Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation & Uncertainty Quantification
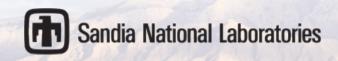
Workshop Footprint Document
June 22nd – 23rd, 2016
Workshop Hosted by Sandia National Laboratories
Albuquerque, NM

Jeff Tsao, Tim Trucano, Steve Kleban, Asmeret Naugle, Steve Verzi, Laura Swiler, Curtis Johnson, Mark A. Smith, Tatiana Flanagan, Eric Vugrin, Kasimir Gabert, Matt Lave, Wei Chen, Daniel DeLaurentis, Alfred Hubler, Bill Oberkampf

ⓜ Sandia National Laboratories

Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation & Uncertainty Quantification

Jeff Tsao (1120), Tim Trucano (1400), Steve Kleban (6132), Asmeret Naugle (1463), Steve Verzi (6132), Laura Swiler (1441), Curtis Johnson (5635), Mark A. Smith (6133),Tatiana Flanagan (6921), Eric Vugrin (6921), Kasimir Gabert (5638), Matt Lave (6112), Wei Chen (Northwestern University), Daniel DeLaurentis (Purdue University), Alfred Hubler (University of Illinois at Urbana Champagne), Bill Oberkampf (WLO Consulting)

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MSXXXX

Abstract

This report contains the written footprint of a Sandia-hosted workshop held in Albuquerque, New Mexico, June 22-23, 2016 on "Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification," as well as of pre-work that fed into the workshop. The workshop's intent was to explore and begin articulating research opportunities at the intersection between two important Sandia communities: the complex systems (CS) modeling community, and the verification, validation and uncertainty quantification (VVUQ) community The overarching research opportunity (and challenge) that we ultimately hope to address is: how can we quantify the credibility of knowledge gained from complex systems models, knowledge that is often incomplete and interim, but will nonetheless be used, sometimes in real-time, by decision makers?

ACKNOWLEDGMENTS

CONTENTS

6

NOMENCLATURE

CS   Complex Systems
VVUQ   Verification, Validation and Uncertainty Quantification
UQ   Uncertainty Quantification
VoI   Value of Information
V&V   Verification and Validation

1. WORKSHOP OVERVIEW

This report contains the written footprint of a Sandia-hosted workshop held in Albuquerque, New Mexico, June 22-23, 2016 on "Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification," as well as of pre-work that fed into the workshop. The workshop's intent was to explore and begin articulating research opportunities at the intersection between two important Sandia communities.

The complex systems (CS) modeling community is gaining importance at Sandia, in response to the importance of a complex systems perspective to addressing many of our nation's pressing national security challenges. The "Resiliency in Complex Systems" Research Challenge has been instrumental in providing an intellectual focus for that community, but has also identified challenges for the future. One particularly important challenge is: how can we quantify the credibility of knowledge gained from complex systems models, knowledge that is often incomplete and interim, but will nonetheless be used, sometimes in real-time, by decision makers? In other words, how can Sandia be an "honest broker" – not only providing for decision makers (a) knowledge derived directly from models, but (b) *meta-knowledge* associated with the degree to which the models are credible and the degree to which decision makers should have confidence in them.

The verification, validation and uncertainty quantification (VVUQ) community has a longer history at Sandia. It is highly regarded for developing and applying sophisticated mathematical and statistical approaches to the VVUQ of the engineering and physics-based models associated with key nuclear weapons components. These models, generally a set of physics-based differential equations with clearly defined quantities of interest, are non-trivial, often characterized by large numbers of elements and parameters -- but only some are uncertain and observational data can often be obtained to narrow their uncertainty. A key question and research opportunity for this community is: can these mathematical and statistical approaches be extended to complex systems models with their much greater uncertainties and "complexity"; or do entirely new approaches need to be developed and what form would these take?

In the six months leading up to the workshop, the CS and VVUQ communities engaged in a fair amount of pre-work, coordinated by a core group of CS staff: Asmeret Naugle (Cognitive Sciences and Systems Department 1463), Mark Smith (System Readiness and Sustainment Technologies Department 6133), Steve Verzi (Systems Research, Analysis and Applications Department 6132), and Tatiana Flanagan (Resilience and Regulatory Effects Department 6921). A small mini-workshop was held March 28, 2016 that widened the engagement to a larger group of CS staff, and especially to a larger group of VVUQ staff. A few of the problem-space framings that emerged from this pre-work and that we brought into the main workshop were:

- Our scope of interest is not limited to models of systems that are "classically complex" in the sense of adaptive, or organized, complexity; it includes models that are not complex in that way, but nonetheless present challenges to current approaches to VVUQ. Our scope of interest is thus operational and somewhat tautological: we are interested in developing VVUQ approaches for those models to which current VVUQ approaches are difficult to apply.
- What exactly we mean by VVUQ is yet to be defined. It certainly includes the possibility of extension of current VVUQ approaches, but it also includes the possibility of developing brand new approaches. Indeed, we should not be afraid to move to new

terminology if we need to, such as the "credibility" of a model, or the "confidence" that a decision maker should have in a model. But quantification is still key, whatever the terminology.

