

ESTABLISHING THE CREDIBILITY OF A BIOTECH SIMULATION MODEL

Lenrick Johnston
Lee Schruben

Arden Yang

David Zhang

Dept. of Industrial Engineering and
Operations Research
University of California at Berkeley
Berkeley, CA 94704, USA

Genentech Inc.
1 DNA Way, South San Francisco,
CA

Bioproduction Group Inc.
P.O. Box 9135, Berkeley, CA
94704

ABSTRACT

One of the key goals for a simulation model is to accurately replicate the real system under consideration. A protocol is proposed to add credibility to the outputs of a simulation, using a double-blind method. Experimental design is outlined to maximize the value of the information obtained. Finally, experiences implementing the method for a large-scale biotech manufacturing facility are discussed.

1 INTRODUCTION

Model accreditation is becoming an increasingly important part in the design and execution of large scale simulation projects. Accreditation of simulations has three broad goals: first, to design simulation models that are ‘correct’ in some statistical sense; second, to encourage confidence in the model so that it will be used in practice; and third, to directly involve system personnel in the development of the simulation model.

The first goal is the subject of considerable research around verification and validation of simulation models (Sargent 2003, 2004, 2005; Whitner 1989). Such tests aim to debug the logic and code of the simulation (verification) and empirically demonstrate that the simulation performs in a manner analogous to the real system (validation). Sargent (2005) classifies a set of techniques – both subjective and objective – for the validation of simulation models including graphical methods, sensitivity analysis, extreme condition tests and predictive validation. Many of these tests are designed partly to address the second goal of accreditation, i.e. to increase the willingness of persons to base decisions on insights or outputs from the simulation model.

In this paper, we discuss some experiences in implementing a simulation model accreditation protocol suggested by Schruben (1980) that is specifically designed to promote the direct participation of decision makers in the

modeling process. This protocol is suggested here as an important step in any accreditation process, since it holistically examines all aspects of a simulation model including data validity, conceptual model construction, accuracy, and output behavior.

2 PROTOCOL

A simple experiment is proposed to complement verification and validation procedures. This experiment is based on comparing a simulated system to a real system, and assumes that output measurements from the real system can be used as an accurate and complete representation of the physical properties underlying that system.

The goal of the protocol is to increase the credibility of the model to users, where ‘credibility’ is loosely defined as a belief by users that the simulated model is an accurate representation of reality. We define one possible empiric metric for credibility in section 4.

The approach treats both simulated and real systems as ‘black boxes’ with a set of identical, exogenous inputs. Both systems produce a number of outputs, which are shuffled and presented to the user in a standardized format. Where possible the forms used should be those actually used in managing the real system; the numbers on them can be real or simulated. Subject matter experts (SMEs) are asked to identify important the differences that distinguish the outputs and to judge whether the output is ‘real’ or ‘simulated’. Most importantly, each SME is asked why they made each of their decisions.

The results of this qualitative exercise to a statistical battery that Schruben called ‘Turing Tests’ to assess how well the SMEs did as compared to someone who was purely guessing. This is because of a historical connection to a situation described by Turing where a human judge engages in a ‘conversation’ – typically a question and answer game – with a human and a machine without knowing which one has produced which answers. If the judge

cannot reliably tell which is which, the machine is said to be indistinguishable from a human (Turing 1950).

Turing-style tests appear in a number of applications in artificial intelligence and other disciplines, see for example (Harel 2005). The advantage of these tests is that they use human ability for pattern recognition and inherent – often un-communicated – knowledge of the process being simulated. Often such inherent knowledge is also difficult to statistically distinguish by tests such as confidence intervals or moments. For example, an operating rule that a particular operation *A* always directly precedes another operation *B*: such a rule may not have any statistical effect on any operating metric in the system if there is a delay in the execution of *B*, but is immediately obvious to a human observer of the simulation.

The key element in Schruben’s model accreditation protocol is that the SMEs are asked to justify their conclusions. They are asked why they felt they could detect the simulated information.

3 EXPERIMENT DESIGN

A detailed simulation model was created using Bioproduction Group’s proprietary simulation design tool. This tool is based on an event relationship graph modeling paradigm. The model was based off of process flow diagrams, process descriptions and documentation, and SME interviews.

