Methodological Changes Needed to Meet the World's Grand Challenges[1]

Alan K. Graham, PhD
PA Consulting Group
One Memorial Drive
Cambridge, Mass. 02142
Tel. +1 617 252 0384
Alan.Graham@PAConsulting.com

Abstract

A companion paper to this one identified four "Grand Challenges" for system dynamics. This paper describes the methodological changes seemingly needed to meet those challenges, describing some of the current precedents and rationale for each. The more technical changes are far more explicit model purpose, use of time series, representation of actors' mental models, and working with multiple models. The changes in professional matters are far more soft science expertise, publication in the language of the public and government, incenting synergistic research, and evolving an open-source online curriculum. These changes do not much alter the core of system dynamics but expand the repertoire of expert modeling activities substantially.

Key words: Grand challenge, methodology, time series, curriculum, model purpose, mental models, incentives

1. Introduction: Grand Challenges

Jay Forrester, in his speech to the 2007 International System Dynamics conference, suggested in so many words that the field of System Dynamics is stagnant. It is reasonably verifiable that the system dynamics approach as articulated in 1961 in *Industrial Dynamics* has changed little over the last twenty years. A more optimistic and more actionable view is that the field is following a long wave pattern, of the sort that many basic innovations seem to have shown.[2] In this view, the original system dynamics is at the maturity point in that long wave of innovation, where the seeds of technologies boldly sown forty and fifty years, like computer simulation and knowledge about how to quantify management and consumer decision-making, have matured into a

[1] The views expressed herein are the author's own, and do not necessarily reflect those of PA Consulting Group or its customers around the world.

field with well-established teaching and software infrastructure, and further progress along the same lines lies mostly in very incremental improvements. But even during the period when mature technologies are essentially stagnant, the seeds of the next fifty-year wave of innovation have already been planted, and although they are not yet prominent, they are being worked on.

I have hypothesized elsewhere that many long-wave innovations combine clusters of technologies, each having separately created a modest improvement, but only together creating a quantum step forward.[3] Under this view, what we should be seeking out is the initial studies and the methodological inventions that when integrated will be able to guide effective actions aimed at the major problems of today.

"Grand challenges" is a process used by the Defense Advanced Research Projects Agency (DARPA) to bring important technologies out of (sometimes many) laboratories and into useful application. In a separate paper, I have described four "Grand Challenges" that:

- Impacts the quality of life and even life itself for millions or billions of people through the better or poorer handling of the issue

- Seem to be governed currently by mental models that are seriously in error

- Nearly uniquely, the body of system dynamics methodologies (possibly with methodological extensions) has the means to constructively analyze the problems, design solutions and publicize the results.

The four Grand Challenges I offered are were:

- Insurgency, governance and political stability (which includes corruption, political and economic reforms, and human rights)

- Acting on global warming (starting from the view that the demonstration that the average temperature will keep rising was the easy part of the problem, and much more difficult challenges are now coming to the fore)

- Global financial stability (which subsumes the current worldwide economic meltdown, along with, e.g. the Great Depression, the Japanese "lost decade" and the Asian financial crisis of the late 1990s)

- Harmonious Chinese growth (which history has placed at a point in time when numerous conflicts in many dimensions have become nearly inevitable)

I have watched for several decades system dynamicists solve the conceptually hard part of some major problems in *Urban Dynamics* and *The Persistent Poppy*, and watched the profession nibble around the edges of the problems above.[4] In that time, system dynamics has had two "maybes"—ecology (maybe meeting a grand challenge) and large project management (meeting a challenge that maybe isn't so grand).

I've also watched Pugh-Roberts Associates and its continuation in PA Consulting for over most of its forty-plus years of operation stay more or less at the same size, despite a huge expansion in the number of professors and course offerings around the world. I think we have a problem. Or put another way, we have another set of challenges that are entirely methodological and process-oriented.

I expect Jay Forrester and others are correct in believing that system dynamics can't really penetrate society without penetrating K-12 education, at least as a necessary condition. But it's not sufficient, because we already have a body of system dynamics professionals working, but getting almost no traction in addressing major issues, and it's only cold comfort that, e.g. economics also has almost no traction.[5]

A factoid that's probably more urban legend than expert view: "One definition of insanity is doing the same thing over and over, and expecting something different to result". Practitioners around the world have been doing system dynamics that refines the vision articulated in *Industrial Dynamics*, and practiced for the first few decades and MIT and other academic institutions. Yet the field seems to be in a state of maturity or stagnation. Jay Forrester attributes this state of affairs to practitioners not yet practicing system dynamics with the simple excellence he has so often exhibited. A key passage reads "How often do you see a paper that shows all of the following characteristics", and then lists nine seemingly obvious characteristics, starting with "The paper starts with a clear description of the system shortcoming to be improved".[6] (More about which shortly.)

Let me suggest that Jay's nine criteria are again necessary but not sufficient, for two reasons. First, Jay is way smarter than most of the rest of us, and I know from long experience that some additional process steps and artifacts are useful for the rest of us in trying to meet those nine simple criteria.

