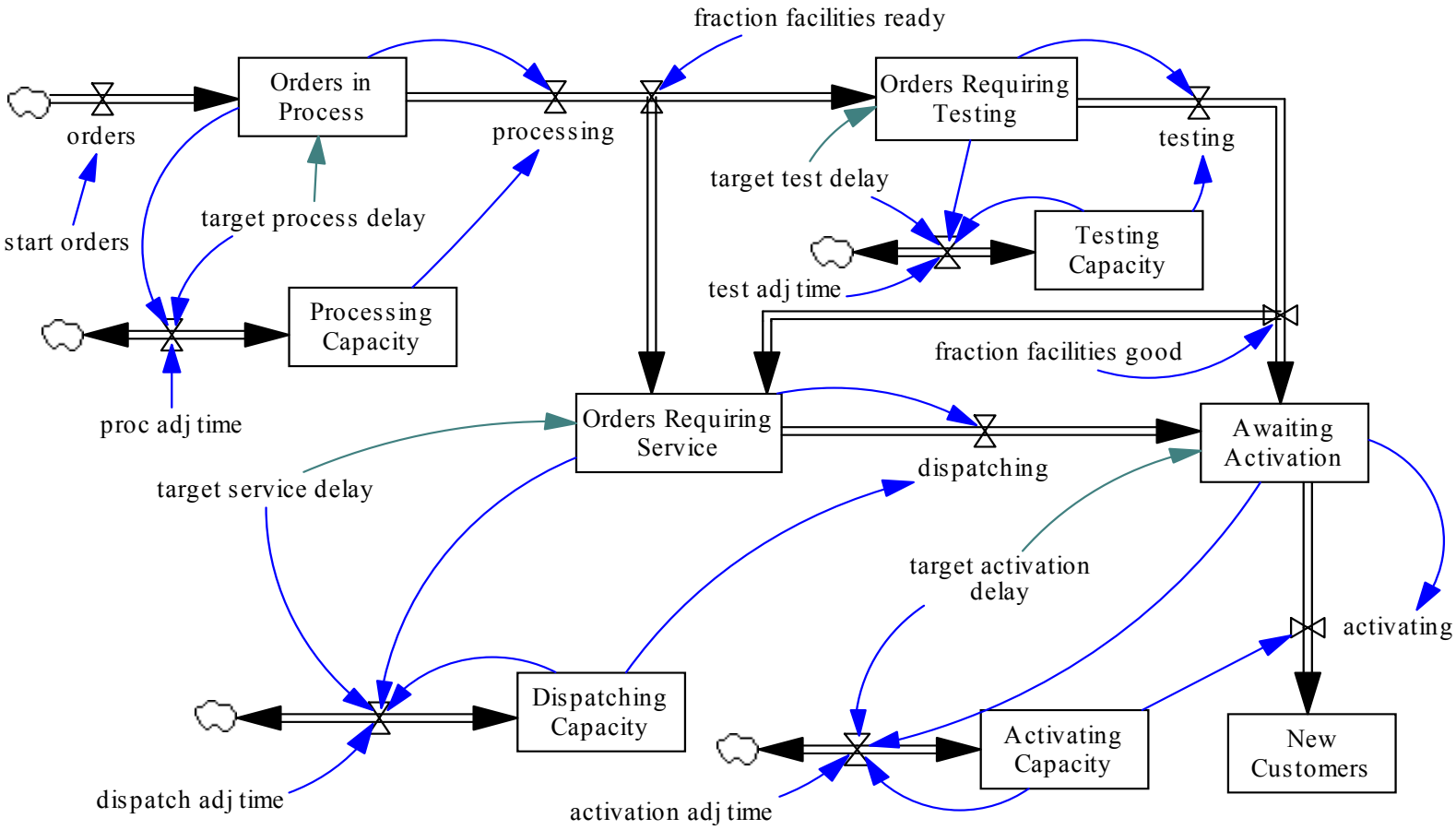
Basic Dynamic Patterns

Indirect Structure Testing Software (SiS)

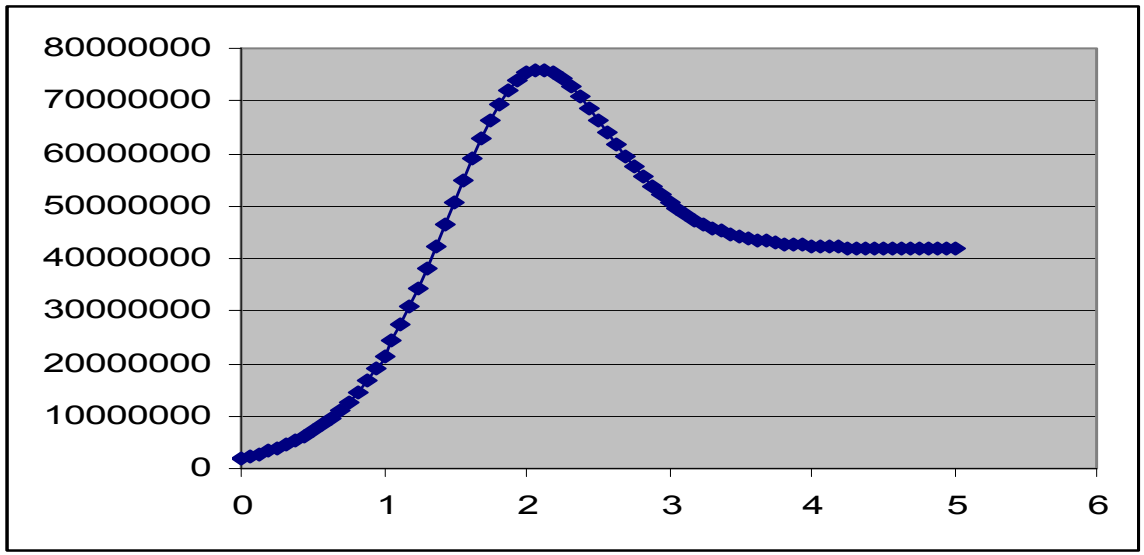
Abbreviation	Description
ZERO	Zero
CONST	Constant
PLINR	Linear with positive slope
NLINR	Linear with negative slope
NEXGR	Negative exponential growth
SSHGR	S-shaped growth
PEXGR	Positive exponential growth
GR1DA	Growth with decreasing rate followed by decline to equilibrium (growth level is less than decline level)
GR1DB	Growth with decreasing rate followed by decline to equilibrium (growth level is greater than decline level)
GR2DA	S-shaped growth and decline to equilibrium (growth level is less than decline level)
GR2DB	S-shaped exponential growth and decline to equilibrium (growth level is greater than decline level)
D1PEG	Decline with increasing rate followed by positive exponential growth
D2PEG	S-shaped decline followed by positive exponential decline
NEXDC	Negative exponential decline
SSHDC	S-shaped decline
PEXDC	Positive exponential decline
D1GRA	Decline with increasing rate followed by growth to equilibrium (decline level is less than growth level)
D1GRB	Decline with decreasing rate followed by decline to equilibrium (growth level is less than decline level)
D2GRA	S-shaped decline and growth to equilibrium (decline level is less than growth level)
D2GRB	S-shaped decline and growth to equilibrium (decline level is greater than growth level)
G1PED	Decline with decreasing rate followed by positive exponential decline
G2PED	S-shaped growth followed by positive exponential decline
OSCCT	Oscillation around constant mean
OSCGR	Oscillation around linearly growing trend
OSCDC	Oscillation around linearly declining trend