We used the following experimental design to maximize the information obtained from a number of subject matter experts (SMEs), all of whom take the test at the same time. Typical tests involve a vertical cross-section of staff including operators, supervisors and managers: as such it is important to control for ‘group-think’ and other such effects which may cause an SME to change their vote as a response of someone else participating in the test. Multiple rounds of this test may be conducted to iteratively improve the model’s accuracy.

3.1 Reporting Formats

The following considerations should be made when presenting reports to the subject:

1. **Standardized reporting format.** Reports are presented to the SME in the standardized format used by the company for reporting metrics. This avoids confusion about the structure of the output that may cause the subject to guess randomly when they may have been able to distinguish real from simulated output, or unfairly identify a simulated output due to company report formatting specifics.
2. **Localizable information removed.** Outputs are stripped of information which may allow the SME to localize the simulation output to a spe-

cific date/time, to a specific user, or to a specific batch/entity. This avoids historical knowledge of particular events providing additional information to the SME. This may cause the subject to distinguish real from simulated output, when they may otherwise have guessed randomly.

3. **Key performance metrics.** Metrics that report on net throughput or examine key system bottlenecks are more valuable than randomly chosen data since they provide a measure of assurance over the entire simulation model rather than just one aspect of the simulation.

3.2 Phase I: Individual Response

The first phase of testing protocol begins as a double-blind experiment with a collection of reports with data from either the simulation or from the real system. Critically, no one in the room knows whether the data presented are real or simulated. This helps control observer bias and subject expectancy effects.

Subjects are then shown pairs of outputs, one as a simulated document and the other as a genuine document. This is to fulfill the requirements of the protocol.

SMEs are required to write down whether they believe a report is ‘simulated’ (containing simulated information) or ‘real’ (a real data set from historical data). Subjects also write the reasons for their decision, to facilitate later discussion. No discussion is allowed between SMEs during this phase.

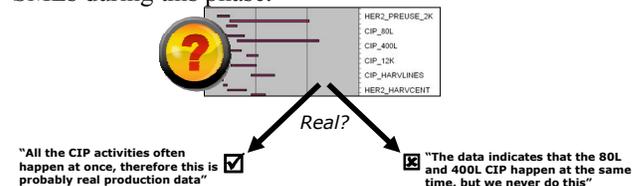


Figure 1: Presenting a report to an SME.

At the end of this phase, answers are collected, copied for reference purposes, and returned to SMEs.

3.3 Phase II: Group Consensus

The second phase of testing is conducted as a group. Subjects are asked to return their responses, which are tallied on the board, and to discuss the rationale for their selections. Finally, the group is asked to provide a consensus on the best possible answer, which is recorded and announced.

The purpose of this phase of testing is to record the best possible answer across the group. This minimizes the effect of experiment bias due to some SMEs not knowing a particular area of system operations and thus contaminating the answer by guessing. Moreover, it identifies the areas that the simulation performs well, and,

most importantly, where the simulation model might require improvement.

3.4 Phase III: Test Comparison

In the third phase, the answers are revealed to the SMEs. All localizable information should be re-incorporated into the outputs at this point, to enable SMEs to isolate the exact time / batch / users involved in the real outputs.

3.5 Additional rounds

The experimenters then analyze the specific reasons the SMEs expressed in their responses and make any needed improvements to the simulation model that were identified. Phases I through III are then repeated with different sets of outputs to iteratively improve the simulation, and the confidence in the model until the SME replies are statistically indistinguishable from guessing. This is the most important aspect of this protocol. However, one needs to address all particular reasons for simulated output identification from Phase II regardless of the mathematical significance in order to gain credibility from the users.

This protocol engages the SMEs within their domains of expertise, using familiar forms and formats. The ultimate model users thus become directly involved in model development, gaining a sense of ownership as model stakeholders.

4 EVALUATING THE RESULTS

The results of a single “black box” test can be evaluated using extensions of the approaches used in Schruben (1980). Here we assume that each individual can detect some subset S of the simulated reports, while the others are classified randomly. The goal is to minimize S; in a completely credible simulation, S is zero. In instances where this is not possible, a Bayesian approach using maximum likelihood estimators is discussed by Schruben (1980).