Second, the evidence I have from decades of consulting is that when a modeler aspires to give answers that are trustworthy enough to guide action, and doesn't have the luxury of walking away from a problem, one invents[7] practices that differ from classic system dynamics. These are often practices that *Industrial Dynamics* or *Business Dynamics* caution against, but which nonetheless are critical for achieving useful results in the real situations. Many of these inventions will be discussed in what follows.[8]

From consideration of the grand challenges listed above, there are several changes likely to be needed to meet the grand challenges:

2. Changes in practice needed

2.1 Far more explicit treatment of model purpose and context.

Forty eight years ago, Forrester wrote:

> Validity and significance are too often discussed outside the context of model purpose. Usefulness can be judged only in relation to a clear statement of purpose. The goals set the frame for deciding what a model must do.[9]

Unfortunately since then, model purpose has been referenced primarily as the justification for model simplification or validation shortcuts, almost always without Forrester's "clear statement of purpose".

It is arguable (although it will not be argued here) that much of the cool to hostile reception that system dynamics has gotten over the decades from wise and experienced practitioners in other disciplines is due to the lack of clearly stated context and purpose. This is not surprising; given that even in well-regarded system dynamics textbooks, there is usually no called-out example of what a "clearly stated purpose" looks like. And the boilerplate purpose of "to understand the dynamics of X" is almost worse than no purpose statement at all. It says nothing about how to judge the validity and usefulness of the model, giving the appearance of purpose without the substance.

Be that as it may, when addressing large problems like the Grand Challenges, clear statements of purpose are the first and necessary step to allow a community of researchers and teachers to understand how that models will (or won't) complement all of the other extant work. Practitioners need to understand in some compact way what a model is and is not supposed to be useful for.

I have argued elsewhere that real-world validation attempts to find flaws not just one thing (the model), but three things: the model purpose, the model itself, and the model results.[10] Without a validated model purpose, model validation and results validation are inconclusive—no matter how many validation-type tests one performs, there is no basis for concluding that the tests are appropriate or sufficient for the client's purpose.

In PA's consulting work, we often start off an engagement with a one-slide statement of what we understand the model purpose to be, which has four components:

- The strategy / policy levers or actions that the client wants to choose among
- Metrics of success—is it profits? NPV? Employee satisfaction? What defines a successful choice?

- Environmental scenarios—what conditions, outside the control or influence of anything happening inside the system, should the policy choices be evaluated against? (An economic recession is usually a good candidate for the first alternative to the "baseline" assumptions….)

- Puzzles—questions that have to be answered if the clients are to be willing and ready to act. Sometimes this is "hoped-for" or "feared" future behaviors (much like dynamic hypotheses). Sometimes this is an inexplicable pattern of behavior. Sometimes this is an internal management dispute about which of two things is more important.

The point of this exercise is to have discussed in tangible, operational terms, right at the start, what the modeling is for, and how the model is to be used. (More often than not, the discussion reveals disconnects between consultant and client—far better corrected at the beginning than after the modeling is finished!)

After the model purpose slide we develop and review a block diagram and time horizon, which starts to define the scope of the model. Only then come causal diagram buildups and the lead-in to explicit dynamic hypotheses (usually more than one). Only at that later point in learning and discussion is it constructive to articulate (even within the modeling team) dynamic hypotheses that have sufficient content and trustworthiness to go ahead and begin construction of a quantitative model.

2.2 Testing the dynamic hypotheses against explicit time series of historical behavior

System dynamicists all seem to espouse the idea that model purpose should shape everything about a model. Logically, this would include choice of validation tests. Yet system dynamicists, especially academic system dynamicists, seem to regard testing against time series as improper and a waste of time, regardless of model purpose. This attitude is prevalent despite Forrester's constant advocacy of using time series comparisons to test qualitative behavior. This is also despite the many pages of reasoning that *Industrial Dynamics* devotes to understanding the circumstances under which duplicating or predicting point-by-point match to series does or does not support a model's purpose.[11] The flavor of this excerpt is typical:

> [By studying economic systems] we are committed to models in which *every* decision function has, at least in principle a noise or uncertainty component. By definition, the exact time pattern of this noise is unknown, and we have not discovered it generating causes. The model acts on the noise components as it acts on all other flows within the system. The structure and characteristics of the model determine the nature of its reaction to noise.
>
> Forrester 1961, pg. 124

One suspects that the prejudice against time series is a convenient rationale for not making the effort to collect the data, qualitative or quantitative.

Two observations are in order. First, Forrester's discussion was much needed during the 1960s, when the combination of computer-supported regression analysis and the newly intensified quantification of economics seemed to promise the world for fitting models to data. And doubtless he was seeing both novice econometricians and novice system dynamicists saying "the model fits the data, so it must be correct." We are now hopefully all aware of the pitfalls of parameter setting via statistical regression.[12]

Second, the systems in *Industrial Dynamics* happened to be the one situation in which attempting to match time behavior even approximately point-to-point would be actively misleading: An oscillatory system subject to continuous small and unknown random input.[13] Many systems of interest have subsequently proven to have different technical characteristics. Even market dynamics models usually have a few large and known events that are the input that dominates the system's response, which implies that matching time series to an appropriate goodness of fit can be a useful validation test.

Rather than drawing the implication that one shouldn't seek out and use time series (which doesn't serve modeler or stakeholders very well), one would better conclude that understanding the technical character of the system and its context should guide the use of time series. Somehow the "this is inappropriate *when*" message has gotten lost over the years. It is perhaps appropriate for introductory courses to caution beginning modelers against attempting to too-closely fit model behavior to a time series. It is inexcusable to take this position, regardless of the model purpose or competence of the modelers.