- Our own CS community works across a wide range of domain areas, hence for the purpose of moving forward we are most interested in approaches to VVUQ or "credibility/confidence" that cut across domains, and are to some extent domain independent. In other words, we are interested in abstracting out general features of CS models, and understanding how to tackle these so that what we learn will be applicable to all domain areas.
- Models don't exist in a vacuum; they exist to facilitate decisions that will be made by a decision maker. Thus, there is a model/decision pair that must usually be considered in tandem, in some cases leading to a desire to "bring the decision maker into the model."
- Models are not cost-free, and modelers routinely weigh the cost/benefit ratio of various modeling approaches. VVUQ is also not cost-free, and understanding how to weigh the cost/benefit ratio of various VVUQ approaches will be equally important.

The workshop itself ran for one and two-thirds days.

Day One emphasized divergent thinking: generating and articulating potential Research Ideas for the workshop to consider and discuss. We opened Day One with five plenary talks, each followed by extensive Q&A and discussions. Laura Swiler (Optimization and UQ Department 1441) talked about current VVUQ approaches as applied to engineering and physics-based systems, with thoughts about how these approaches might be extended to complex systems. Steve Verzi (Systems Research, Analysis and Applications Department 6132) talked about the characteristics of models for which current VVUQ approaches are more difficult. To make the discussion more concrete, Matt Lave (Photovoltaic and Distributed Systems Department 6112) and Kasimir Gabert (Cyber Initiatives Department 5638) talked about examples of models in two domain areas of special interest to Sandia: the electric grid, and cyber. Finally, Curtis Johnson (Analytics and Cryptography Department 5635) talked about the extreme of adaptive complexity, and some of the special challenges but also special advantages, that those systems might bring.

After these plenary talks, the workshop broke out into three parallel track sessions. These track sessions gave participants the opportunity to hear and discuss the ~18 "seed" Research Ideas that participants had submitted in the month leading up to the workshop. The three sessions were: Concepts in Complexity; New VVUQ; and Decision Making and Applications. At the end of these track sessions, all participants articulated a few "second-pass" Research Ideas. Tim Trucano (Computing Research Center 1400) binned these ~72 ideas into four potential Research Areas, which were the focus of our Day Two discussions.

Day Two emphasized convergent thinking: in-depth discussions of the potential Research Areas that emerged from the binning of the "second-pass" Research Ideas, as articulated and shaped by Tim Trucano. These Research Areas were:

- Value of Information (VoI). This potential Research Area considers information as the means by which the three "entities" – the model, the real world, and the decision maker using the model – communicate and calibrate each other. Information from the real world calibrates the model; information from and about the model calibrates the credibility of

the model; and joint information about the decision maker and the model calibrates the confidence that the decision maker might have in the decision that is being made based on the model. And information is broadly construed to encompass: "hard" data in engineering and physics-based units as well as seemingly-difficult-to-quantify "soft" human-centric data; meta-information about the information as well as the information itself; and information over all time scales, including, in the case of systems that are adaptively complex, information about what the system has adapted to in its historical past. How might one quantify VoI for these various types of information in the various "impedance matching" (or mismatching) roles they play between model/world/decision-maker? Or, for example, how might one piece of information's VoI affect another's: what is the role of diversity of information in the design of a model that optimizes for "highest credibility at lowest cost"?

- Strong Tests. This potential Research Area asks whether strong tests, those that are used in current VVUQ approaches to engineering and physics-based models, can be extended to complex systems models. Perhaps they could take advantage of the continuing revolution in data mining/analytics and machine learning; perhaps they could be extended to include qualitative dimensions; perhaps they could be applied selectively and artfully to *pieces* of hierarchically modular complex systems models; and perhaps even if incomplete they could be part of a larger "body of evidence" that quantifies credibility of a complex systems model?

- Risk. This potential Research Area considers that, between understanding the rewards versus understanding the risks of particular decisions, it is often risk that is elevated to the fore. Thus, defining, understanding and managing a decision-maker's risk is central, and must "roll up" the various uncertainties and risks associated with all three entities discussed above – the model, the real world, and the decision makers themselves.

- VVUQ-Informed Model Design. This potential Research Area recognizes the inseparability of the model and its credibility, not only at the macro level of the decision maker who might use knowledge from the model, but at the micro level of the approaches used to model and to quantify the model's credibility. In other words, can we incorporate VVUQ and credibility/confidence considerations into the modeling process right from the beginning? For example, if a system is known to be adaptively complex, and to have a strong homeostatic response to the environmental variable in question, can the desired degree of confidence in that continued response determine at what level of detail the system should be modeled?

These potential Research Areas are clearly not orthogonal. Each informs the other three and arguably there might ultimately be a more logical "basis set" that a coherent research agenda partitions into. Indeed, there are at least two cross-cutting themes that one might want to meld in some manner with these potential Research Areas:

- Error Cross-Cut. This cross-cut would map various aspects of the potential Research Areas onto the four "classic" model/decision pair error types: Type I = Belief model is "wrong" when model is "right"; Type II = Belief model is "right" when model is "wrong"; Type III = Solved wrong problem; and Type IV = Problem solution used incorrectly.