All of the outcomes of the experiment as seen in Figure 2 are positive for simulation builders. SMEs who correctly distinguish between real and simulated outputs give the opportunity to improve the simulation, while the other case increases simulation credibility.

A number of inferences can be made from different responses, for example if a user says “simulated” to real data, it may be inferred that the users’ knowledge of the domain is poor. We minimize this effect - and maximize the value of the responses obtained - by careful selection of subject matter experts (SMEs) with detailed knowledge of the process under consideration.

In both cases where the user guesses correctly, there may be a chance to improve the simulation design since the user may have been able to correctly identify that the output was / was not simulated. In a perfect simulation (simulated output indistinguishable from real output) even

an experienced domain expert would guess incorrectly half of the time.

		User says:	
		“Simulated”	“Real”
The data is:	Simulated	Chance to improve simulation	Improved Model Credibility
	Real	Improved Model Credibility	Chance to improve simulation

Figure 2: Matrix of possible outcomes to each experimental question

In the experiment conducted, two groups of SMEs were selected corresponding to each part of the facility that the simulation was built upon. These groups consisted plant operators, who used the plant equipment on a day-to-day basis, plant managers, who oversaw plant operations, and engineers, who analyzed data, developed new processes, and helped troubleshoot and improve existing plant issues.

The first group consisted of two plant floor managers, one plant floor technician, and two plant engineers. The second groups consisted of the same two plant engineers and three plant floor technicians.

5 EXPERIMENTAL RESULTS

A detailed simulation of a manufacturing plant for a major biopharmaceutical company was created as part of a larger project. In biopharmaceutical manufacturing, the entire production process is highly regulated and completely documented. No variation from the regulated process is allowed, and automation systems implement actions on the manufacturing floor using a set of pre-programmed rules. This makes a ‘conceptual simulation model’ - i.e. the theories and assumptions underlying the model and its representation of the physical plant – relatively straightforward to validate, but involving a large amount of detailed information. The simulation was implemented exactly matching the documented rules in the automation systems, including logic gates controlling the processes.

Similarly, the use of automation systems provide highly accurate information about timing data and resource utilization for a simulation. Since the systems are highly automated and regulated by governmental agencies such as the FDA, the timing data for production batches is not subject to a lot of error. As such, ensuring data validity is not as significant a problem in the biotech field as it is in some areas.

A serious modeling issue for biotech manufacturing lies in complex process automation. Because of the detail

involved in automating a process, users have expressed doubt that a simulation can accurately represent a facility without modeling at a very fine level of granularity. This is a pattern in a number of highly technical manufacturing industries, where systems are increasingly automated and data-rich – refer Figure 3 for a conceptual data outline. In such an environment, advanced decision support tools (automated production systems, data historians, manufacturing execution systems) actually make the process of validation and verification easier by providing second-by-second reporting of the entire plant operations. This makes the process of logic verification and data validation more simple than classical simulation problems.

An accredited simulation, however, must demonstrate to management that it can successfully replicate such a complex system. The very existence of these advanced manufacturing systems ironically seems to make model accreditation a more difficult task since the automation systems are often seen as a complex ‘black box’. As such simulation credibility becomes of key importance in establishing a model that will be used to make decisions.

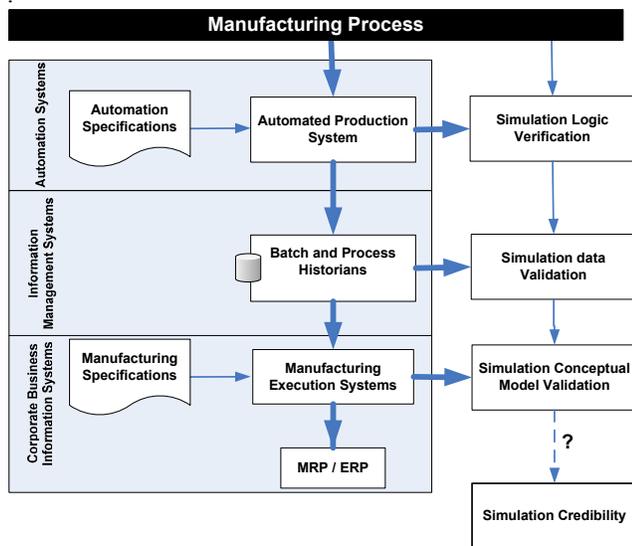


Figure 3: Detailed knowledge of automation systems does not imply a credible simulation