It is often asserted by non-practitioners [14] that "curve fitting" is trivial and proves nothing, while referring to the number of parameters compared to the number of data points. This is a transfer of learning from the realm of regression equations, where it is valid, to dynamic systems, where it is not valid.[15] The mathematical framework is completely different. It is quite possible to have a model with many parameters and almost no data points, and still have it be impossible to fit simulated behavior to the data. Empirically, fitting behavior simultaneously to even just a handful of time series is usually quite challenging, and results in better formulations (and data that is both better debugged and better understood).

Be that as it may, Grand Challenge modeling will require different validation testing than groundbreaking models like *Urban Dynamics* or *World Dynamics:*

- Models that guide action on high-stakes questions need to be held to a higher standard of validation than "insight" models, so neglecting time-series comparison for expediency cannot be justified. 40 years of modeling high-stakes problems (although admittedly not as high stakes as the Grand Challenges) and using time series data have convinced us at PA that intelligently fitting to time series really does challenge the model structure (and the data) and yield better models. This same argument, of course, also argues for more intensive use of the numerous other validation tests of system dynamics.

- Grand Challenge models will have a significant fraction of "soft variables", and correspondingly less information about what plausible numerical relationships among them would be. This puts a greater burden on time series to establish parameter values.[16]

- The participants in Grand Challenge systems are more diverse and less accessible than for many "classic" SD applications, so the traditional dependence on firsthand observation will be more problematic than usual. Indeed, some of the Grand Challenge dynamics *depend* on significantly different perceptions among stakeholders.

- Much of the problematic nature of Grand Challenge systems is feedback loops of course, but much is also simply keeping track of the magnitudes. In today's economic debates, much of the rhetoric is asking, but not answering, what the balance is between growing the economy via fiscal stimulus, and ultimately dragging the economy down by increasing the national debt. These are very numerical questions, and they require a model that's reasonably close to numerically correct.

- Part of a Grand Challenge is convincing broad audiences that the work is plausible. John Sterman states the situation nicely:

  > Validation is also intrinsically social. The goal of modeling, and of scientific endeavor more generally, is to build shared understanding that provides insight into the world and helps solve important problems. Modeling is therefore inevitably a process of communications and persuasions among modelers, clients, and other affected parties. Each person ultimately judges the quality and appropriateness of any model using his or her own criteria.
  >
  > (Sterman 2000, pg. 850)

  To this I add that showing behavior that approximates known behavior is probably the fastest available means of establishing validity and starting to convince a wide audience of stakeholders. The academic practitioners of the field have been trying for fifty years to convince stakeholders that they shouldn't be looking at time series, and it's just not working.

In brief, refusing to use data that can disconfirm or confirm models is simply not a tenable stance for "grand challenge" research.

2.3 Working within an ecology of models

Acting on global warming is one challenge that will surely be met by an ecology of models, ranging from simple teaching models to complex political and economic models to still more complex climate models. Likewise, tackling harmonious Chinese growth probably rests on using models of complex financial and energy markets that will be simplified to understand how energy and finances (very much including balance of trade and holding reserves in dollars) interact with the politics of perception and reaction. Modeling economic and financial stability will also likely require multiple models, since models complex and detailed enough to capture the known complexities of financial markets will be far to large for analysis or communicability. For any of the grand challenges then, we would expect more complex models to validate relationships in simpler or more broadly-scoped models.

We should also expect to develop simplified models to allow computation-intensive analysis, e.g. comprehensive behavior and policy sensitivity analysis, potentiation, and data-constrained outcome sensitivity. Some situations will need to model "several moves ahead" thinking, like insurgency or political positioning (see later discussion in this Appendix). For these situations, some form of optimizing search is likely to be needed to understand the behaviors, and such optimization will quickly exceed computational feasibility if a model is too large.

Going in the other direction, analysis of more broadly-scoped models will define the appropriate assumptions about operating environment used by narrower but more detailed models.

In effect showing that multiple validated models are consistent with one another is a type of validation (i.e. failure of any one model to disprove the assumptions of another). And an ecology of models can be validated within multiple appropriate communities to a thoroughness that no single model could approach.

As might be expected, there are already some precedents for simpler models being validated against more complex models.[17] (Eberlein 1984) has worked out a rigorous approach to simplifying dynamic models, and the Reality Check™ feature of Vensim foreshadows software support for automating the manipulations that are required to check or modify one model's assumptions against another's output. But as a practice of using multiple models that is part of a discipline, the practice is in its infancy.

2.4 Mental model-based forward-looking model formulations.

Industrial Dynamics, *Urban Dynamics* and *World Dynamics* have generated significant insights about the operation of the systems in question. But they have not addressed the organizational and political considerations and beliefs that resist change. Insight about the physical system is necessary but not sufficient. To understand what interventions can succeed, we must analyze the complex of beliefs and incentives that hold the present practices in place. Until then, we are attempting to intervene intuitively in a counterintuitive system.