List of dynamic behavior pattern classes

Sample Model Used with SiS



Validity Testing with Default Parameters

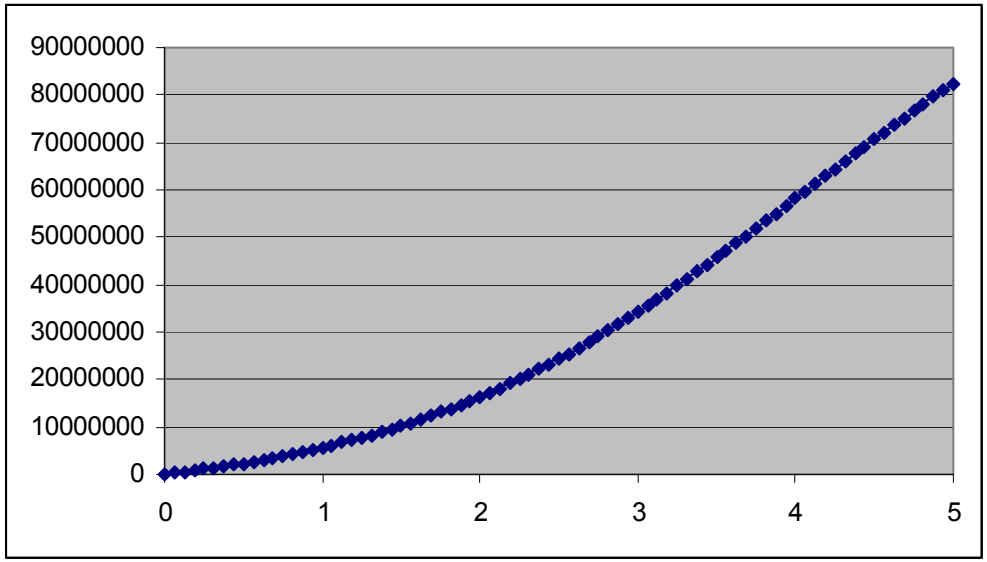


Simulation Output (with default base parameters)

ZERO0	-10	GR1DA	-16,7075	SSHDC	-25,4027	G2PED	-8,20289
CONST	-10	GR1DB	-7,26676	PEXDC	-34,4607	OSCCT	-21,6809
PLINR	-10	GR2DA	-7,08969	D1GRA	-13,9766	OSCGR	-21,6809
NLINR	-10	**GR2DB**	**1,02006**	D1GRB	-10,9594	OSCDC	-10
NEXGR	-15,7189	D1PEG	-11,0744	D2GRA	-9,18981		
SSHGR	-15,8379	D2PEG	-15,9738	D2GRB	-8,76851		
PEXGR	-29,1582	NEXDC	-17,7455	G1PED	-10,4806		

Likelihood Values of simulation behavior correctly classified as the GR2DB pattern

Parameter Calibration with Specified Pattern

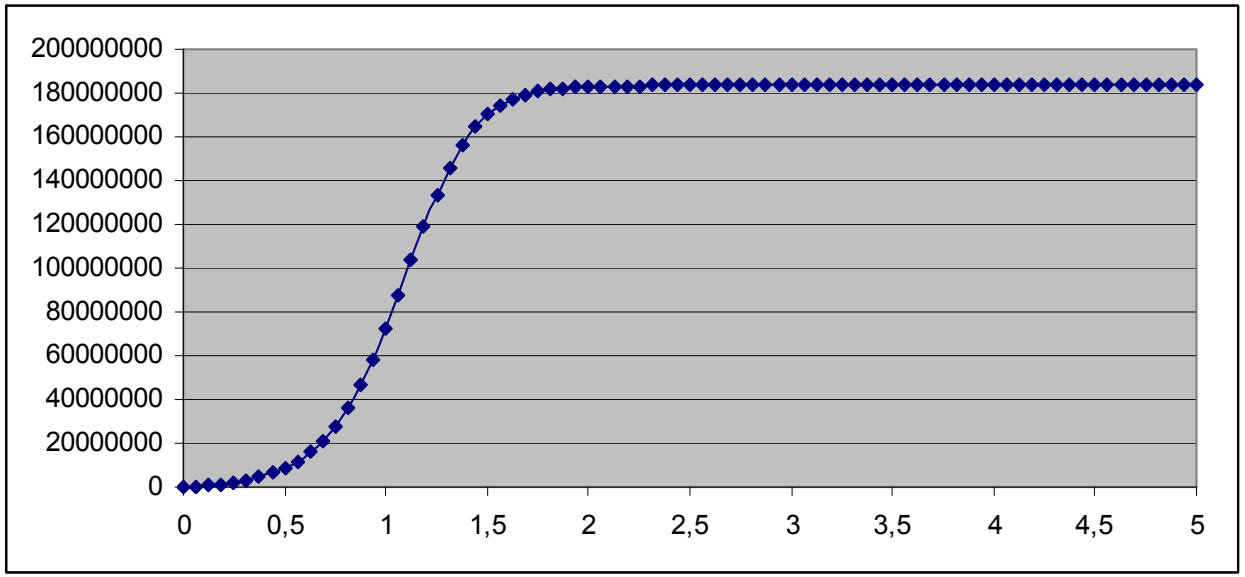


Simulation Output (with base parameters)

Selected Parameters	Min	Max	Number of Values In the Interval
1. advertising effectiveness	0	1	5
2. customer sales effectiveness	0	8	5
3. sales size	1	5	5