Schruben’s model accreditation protocol was initiated using a vertical cross-section of SMEs. This included technicians from the manufacturing floor, their immediate supervisors, and a set of experts in the process under question. SMEs were given a set of thirteen reports in each round, of which only four were correctly identified by the group in the first round and two in the second round.

Notably, two users when calling out their answers to questions at the start of phase II changed their answer three times to match that of their supervisor or manager. This proved to be a bad choice on a number of occasions where individuals guessed correctly but the group consen-

sus was different. This highlights the need to accurately record the independent answers of SMEs in phase I of the protocol.

After some model modifications as a result of detailed comments from the first round, a second round of experiments was initiated. In both cases the maximum likelihood estimator of the number of simulated documents actually identified was zero, indicating that there was no statistical ability for users to distinguish the simulation from real information. The model has been subsequently used for a number of key investment decisions relating to technology implementation at the plant and enjoys a high degree of credibility due to SME buy-in of the simulation validation protocol. For decision-makers without the knowledge and intimacy of the plant operations, the use of SMEs is imperative for their approval.

6 CONCLUSIONS

The protocol employed here to promote credibility in simulations seems to have a number of key advantages. First, it targets one of the key issues in highly automated plants such as the biopharmaceutical industry: the disbelief that a simulation could successfully emulate such a complex system. In fact, the converse is true since increasing automation inevitably lowers the number of manual operations and increases the likelihood that simulators can replicate the set of automated actions in the plant. Convincing managers and decision makers of this fact however is a difficult task, but a key to ensuring that simulation models are trusted.

Second, all outcomes of the protocol are advantageous to a simulation model builder. If experts are able to tell the difference between the simulation and real information, the model builder then knows and can usually change the specific model assumptions and elements that are causing the problems. Conversely, if experts are unable to distinguish real outputs from simulated ones, then their confidence in the tool increases. Both outcomes are highly valuable in establishing credible simulation models.

REFERENCES

- Balci, O. 1989. How to assess the acceptability and credibility of simulation results. In *Proc. 1989 Winter Simulation Conference.*, eds E. A. MacNair, K. J. Muselman, P. Heidelberger, 62-71. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Herel, D. 2005. A Turing-like test for biological modeling. *Nature Biotechnology*. 23(4): 495-496.
- Law, A. M. and Kalton, W. D. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill

- Sargent, R. G. 2004. Validation and verification of simulation models. In *Proceedings of the 2004 Winter Simulation Conference*, eds R. G. Ingalls, M. D. Rossetti, J. S. Smith, H. A. Peters, 17-28. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Sargent, R. G. 2005. Verification and validation of simulation models. In *Proceedings of the 2004 Winter Simulation Conference*, eds M. E. Kuhl, N. M. Steiger, F. B. Armstrong, J. A. Joines, 130-141. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Schruben, L. W. 1980. Establishing the credibility of simulation models. *Simulation* 34(3): 101-105
- Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind*

AUTHOR BIOGRAPHIES

LENRICK JOHNSTON is a PhD Candidate in Industrial Engineering and Operations Research at the University of California at Berkeley. His research interests focus on large scale decision support systems, including simulation, for biotechnology. Email <rickj@berkeley.edu>

LEE SCHRUBEN is a Professor in the IEOR Department at Berkeley. His PhD is from Yale and his research and teaching interests are on simulation modeling and analysis methodologies.

ARDEN YANG is Principal, Capital Asset Planning & Technology, at Genentech, Inc. He has a PhD in Economics from the University of California at Berkeley and has been Economic Lead for Genentech's Technology of the Future program.

DAVID ZHANG is CFO and a founding member of the Bioproduction Group Inc, a decision support company specializing in Biopharmaceutical production and supply chains. Email <david@bioproductiongroup.com>