Although history shows that nearly any phenomenon can be represented by the classic levels-rates-auxiliaries formulation, it seems likely that a succinct treatment of political motivations in particular will require new generic structures to deal explicitly with political actor's beliefs and expectations "simulated ahead" to foresee consequences in the future of alternative actions and thus guide present choices, and pick the best choice.

For example, a realistic political model must represent the motivations for leaders to continue suppressing freedom of speech and political dissent, in terms of expected outcomes. Alliances of convenience are another example. A third example would be a country's leadership pretending to have nuclear weapons. There is a school of thought and research within political science that demonstrates that even despots and failed states are making rational decisions, but of a sort that recognizes the whole spectrum of personal, political and economic calculations about the consequences of future actions that govern those actions.[18]

As with political dynamics, understanding the financial markets in crisis situations seems likely to require more explicit representation of mental models and expected consequences of alternative actions. For example, in crisis situations, some market conditions that would normally cause financial institutions to sell and rebalance the market will flip, and inhibit selling and further destabilize the market. Such dynamics are so central to the whole problem that the expectations formation processes that drive them should be modeled explicitly (at least in some models within an ecology of models).

Such "looking ahead" within a simulation would be a form of endogenous optimization (even though it may well turn out to be suboptimization). And simulation ahead of multi-stakeholder actions probably involves elements of the formalisms of agent-based simulation as well, to explicitly represent what goes on in the "I think that he thinks that I think…." situations.

Although there has been research that uses quantitative, repeatable methods to make inferences about simple expectations and decision-making[19], very few people apart perhaps from John Sterman has inferred an explicit quantitative mental model and then used that model within a larger model. Explicitly "forward looking" decision-making has been used successfully in a System Dynamics model, but only in a purely microeconomic setting, and where the system participants knew the system well enough that the model itself could be used to simulate and optimize of future actions (Graham and Godfrey 2005).

2.5 Far more subject matter expertise and rigorous methods for soft variables and their dynamics

System dynamicists have always dealt with the “soft variables” surrounding human decision-making. However, those decisions are usually surrounded by “hard” variables, whose behavior is very knowable at least qualitatively, if it they are not measured and available as time series. *Industrial Dynamics* very much focused on physical events, with human actors behaving in boundedly-rational ways with simple decision mechanisms. *Urban Dynamics* and *World Dynamics* have followed much the same mold of humans making decisions in response to physical surrounds.

In the context of serving a single group of industrial clients, or publishing ground-breaking research, the traditional SD methodology of relying on first-hand observers was quite sufficient for the purposes of those works. It helped a lot that the first-hand observers were observing behaviors that, for the most part, any of the client group or readership would also have observed. It’s a validation to point to commonly-shared knowledge. Moreover, in systems with less-knowable variables scattered amongst harder, more knowable variables, correct model behavior implicitly provided some validation of the formulations for the less-knowable variables—driven by hard variables, they drive hard variables in the right way.

The Grand Challenges, however, involve systems different in character, and research and teaching to a wider audience. With the possible exception of economic and financial stability, all the grand challenge dynamics are predominantly very soft variables, and indeed many of the dynamics revolve around different perceptions of the system. This means that validation by appeal to common knowledge simply doesn’t fly; because the knowledge isn’t common. Worse, this means that the informal methodology of conversation with experts doesn’t necessarily ensure that even the modeler knows what’s going on. Still worse, the need to speak in the language of diverse audiences means that simply appealing to common sense doesn’t speak in a language that experts see as showing knowledge.[20]

This is not an argument for “throwing in some references” when modeling activities finally result in writing and publication. This is an argument for needing to actually use the methods and knowledge of fields heretofor ignored by system dynamics, and subject models to validation tests against knowledge in a far greater diversity of fields than is generally the case.[21] And in practice, subjecting models to, e.g. political science types of validation testing implies actually collaborating in performing research using political science methods.

Generally, system dynamics modelers seem to stay away from building or accessing substantial expertise in the “softer” subjects. Naturally, there are precedents. Modelers have systematically used market research and marketing formulations.[22] Political science sometimes informs dynamic modeling and the empirical research it engenders.[23] Modeling insurgency and political stability have driven study of social science literature and consultation with experts.[24] However, to address these Grand Challenges, the overall tendency for dynamic modelers to associate themselves predominantly with “good” quantitative thinkers (in management science departments and the like) must change.

2.6 Publication in the language of the public and government

If political speeches must use the language complexity of an average American teenager, world-changing publications cannot be limited to academic format and language. Success in understanding and managing actions to mitigate global warming hinges as much on using very different communications strategy for research as it does on the research itself. Likewise, harmonious Chinese growth very much depends on achieving a public understanding of not only economic issues, but economic issues as they play out in a setting very different from that of the developed world. The ideas must be not only just accessible to politicians, industry leaders, students and their teachers. The ideas must be compelling, to motivate the kinds of sacrifices that may be needed to solve any of these grand challenge problems.

Ways need to be found for professors at researcher universities who reach out to the public to be seen as successful within the university community (as they are not, at least within research universities). Well-rounded publication efforts and curriculum creation at below the university level should not impair suitability for tenure and promotion.

One *modus vivendi* already used at least by academic “stars” is participation in think-tanks, which derive funding from other sources, and can hire staff not subject to tenure considerations. That said, the organizational model needs further refinement, as think tanks, although they do very well at gathering qualitative knowledge about complex issues, generally seem disinclined or unable to do ground-breaking quantitative modeling.