The ranges and number of values tried for each parameter

Result of the Parameter Calibration



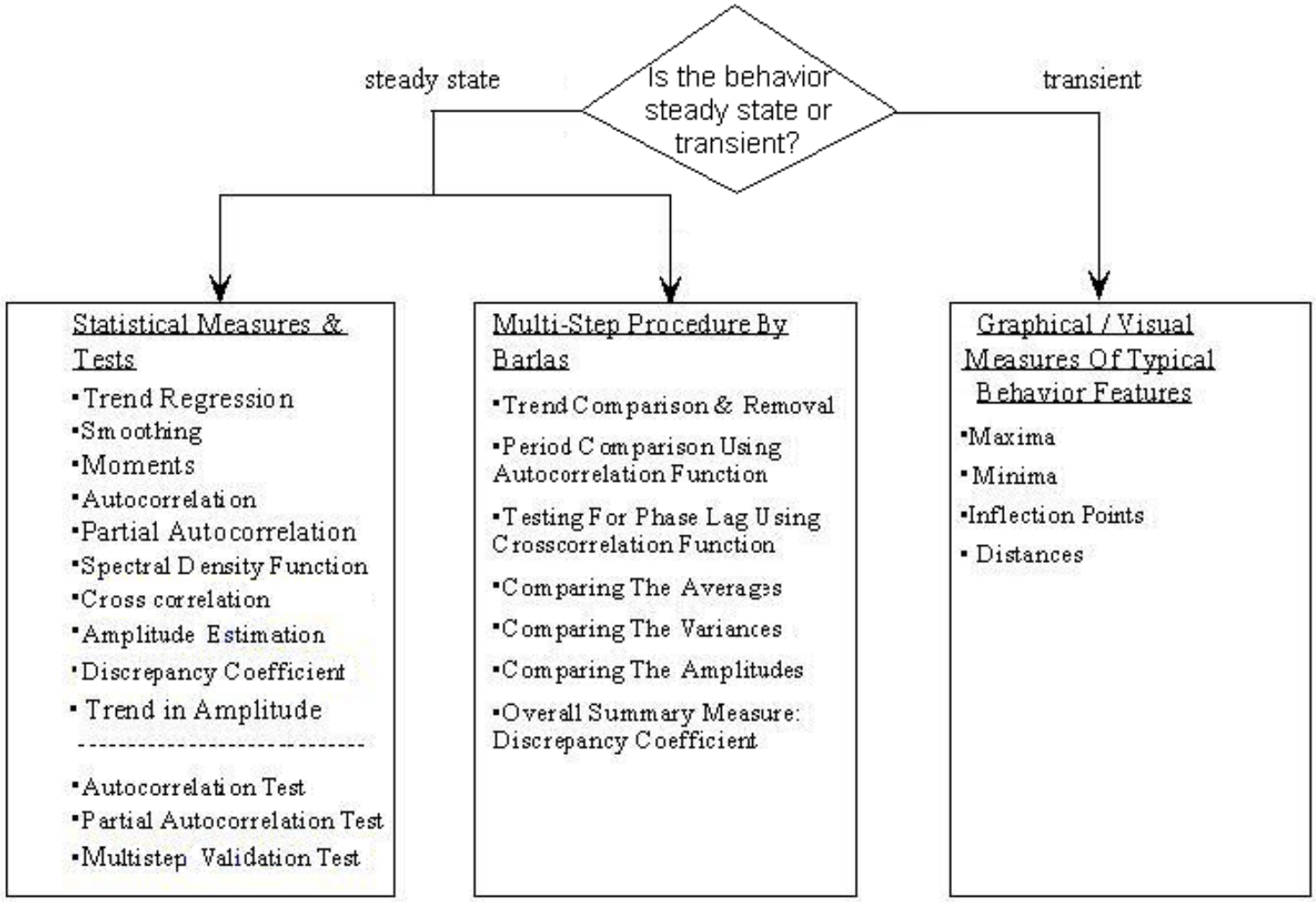
Simulation Output as Desired (after automated parameter calibration)

- Best parameter set is 41
- Best Likelihood Result: 1.2119776136254248
  Best Parameter Set:
- 1. advertising effectiveness: 0.25
- 2. customer sales effectiveness: 6.0
- 3. sales size: 1.0

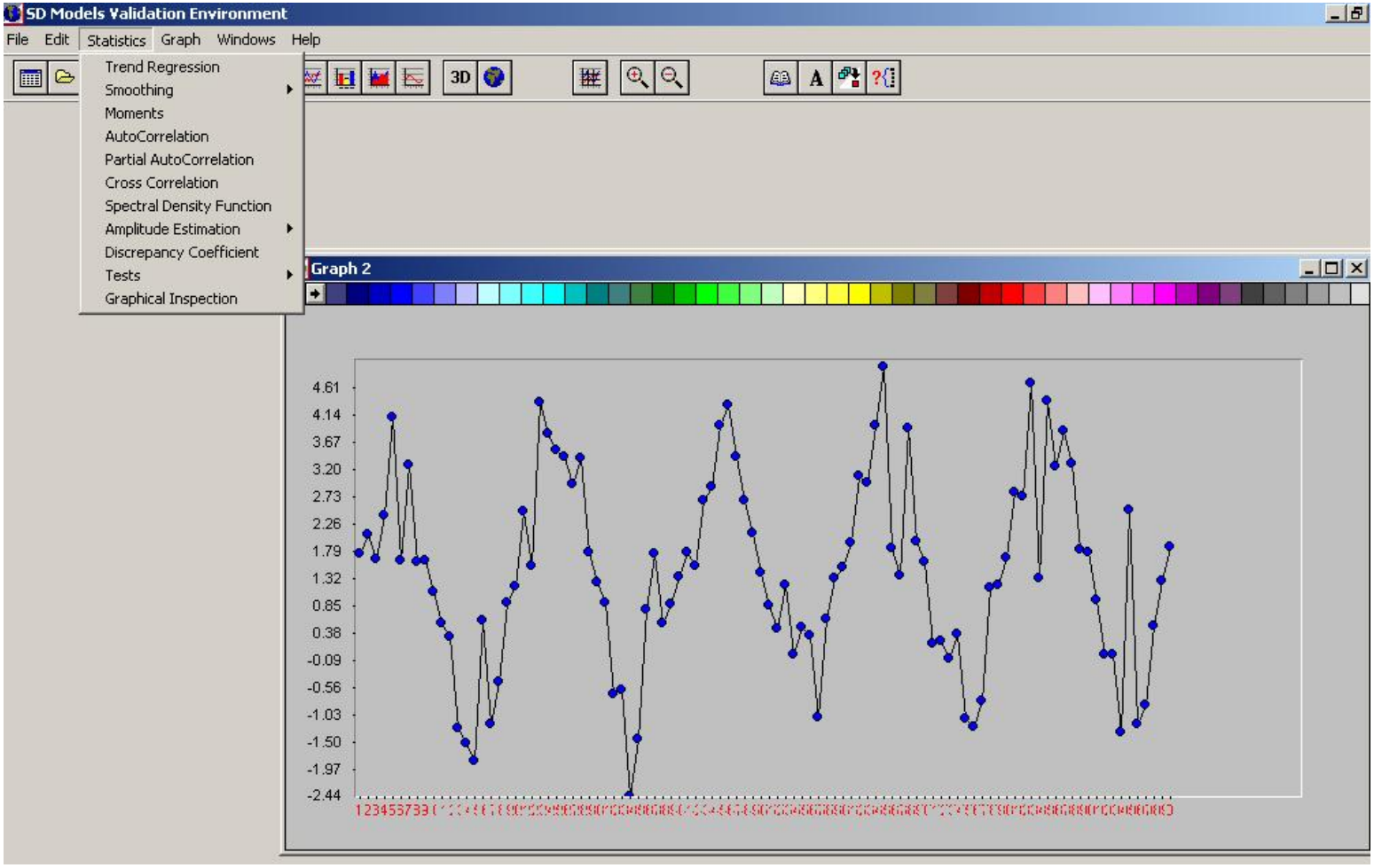
Output Behavior Credibility

- Two types of patterns
  - Steady state
  - Transient
- Major pattern components
  - Trend, periods, amplitudes, ...