Ways need to be found to counter the academic publication incentives to work on very narrow and “academic” topics (as Forrester describes them), as opposed to topics that contribute to real-world issues.

2.7 Incenting research that is important and synergistic

There is competition for funding and visibility that is constructive to a point, but competition can create perverse incentives. Too often, the path to success lies in rejecting all or part of other researcher's work, rather than integrating it with one's own work. Also, communication about research topics is usually via the "next steps" sections of published papers, which tend to be either poorly-defined exhortations that X needs further study (as in "give me funding and I'll fill in the details"), or specification of research areas in which the author already has a marked advantage over other researchers, by virtue of having gathered the data and completed the work reported. These are not ways to expand the number of researchers working productively on a problem.

Promotion criteria incent academic research to stay within a single area, to achieve professional uniqueness. There are no particular incentives to fill in gaps in knowledge, especially if the gaps are persisting in part because they are not amenable to traditional methods. Academic researchers may find it hard to believe, but when measured against the criterion of modeling to solve real-world problems, the academic literature typically has many gaps, to the point where Forrester's original approach of just asking first-hand observers is actually the most efficient strategy.[25]

The need to focus on publication also mitigates against working to solve a real-world problem. Tenure committees and professional standing both incent doing many publications in one particular area. In contrast, addressing real problems often calls for original work in a variety of subject matter and methodological areas.

Special Interest Groups of professional societies like the System Dynamics Society may be a nucleus for more synergistic research activities, given some evolution. Perhaps the next step in evolution is co-authoring papers for publication that specifically describe opportunities for research topics that will directly contribute to a grand challenge. We need some exemplars of papers that not only point the way but give directions.[26] But note that SIG work is vulnerable to the same perverse incentives and hijacking by individuals with established research specialties.

Some science-based government agencies presumably try to counteract these incentives. But there are two strikes against them: The deep experts in a field aren't disseminating high-quality, detailed thinking on what research needs to be done, and promotion committees and academic respectability still wield a might influence over what work gets proposed for funding.

And government agencies most responsible for many facets of the Grand Challenges are staffed with devoted public servants who know almost nothing of either good research or research relevant to their concerns. As a result, government-funded research can easily end up misfocused and fall short, not by a little but by a lot, of the rigor system dynamicists normally associate with good public policy research. I have seen a report to the US Congress on a critical financial regulation question—a headline issue—that seemingly could have been produced by well-funded teenagers, if teenagers could be found with suitably overriding ideological biases.

If these problems are addressed within universities, and it will starts with a much broader definition of what constitutes good teaching and research. If this goes too slowly, it seems likely that some other type of organization—perhaps think tanks with somewhat different behaviors regarding sharing research—will step into the vacuum. Certainly the forty years that Pugh-Roberts and then the SD practice of PA Consulting demonstrate that different organizational forms can achieve modeling of a thoroughness and usefulness far in excess of typical models found in academic publications.

2.8 Open-source online curriculum development

The market has spoken: The world needs more system dynamics training than most universities can support, and the University of Bergen offers wholesale system dynamics training to a large number of retail degree-granting universities. And the WPI distance learning program continues to do very well.

That said, Pål Davidson, Khalid Saeed and Mike Radzicky aren't going to be able to flesh out the kind of curriculum needed for intensive training in modeling in many different subject matter and methodological disciplines, and to an intensity and breadth of experience that only a very few consultants are lucky enough to get. Extensive and efficient SD training is made even more necessary if modelers are supposed to come to a professional-level understanding of Grand Challenge-related professional disciplines as well.

As massive as Sterman's *Business Dynamics* is, there are many more topics, and many more examples and exercises of the same concepts in different application areas. And the control theory and statistical foundations of system dynamics have almost entirely vanished from the extant SD curriculum materials.[27] And Andy Ford's textbook may be the only one dedicated to a specific topic area, environmental issues.[28]

These individuals (and many others) have done prodigious amounts of work to put together those curriculum materials. But relative to the task at hand, the model of the individual contributor creating a new textbook or online course seems to have reached its own limit to growth.

The Internet and Wikipedia offer first glimpses of how an open-source online set of curriculum materials could evolve, including not only software but also social mechanisms for disagreement and recognition—the very soul of scholarly life. Although it may be prudent to impose some restriction on who can contribute, the number of potential contributors is very large (all experienced SD consultants and professors, and anyone they supervise).

In my mind, the best format is yet to be determined. The WPI distance learning format still requires videotaping, and some continuing personal tutoring on an ongoing basis. Back in the days when (the unfortunately named) "programmed learning"[29] was popular, Jay Forrester's *Principles of Systems* was set up for completely self-contained teaching. That was my first exposure to system dynamics at MIT.

The overarching point here is that there is vast potential that is not being taken advantage of, and we will need to if we really want system dynamics to change the world.

3. Conclusion: Expand from the core

After a dozen pages of noting shortcomings in the dominant practices of system dynamics, it needs to be clarified that I do not believe that those core practices are incorrect. Indeed, I believe that the "classic" system dynamics practices are the right way to start out: simple models, whose purpose is improved performance, whose structure is developed from discussions with firsthand observers, and whose baseline behavior and policy outcomes are tested with many different validation tests.[30]

But especially academic instructors that teach introductory courses shouldn't confuse what's taught to beginners with what's needed to tackle Grand Challenges successfully. That beginning "core" simply isn't broad or deep enough.