Behavior Credibility Testing Software: BTS II

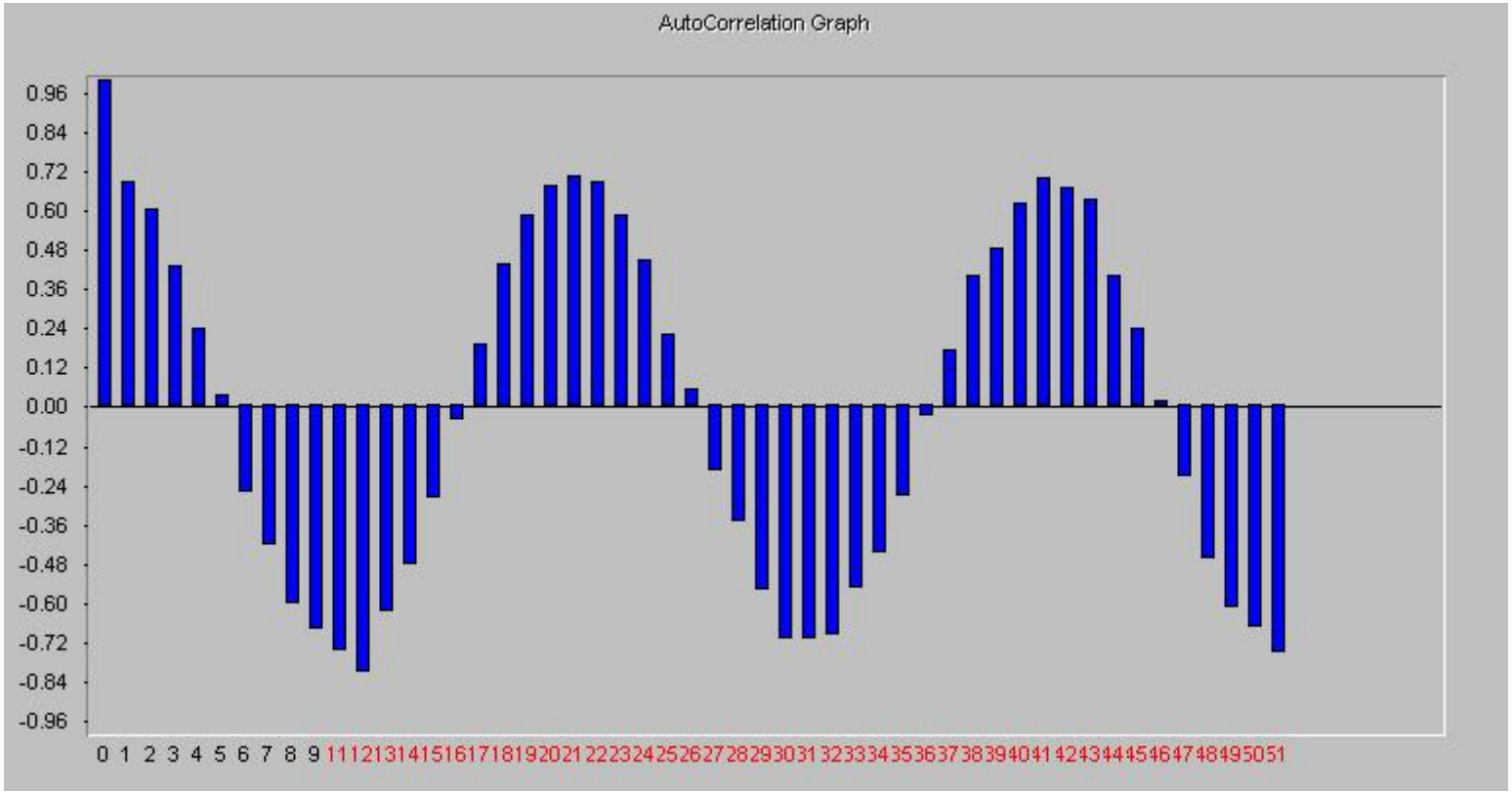


Behavior Validity Testing Software: BTS II



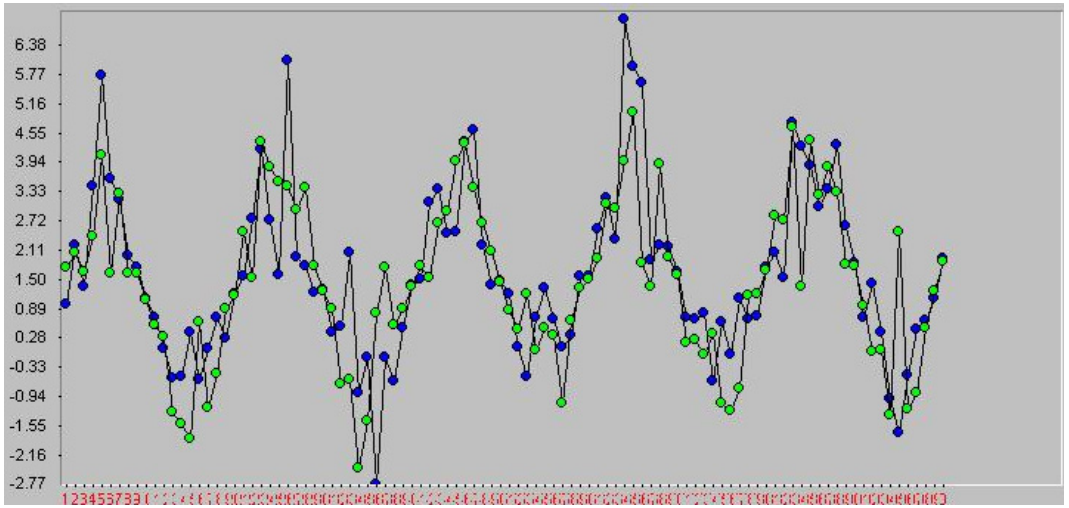
BTS II Tools

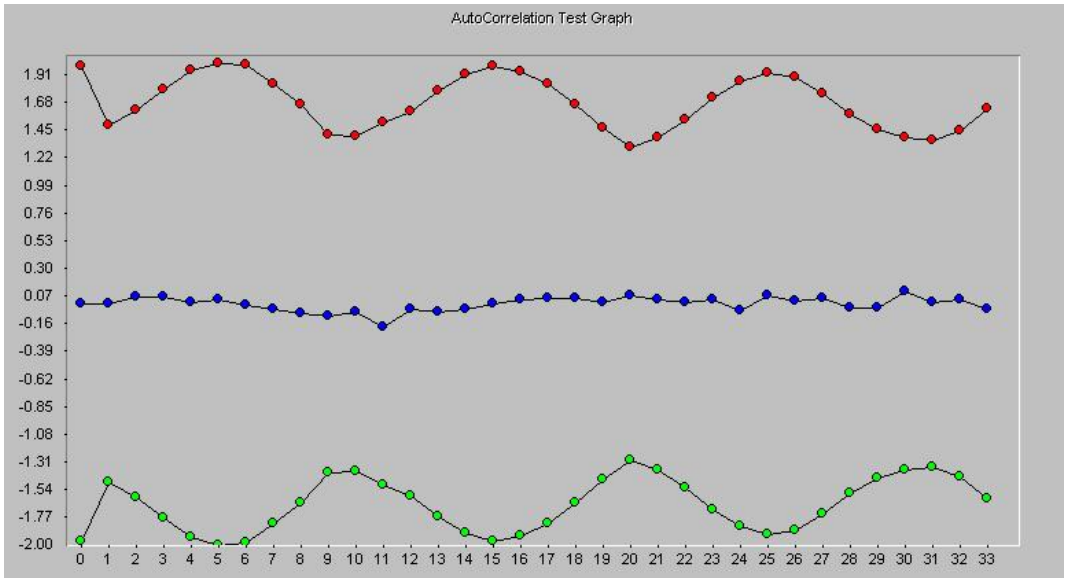