4. References

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson and Alastair Smith 2002. “Political Institutions, Policy Choice and the Survival of Leaders”. *British Journal of Political Science* **32**, 559-590.

Eberlein, Robert L. 1984. *Simplifying Dynamic Models by Retaining Selected Behavior Modes.* PhD Thesis, Sloan School of Management. Cambridge, Mass.: Massachusetts Institute of Technology.

Ford, Andrew 1999. *Modeling the Environment.* Washington DC: Island Press.

Forrester, Jay W. 1961. *Industrial Dynamics.* Waltham Mass.: Pegasus Communications.

Forrester, Jay W. 1969. *Urban Dynamics*. Waltham Mass.: Pegasus Communications.

Forrester, Jay W. 2007. “System Dynamics—the Next Fifty Years”. In (Sterman 2007), pp. 359-370.

Graham, Alan K. 1977. *Principles on the Relationship between Structure and Behavior in Oscillatory Systems.* MIT Electrical Engineering PhD dissertation. Cambridge, Mass.: Massachusetts Institute of Technology.

Graham, Alan K. 1982. “Software Design: Breaking the Bottleneck”. *IEEE Spectrum* (March), 43-50.

Graham, Alan K. 2002. “On Positioning System Dynamics as an Applied Science of Strategy”. *Proceedings of the 2002 International System Dynamics Conference, Palermo, Italy.* Available through the System Dynamics Society at http://www.albany.edu/cpr/sds/

Graham, Alan K. 2000. “Beyond PM101: 20 Lessons for Managing Large Development Programs”. *Project Management Journal* **31**(4) 7-18.

Graham, Alan. K. 2004. “Strategic Disconnects and How to Bridge Them” *Proceedings of the Strategic Management Society 2004 Annual Conference, San Juan Puerto Rico.* Strategic Management Society http://strategicmanagement.net/.

Graham, Alan K. 2005. “Frontiers of Validation” in *Workshop on System Dynamics Modeling of Physical and Social Systems for National Security*. Chantilly, VA: US Department of Defense. Slides and video available from Worcester Polytechnic Institute System Dynamics Group.

Graham, Alan K. 2009b. "Economic Effects", in Kott, Alexander and Gary Citrenbaum, eds., *Estimating Impact: A Handbook of Computational Methods and Models for Anticipating Economic, Social, Political and Security Effects in International Interventions.* Heidelberg: Springer. (forthcoming)

Graham, Alan K. and Carlos A. Ariza 2003. "Dynamic, hard and strategic questions: Using optimization to answer a marketing resource allocation question". *System Dynamics Review* 19(1), pp. 27-46.

Graham, Alan K. and Jeremy Godfrey 2005. "Achieving Win-Win in a Regulatory Dispute: Managing 3G Competition". *Proceedings of the 2005 International System Dynamics Conference in Boston.* Albany NY: System Dynamics Society.

Graham, Alan K. and Peter M. Senge 1980. "A Long-Wave Hypothesis of Innovation". *Technological Forecasting and Social Change.* **17**: 283-311.

Kadoya, Toshihisa, Tetsuo Sasaki, Satoru Ihara, Elizabeth LaRose, Mark Sanford, Alan K. Graham, Craig A. Stephens and C. Keith Eubanks 2005. "Utilizing System Dynamics Modeling to Examine Impact of Deregulation on Generation Capacity Growth", *Proceedings of the IEEE* **93**(11), pp. 2060-2069.

Levin, G., Gary B. Hirsch, and Edward B. Roberts 1975. *The Persistent Poppy: A Computer-Aided Search for Heroin Policy.* Cambridge MA: Ballinger.

Lofdahl, Corey X. 2002. *Environmental Impacts of Globalization and Trade: A Systems Study.* Cambridge, Massachusetts: MIT Press.

Mayo Donna D., Martin J. Callaghan and William J. Dalton 2001. "Aiming for restructuring success at London Underground". *System Dynamics Review* **17**(3): 261–289

Mensch, Gerhard 1979. *Stalemate in Technology: Innovations Overcome the Depression.* Cambridge, Mass.: Ballinger.

Peterson, David W. 1980. "Statistical Tools for System Dynamics". In Randers, Jorgen, ed. *Elements of the System Dynamics Method.* Portland, Oregon: Productivity Press.

Peterson, David W. and Fred C. Schweppe 1974. "Code for a General Purpose System Identifier and Evaluator: GPSIE". *IEEE Transactions on Automatic Control* (December).

Schmidt, Markus J. and Michael Shayne Gary 2002. "Combining system dynamics and conjoint analysis for strategic decision making with an automotive high-tech SME" *System Dynamics Review* **18**(3): 359–379.

Schweppe, Fred C. 1973. *Uncertain Dynamic Systems*. Engelwood Cliffs, NJ: Prentice-Hall.

Senge, Peter M. 1974a. "Evaluating the Validity of Econometric Methods for Estimation and Testing of Dynamic Systems". Cambridge, Mass.: Massachusetts Institute of Technology Sloan School of Management System Dynamics Group Working Paper D-1944-2.

Senge, Peter M. 1974b. "An Experimental Evaluation of Generalized Least Squares Estimation". Cambridge, Mass.: Massachusetts Institute of Technology Sloan School of Management System Dynamics Group Working Paper D-1944-6.

Senge, Peter M. 1990. *The Fifth Discipline.* New York: Doubleday.

Sterman, John D. 1986. "The Economic Long Wave: Theory and Evidence". *System Dynamics Review* **2**(2): 87-125.

Sterman, John D. 1989a. "Misperceptions of feedback in dynamic decision making". *Organizational Behavior and Human Decision Processes.* **43**(3): 301-335.

Sterman, John D. 1989b. "Modeling Managerial Behavior: Misperceptions of feedback in a dynamic Decision Making Experiment". *Management Science* **35**(3): 321-339.

Sterman, John D., ed. 2007. *Special Issue: Exploring the Next Great Frontier: System Dynamics at 50. System Dynamics Review **23**(2-3): 89-370.*

Van Duijn, J. J. 1983. *The Long Wave in Economic Life.* London: George Allen & Unwin.

5. Endnotes

[2] See, e.g. (Mensch 1979, Graham and Senge 1980, Van Duijn 1983, Sterman 1986)

[3] (Graham 1982), retold in (Senge 1990).

[4] (Forrester 1969) and (Levin, Hirsch and Roberts 1975)

[5] At least economics has achieved the status where politicians at least listen to them before shaping actions toward their own political agenda. By contrast, system dynamics hasn't achieved a body of work about the major problems of the day, let alone the kind of consensus the economics profession has developed.

[6] (Forrester 2007, pg 365)

[7] In the parlance of technology tracking and forecasting, the practices I'll discuss would be called "inventions". Only after a technology becomes widely available and used does it become an "innovation" in this parlance. I don't believe that any of these practices will distort system dynamics beyond recognition, but nonetheless, they represent doing some very different things, or stated more precisely, executing aggressively and at large scale some practices which heretofore have been used only in limited applications and one at a time.

[8] One invention not mentioned elsewhere that deserves mention is using a review and scoring process to draw conclusions from a causal diagram. Acknowledging that the cautions in (Richardson 19XX) and

elsewhere are well-taken, in the hands of experienced quantitative modelers, this process has yielded useful results in time- or budget-constrained situations where quantitative modeling was impossible, e.g. (Mayo, *et al.* 2001). Geoff Coyle raised a firestorm of protest at one system dynamics conference when he suggested that there would be situations in which quantitative modeling would be wasteful or counterproductive. Yet we know those situations happen, and a modeler either has to do the best with what's available, or maintain methodological purity and walk away from the problem. Consultants for hire like PA can't afford this luxury, so we find ourselves in the perplexing condition of being criticized for using methods that are too lax (the diagrammatic scoring exercises) and too data-intensive and rigorous (the large project models are prominent, but many market strategy models are also calibrated against historical behavior).

[9] (Forrester 1961, Section 13.7)

[10] (Graham 2009) and (Graham 2005). One paper (Stephens, Graham and Lyneis 20XX) argues for just two targets of validation (system and outcomes), but that paper addresses itself purely to dispute resolution cases, in which the purpose of the model is usually quite clear.

[11] (Forrester 1961, Section 13.7 and Appendix K)

[12] Textbook expositions of statistical regression usually acknowledge the assumptions underlying the derivations, specifically non-colinearity of inputs and perfect data, and very experienced statisticians look for these problems and can often use *ad hoc* methods to correct for them. (Senge 1974ab) gives synthetic data experiments in a system dynamics context where very modest imperfections in the data give regression results that both appear quite good, and are completely wrong.

[13] Forrester points out, in Section 13.7 of *Industrial Dynamics* that an exercise in fitting model behavior to time series data via weighted least squares can create absurd results, such as a model producing a straight line being superior to a model that produces a slightly different oscillation than the data show. He is quite correct for the class of systems with relatively undamped oscillation responding to ongoing random events that are unknown but have consistent characteristics, for which that Section also gives an example. The proper methodology for dealing with data on such systems was not yet invented when Forrester wrote *Industrial Dynamics.* That method is of course Kalman filtering, and its extension to parameter estimation, described by (Schweppe 1973, Peterson and Schweppe 1974, Peterson 1980). As pointed out in (Graham 2002), for systems subject to large and known disturbances, even oscillatory ones, matching a simple simulation gives approximately the correct results of Kalman filtering.

[14] (Forrester 1961, pg 122) also raises a warning, but with the correct modeling context: "A model built on historical, statistically-derived relationships between variables might reproduce a pattern similar to the actual system. It explains the behavior of the system only if a separate defense is made that the model relationships represent the true causes for system actions", i.e., fit to data is not a substitute but a supplement to "classic" SD validity tests.

[15] The form of regression equations and the statement of the parameter over- or under-determination question are very different from the form of a dynamic model and the corresponding statement of the unique parameter identification problem. Dynamic models implicitly parameterize an enormous number of potential parameters at zero, and insist that the parameters to be varied lie within plausible limits. Also, posing the problem as a dynamic system means that the behavior depends on parameter values in extremely complex and nonlinear ways (as opposed to the simple linearity of the regression case). The large number of parameter constraints, plus the complexity of the mapping from parameter values to behavior in practice means that the computation of whether a system is over- or under-determined is complex to the point of being solvable only numerically—not by a simple count comparison of parameters versus data points.