Autocorrelation



BTS II Tools

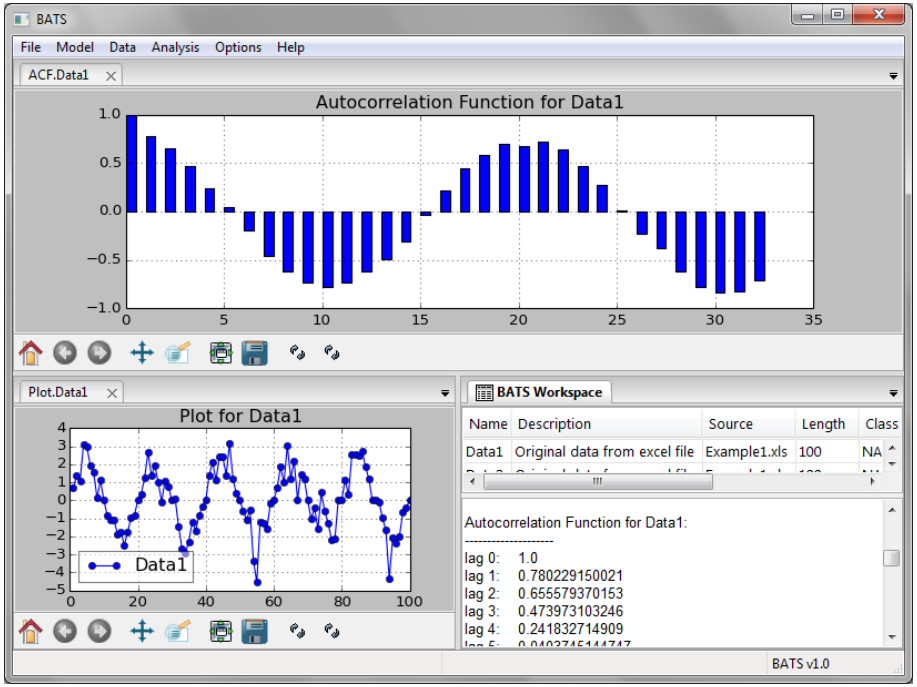
Autocorrelation Test

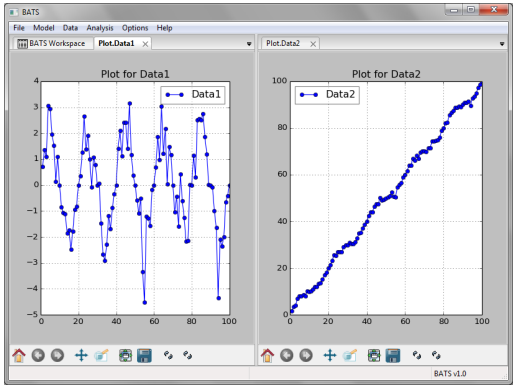


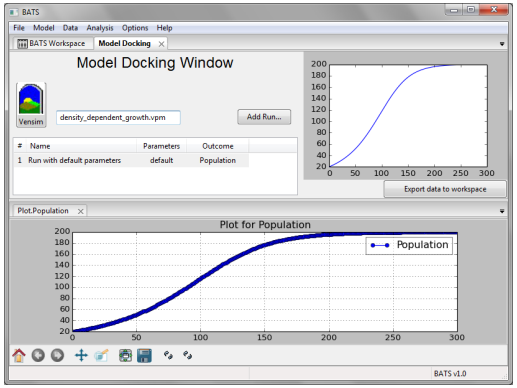


BATS
(Behavior Analysis and Testing Software)

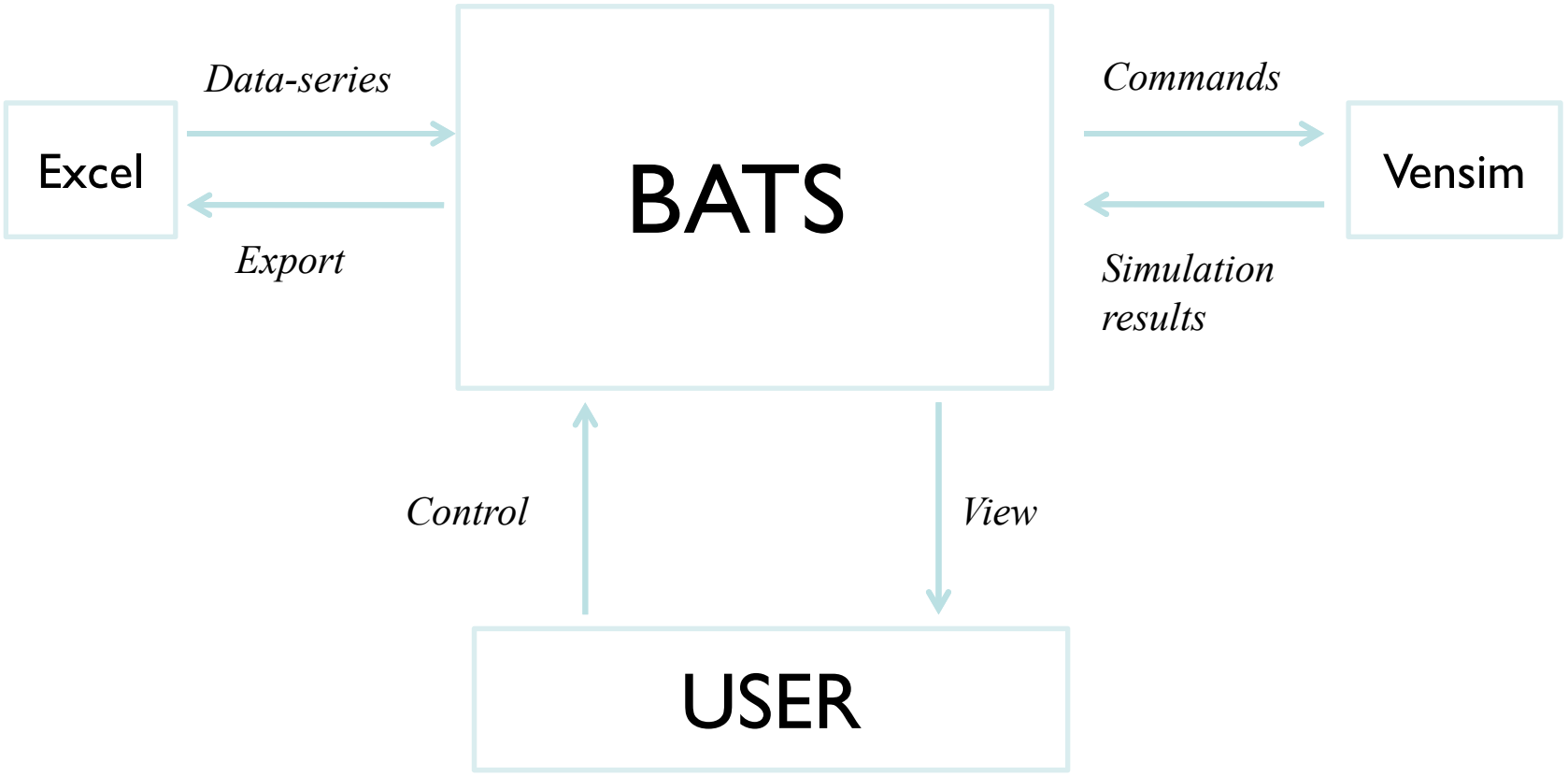
Combines BTS and SiS in single platform







BATS Interfaces



Overview of Features of BATS

Data Importing

- Load From File
- * Model Docking Window
- Draw

Data Preparation

- Split
- Select
- Exponential Smoothing
- Moving Average
- < Trend >

Data Visualization

- Plot

Data Analysis

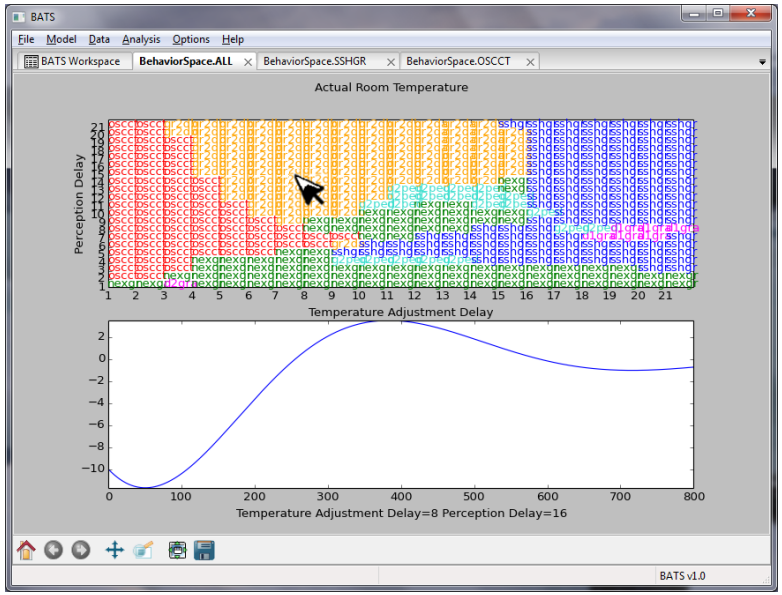
- Classify
- Trend
- Autocorrelation
- Autocorrelation Test
- Spectral Density
- Amplitude Estimation
- Crosscorrelation
- Summary Stats
- * Graphical Comparison

Model Analysis

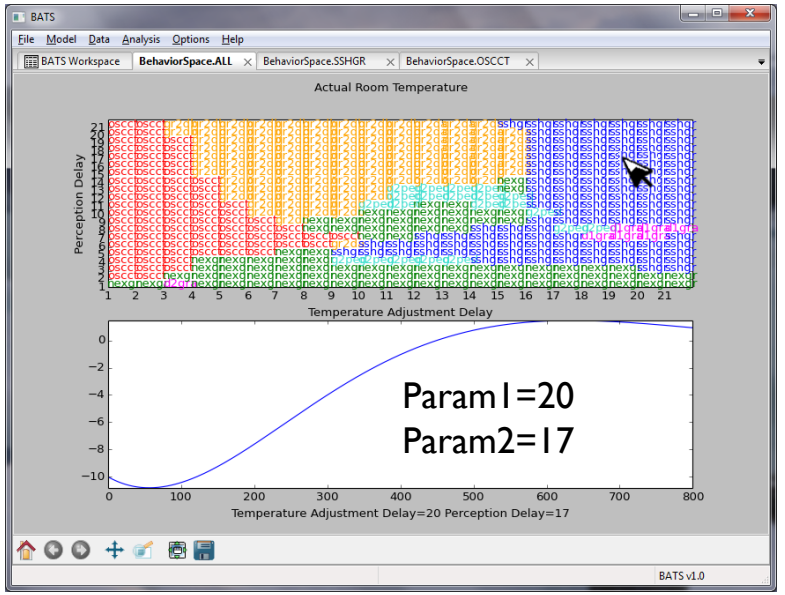
- * Hypothesis Tester
- * Behavior Space Classifier
- * Behavior Class Mapper

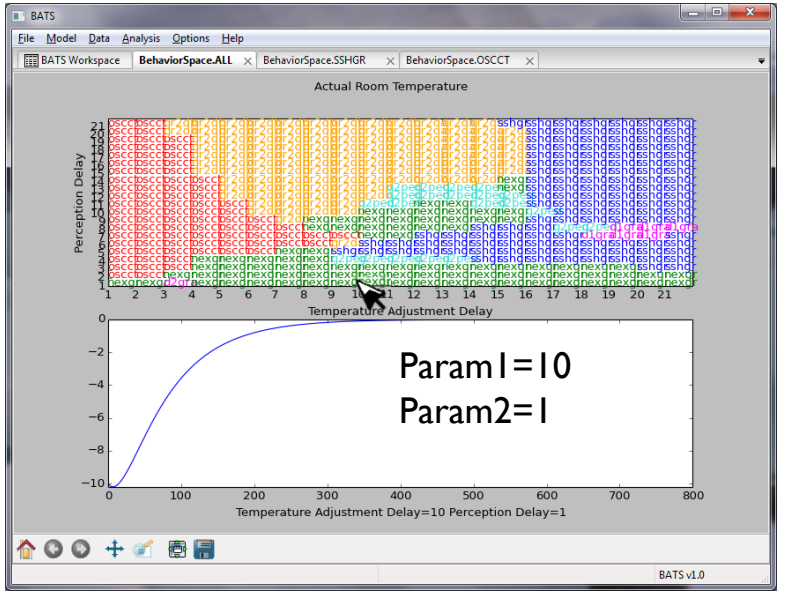
BATS Behavior Space Classifier

(The plot is dynamic)