[16] "Time series" do not need to be "hard" data like economic statistics, or even "hard" data like attitude surveys. Often, expert perceptions of "was high, dipped in the late 1990s, but then went back up in 2001",

if cross-checked, can be very valuable in improving model behavior. PA has used cross-checked qualitative time series extensively in its work on lawsuits and arbitrations for large projects (Graham 2000).

[17] Particularly in market models of network-based industries (mobile phones and electricity), the network portions have been validated against much more detailed models of network operation, in, e.g. (Graham and Godfrey 2005 and Kadoya *et al.* 2005). There is also work going on to link models arising from very different methodologies to address military / political / economic policy planning (Graham 2009).

[18] (Bueno de Mesquita 2002). Political scientists have evolved a concept for political decision-making (which my scholarship skills haven't successfully recovered) that is remarkably parallel to the bounded rationality concept from management science.

[19] Data for expected energy demand and inflation are modeled by relatively simple structures in (Sterman 2000, Section 16.2 and 16.4) respectively. Decision-making in a laboratory / gaming setting is rigorously quantified in (Sterman 1989ab).

[20] Forrester's works are not a counterexample. Forrester's works (*Industrial Dynamics, Urban Dynamics* and *World Dynamics)* are ground-breaking, in the sense that they for the *first time* address a set of issues in a holistic and fact-based way. But, to continue the metaphor, breaking ground isn't the same as nourishing crops, harvesting them, and putting food on the table. Breaking ground needs one set of tools (and expertise judged by one set of standards), and delivering nutrition requires a different and more extensive set of tools, and expertise will be judged by several additional and different sets of standards. Fully meeting a Grand Challenge requires all those additional tools and standards.

[21] I am thinking of political science in particular, where empirical studies shed light and test hypotheses with a precision impossible within traditional system dynamics, for example on dissident groups turning to terror tactics, or democracies deciding to wage war (Bueno de Mesquita 2002).

[22] (Schmidt and Gary 2002) uses results of conjoint analysis to parameterize the pivotal market share portion of a market dynamics model. Market research often informs PA market models, e.g. (Graham and Ariza 2003).

[23] (Lofdahl 2002) is a good example of research that integrates fact-finding, political science theory and system dynamics modeling. Doubtless there are other examples, but this author's roots in political science are shallow indeed.

[24] All of PA's extensive work in this area is For Official Use Only by US government agencies or client-sensitive.

[25] The best worked-out example is documented in the MIT System Dynamics Working Papers, when Gilbert Low attempted to build a model strictly from the economics literature that would account for the 3-7 year business cycle. At several points, when commonsense knowledge said that a relationship had to exist, a determined search found no studies that quantified those relationships. I've had similar experiences trying to find academic support in marketing, immunology and endocrinology, and economics.

[26] The 50-year Special Issue of the *System Dynamics Review* (Sterman, 2007) contains several papers that indeed point the way, although they are not coupled to this decade's major issues, and they are generally far from specific about the research topics that would contribute. Jay Forrester's (2007) paper falls in somewhat a different category, as it ascribes most of the current stagnation of the field to low standards in execution of traditional system dynamics, and identifies some of the institutional shortcomings that are the environment for doing work at less than high standards. Although Forrester identifies some institutional changes that would hypothetically deliver more high-quality work, at the end of the day (and the special issue) the infrastructure of mechanics and exemplars for doing higher-quality work aren't yet in place.

And researchers assisting each other in focusing on doing research that is important to the field and to society is a key piece of that infrastructure.

[27] The discussions of time series, for example, are informed by the mathematics of probabilistic dynamic system and Kalman filtering, which build on standard state-space dynamics. MIT formerly offered a version of Principles of Systems II in which students were taught not just about the concepts of phase and gain, but how to compute them in open-loop and closed-loop systems, along with other basic control theory concepts like transition matrices, and work with them enough to be useful in understanding the link between structure and behavior for oscillatory systems (Graham 1977). These subjects can and have been made reasonably intuitive, but with reasonably hard work expected from students to systematically build that mathematical intuition.

[28] (Ford 1999)

[29] The "programming" involved was to break the knowledge or skill into very small elements, and give the student problems of a paragraph or less, and multiple choice answers. Every answer would take the student either to the next bit if the answer was correct, or an explanation of why the answer was incorrect, and another similar problem to try. The idea is feedback and correction within seconds, which is faster than a tutor can give feedback. As a pre-teenager, I took a year of English grammar in that format, and found it painless and if not actually fun, at least satisfying—who doesn't like to be told that they're right 30-40 times per hour?

[30] That isn't to say I necessarily accept the traditional pedagogy of teaching first about the elements of models and working up to solving problems with validated models and results at the end. A completely different approach would be to start with real problems and MBA-ish sorts of spreadsheet models, and frame the situation as an hypothesis test: Is this a good model? If not, here are some modifications to make the model more realistic. And here are the properties of those additions (like delays). This "spiral" approach has the advantage of always keeping model purpose at the front of every lesson, where it belongs. (Graham 2004) gives a typology of the disconnects in the normal strategy modeling processes that need to be bridged.