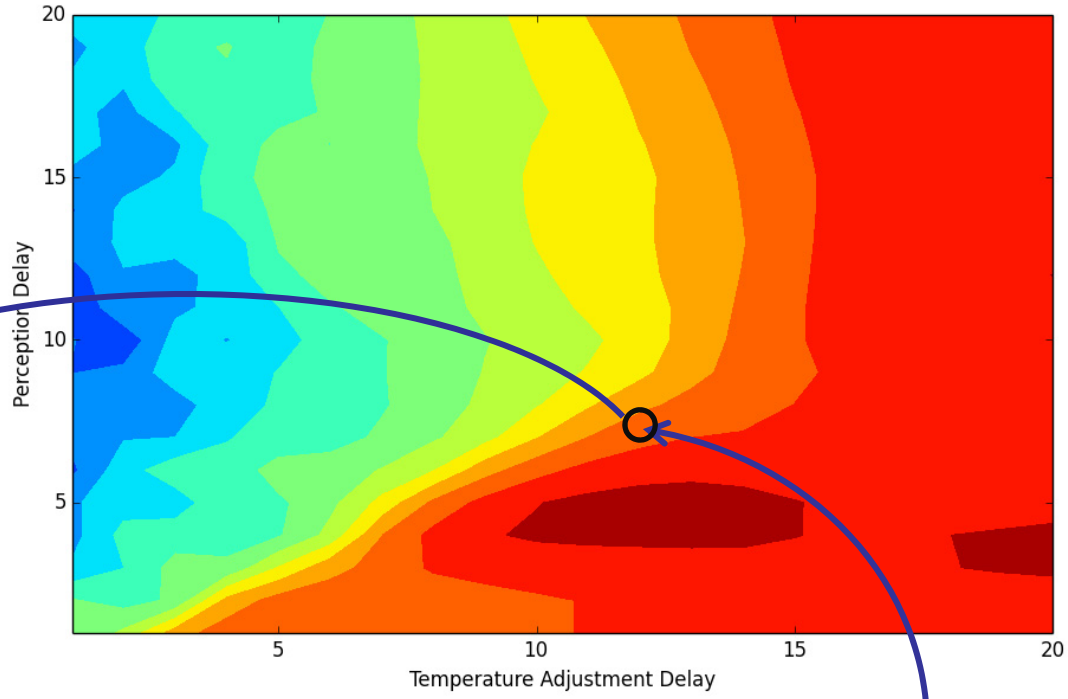
Param1=8
Param2=16





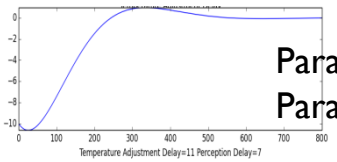
BATS Behavior Class Mapper

Pre-specified class is **sshgr** (s-shaped-growth)



For each point, a simulation run is made

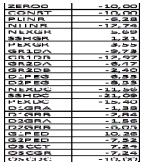
Model output is taken from Vensim



Param1=8
Param2=16

ISTS Algorithm is called

25 likelihoods are obtained



The **likelihood value for pre-specified** class is plotted on the graph

-3.99

Usage Modes of BATS

- Structure-oriented behavior validity testing
  - Extreme condition tests
- Behavior pattern validity testing
  - Barlas' Multi-Step procedure and Graphical Inspection
- Sensitivity analysis
  - Behavior pattern sensitivity with respect to parameter changes
- Model calibration
  - Parameter calibration based on behavior patterns
- Policy design
  - Policy parameter specification based on behavior patterns
- Policy analysis
  - Policy structure specification based on pattern characteristics of model behaviors

Practical Implementation Issues

- More standardized and automated tools
- User friendliness
- Better integration with simulation software
- More standardized validation procedures
- ‘Credibility of Implementation’

  (A justified model means just a reliable laboratory; ‘implementation validity’ does *not* automatically follow at all. It should be taken as project in itself)

Concluding Observations

- Model credibility as a process, rather than an outcome
- Continuous (prolonged) credibility testing
- Model testing, analysis and policy design all integrated
- Credibility of Policy Recommendations
  (Robustness, timing, duration, transition...)
- From validity towards quality
- Quality ‘built-in versus inspected-in’
- Group model building
- Testing by interactive gaming (stakeholders)

References

- Akkermans, H.A., and J.A.M. Vennix. 1997. “Clients' opinions on group model-building: an exploratory study.” System Dynamics Review 13(1): 3-31
- Andersen, D.F. 1980. “How Differences in Analytic Paradigms Can Lead to Differences in Policy Decisions.” In *Elements of the System Dynamics Method*, ed. J. Randers. Portland, OR: Productivity Press
- Andersen, D.F., G.P. Richardson and J.A.M. Vennix. 1997. “Group model building: adding more science to the craft.” System Dynamics Review. 13(2): 187-201
- Ansoff, H.I and D.P. Slevin. 1968. “An Appreciation of Industrial Dynamics.” *Management Science* 14: 383-397.
- Balci O. 1994. Validation, verification and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research* 53: 121–173.
- Balci O, Sargent RG. 1984. A bibliography on the credibility assessment and validation of simulation and mathematical models. *Simuletter* 15: 15 –27
- Barlas Y, Suat Boğ. 2005. Automated dynamic pattern testing, parameter calibration and policy improvement. *Proceedings of international system dynamics conference*. (CD). Albany, NY, USA
- Barlas Y. 2002. ”System Dynamics: Systemic Feedback Modeling for Policy Analysis” in Knowledge for Sustainable Development - An Insight into the Encyclopedia of Life Support Systems Vol.1, UNESCO-Eolss Publishers, Paris, France, Oxford, UK. pp.1131-1175.
- Barlas, Y. 1996. “Formal Aspects of Model Validity and Validation in System Dynamics”, System Dynamics Review, Vol.12, no.3, pp. 183-210.
- Barlas, Y. 1992. Comments on “On the Very Idea of a System Dynamics Model of Kuhnian Science.” *System Dynamics Review* 8(1): 43-47.
- Barlas, Y. 1990. “An Autocorrelation Function Test For Output Validation.” *Simulation* 55(1):7-16.
- 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. 1989b. “Tests of Model Behavior That Can Detect Structural Flaws: Demonstration With Simulation Experiments.” In *Computer-Based Management of Complex Systems: International System Dynamics Conference.* P.M.Milling and E.O.K.Zahn, eds. Berlin: Springer-Verlag.

- Bell, J.A and M.F. Bell. 1980. “System Dynamics and Scientific Method.” In *Elements of the System Dynamics Method*, ed. J. Randers. Portland, OR: Productivity Press
- Carson, E.R.and R.L.Flood. 1990. “Model Validation: Philosophy, Methodology and Examples.” *Trans Inst MC.*12(4): 178-185.
- Coyle, G. And D. Exelby. 2000. “The validation of commercial system dynamics models.” System Dynamics Review 16(1): 27-41
- *European Journal of Operational Research*. 1993. Special Issue on Model Validation. 66(2).
- Forrester J.W. and P.M.Senge. 1980. “Tests For Building Confidence in System Dynamics Models.” In *System Dynamics.* Legasto, A.A., J.W. Forrester and J.M. Lyneis, eds. Amsterdam: North-Holland
- Forrester, J.W., G.W. Low and N.J. Mass.1974. “The Debate on World Dynamics: A Response to Nordhaus.” *Policy Sciences* 5: 169-190.
- 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.
- Forrester, J.W. 1968. “A Response to Ansoff and Slevin.” *Management Science* 14: 601-618.
- Forrester, J.W. 1961. *Industrial Dynamics.* Portland, OR: Productivity Press
- Graham, A.K., J.D.W. Morecroft, P.M. Senge and J.D. Sterman. 1994. “ Model-supported Case Studies in Management Education.” In *Modeling For Learning Organizations.* Morecroft, J.D.W and J.D. Sterman, eds. Portland, OR: Productivity Press
- Kanar K and Y. Barlas. 1999. "A Dynamic Pattern-oriented Test for Model Validation”, Proceedings of 4th Systems Science European Congress, Valencia, Spain, Sept. 1999, pp. 269-286
- Kleijnen, J.P.C. 1995. “Verification and Validation of Simulation Models.” *European Journal of Operational Research* 82: 145-162.
- Lane, D.C. 1995. “The Folding Star: A Comparative Reframing and Extension of Validity Concepts in System Dynamics.” In *Proceedings of International System Dynamics Conference* (Tokyo, Japan). Vol.1: 111-130.
- Lane, D.C. 2001. “Rerum cognoscere causas: Part I - How do the ideas of system dynamics relate to traditional social theories and the voluntarism/determinism debate?” System Dynamics Review 17(2): 97-118
- Lane, D.C. 2001. “Rerum cognoscere causas: Part II - Opportunities generated by the agency/structure debate and suggestions for clarifying the social theoretic position of system dynamics.” System Dynamics Review 17(4): 293-309

- Mass, N. and P.M. Senge. 1980. “Alternative Tests for Selecting Model Variables.” In *Elements of the System Dynamics Method*, ed. J. Randers. Portland, OR: Productivity Press
- Meadows, D. 1989. “Gaming to Implement System Dynamics Models.” In *Computer-Based Management of Complex Systems: International System Dynamics Conference.* P.M.Milling and E.O.K.Zahn, eds. Berlin: Springer-Verlag.
- Meadows, D. H. 1980. “The Unavoidable A Priori.” In *Elements of the System Dynamics Method*, ed. J. Randers. Portland, OR: Productivity Press
- Miller, R.G. 1981. *Simultaneous Statistical Inference.* New York, N.Y: Springer-Verlag
- Mitroff, I. 1969. “Fundamental Issues in the Simulation of Human Behavior.” *Management Science*. 15: 635-649.
- Morecroft, J.D.W. and J.D. Sterman, eds. 1994. *Modeling For Learning Organizations.* Portland, OR: Productivity Press
- Morrison, D.E. and R.E. Henkel, eds. 1970. *The Significance Test Controversy.* Chicago, IL: Aldine Press
- Nordhaus, W.D. 1973. “World Dynamics: Measurement Without Data.” *Economic Journal* 83: 1156-1183.
- Naylor, T.H. and J.M. Finger. 1968. “Verification of Computer Simulation Models.” *Management Science*. 14: 92-101.
- Oral, M. and O. Kettani. 1993. “The Facets of the Modeling and Validation Process in Operations Research.” *European Journal of Operational Research* 66(2): 216-234.
- Ören, TI and L Yilmaz. 2013. ‘Philosophical aspects of modeling and simulation’ in *Ontology, Epistemology, and Teleology for Modeling and Simulation,* Springer Berlin Heidelberg
- Ören, TI. 1981. ‘Concepts and criteria to assess acceptability of simulation studies: a frame of reference’ *Communications of the ACM* 24 (4), 180-189
- Peterson, D.W. and R.L. Eberlein. 1994. “Reality Check: A Bridge Between Systems Thinking and System Dynamics.” *System Dynamics Review* 10(2-3): 159-174
- Radzicki, M.J. 1992. Comments on “On the Very Idea of a System Dynamics Model of Kuhnian Science.” *System Dynamics Review* 8(1): 49-53.
- Radzicki, M.J. 1990. “Methodologia Oeconomiae et Systematis Dynamis.” *System Dynamics Review* 6(2): 123-147.
- Roberts, N., D. Andersen, R. Deal, M. Garet, W. Shaffer. 1983. *Introduction to Computer Simulation: A System Dynamics Approach*. Portland, OR: Productivity Press
- Rouwette, E.A.J.A., J.A.M. Vennix and T. Mullekom. 2002. “Group model building effectiveness: a review of assessment studies.” System Dynamics Review 18(1): 5-45

- Saysel A.K and Y Barlas. “Model simplification and validation with indirect structure validity tests” *System Dynamics Review*, Vol. 22, No.3, 2006; 241-262.
- Sheng G , MS Elzas, TI Ören, BT Cronhjort. 1993. ‘Model validation: a systemic and systematic approach’ *Reliability Engineering & System Safety*, 42(2), pp. 247-259
- Schruben, L.W. 1980. “Establishing the Credibility of Simulations.” *Simulation.* 34(3): 101-105.
- Senge, P.M. 1977. Statistical Estimation of Feedback Models. *Simulation.* 28: 177-184.
- Schlesinger, S. et al. 1979. “Terminology for Model Credibility.” *Simulation.* 32(3):103-104.
- Sterman, J.D. 1992. Comments on “On the Very Idea of a System Dynamics Model of Kuhnian Science.” *System Dynamics Review* 8(1): 35-42.
- Sterman, J.D. 1987. “Testing Behavioral Simulation Models by Direct Experiment.” *Management Science* 33(12): 1572-1592.
- Sterman, J.D. 1985. “The Growth of Knowledge: Testing a Theory of Scientific Revolutions with a Formal Model.” *Technological Forecasting and Social Change* 28(2): 93-122.
- Sterman, J. D. 1984. Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models. *Dynamica.* 10(2):51-66.
- Vennix, J.A.M. 1999. “Group model-building: tackling messy problems.” System Dynamics Review. 15(4): 379-401
- Vennix, J.A.M., H.A. Akkermans and E.A.J.A. Rouwette. 1996. “Group model-building to facilitate organizational change: an exploratory study.” System Dynamics Review. 12(1): 39-58
- Wittenberg, J. 1992. “On the Very Idea of a System Dynamics Model of Kuhnian Science.” *System Dynamics Review* 8(1): 21-33.
- Wright, R.D. 1972. “Validating Dynamic Models: An Evaluation of Tests of Predictive Power.” In *Proceedings of Summer Computer Simulation Conference.* 1286-96.
- Yücel, G. and Y. Barlas. 2011. Automated parameter specification in dynamic feedback models based on behavior pattern features. *System Dynamics Review*, Vol. 27, No.2, pp. 195-215.
- Zellner, A. 1980. Comment on Forrester’s “Information Sources for Modeling the National Economy.” *Journal of the American Statistical Association* 75: 567-569.