- Polarities Cross-Cut. This cross-cut would map various aspects of the potential Research Areas onto various binary extremes that are often discussed, e.g.: qualitative versus

quantitative; objective versus subjective; knowing what is wrong versus knowing what is right; complex models versus models of complex systems; fast versus good response to application drivers.

Our hope is that this report, the written footprint of our workshop and of the pre-work leading to the workshop, can be the starting point for defining and articulating a long-term research agenda for Sandia and the larger community in this area.

1.1 Group Photo of Workshop Participants

Figure 1 Workshop Participants

From Left to Right: Steve Kleban, Bill Oberkampf, Sondra Spence, Tim Trucano, Matt Lave, Jim Stewart, Jackson Mayo, Mark Smith, Jeff Tsao, Vicente Romero, Drew Levin, Dan DeLaurentis, Ben Cook, Kasimir Gabert, David Stracuzzi, Linas Mockus, Asmeret Naugle, George Backus, Marcey Hoover, Roshanak Nateghi, Dennis Engi, George Barr, Brian Adams, Daniel Appelo, Thomas Kajder, Laura Swiler, Chris Frazier, Bruce Thompson, Sasha Outkin, Jacob Caswell, Wei Chen, Jessica Turnley, Dean Jones, Curtis Johnson, Lori Parrott, Emma Johnson, Craig Lawton, Tammy Brown, Shreyas Sundaram, Alfred Hubler, Nat Brown, Dan Pless, Steve Verzi

1.2   Workshop Announcement

Exceptional service in the national interest

Sandia National Laboratories

Complex Systems Models and Their Applications:
Towards a New Science of Verification, Validation and Uncertainty Quantification

Sandia National Laboratories ("Sandia") invites you to participate in "Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification." This Sandia-hosted workshop will be held on June 22-24th, 2016 in Albuquerque, NM.

Sandia has a long history of research and development in validation, verification (V&V), and uncertainty quantification (UQ) methods and capabilities. We also have a substantial body of research and development in complex systems science, modeling and simulation, and analysis. Sandia has developed an initial framework for addressing V&V and UQ of complex systems critical to our national security applications. The goal of the workshop is to share this framework more broadly with workshop participants, identify key research opportunities and approaches, and define potential future research directions in complex systems, V&V, and UQ. Sandia is inviting a select group of leading scientists and researchers in complex systems science, V&V, and UQ to collaborate on the development of this research agenda.

The following are research challenges that Workshop attendees will explore at the Workshop:

- Can the reliability, credibility, predictive capability and/or fidelity of complex systems models be characterized more quantitatively, rigorously and/or decisively?
- Can the ways in which models are currently "VVUQ'ed" be extended to apply to models whose purposes go beyond prediction of well-defined quantities of interest, which involve huge uncertainties, and in extreme cases whose underlying systems adapt and self–reorganize?
- Must new VVUQ methodologies be developed, and what are some of the forms these might take?

In addition, the following planned topics will also be addressed:

- Understanding non-standard models and their applications: models which make use of soft quantities such as credibility, insight, and belief; models intended to falsify rather than confirm; models which treat non-average rather than average behavior; models which assure "never" rather than "always"; models for understanding robustness and/or resiliency rather than efficiency; models based on correlation rather than causation; models intended to be used in real-time and sometimes "on the fly" rather than slowly and deliberately; models in which decision spaces, not just decisions, change; models of models (meta-models); etc.
- Exemplar complex systems, particularly cyber and the electric grid, with emphasis on current and future applications of models of these systems, and on the reliability, credibility, predictive capability and fidelity necessary to support those applications
- Understanding adaptive complexity, including system self-reorganization via internal competition and cooperation
- Fundamental limits to the application of "classical" VVUQ tools to current complex systems modeling methodologies, fundamental limits to the amenability of "classical" complex systems modeling methodologies to VVUQ tools, and potential ways to address these limits

Participants will be expected to bring ideas and may be asked to give a 5-minute talk on those ideas. Ideas may be informal and speculative. Please register by May 6, 2016; and please submit a 1-paragraph idea title/abstract and 1-paragraph short bio by June 1, 2016. Sandia intends to issue a Workshop Report which will be publicly released.

To register or for more information, please contact Sondra Spence – sspence@sandia.gov – Office: (505) 844-3609 – Cell: (505) 238-0227.

Regards,

Stephen Kleban
Manager, Systems Research, Analysis, and Applications Department
Complex Systems for National Security Group
Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY   NNSA

Sandia National Laboratories

1.3   Workshop Agenda

Sandia National Laboratories Workshop on
Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification
Albuquerque, New Mexico, 22-23 (,24) June, 2016

Agenda

Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification

Workshop Location:
Center for Global Security & Cooperation (CGSC)
10600 Research Rd. SE Albuquerque, New Mexico

Workshop Attire:
Casual

Security Notice: The following items are prohibited on Sandia premises: Removable computer media (e.g., CDs, USB Drives, etc.), audio and visual recording devices, and Bluetooth. Intoxicants, illegal drugs, drug paraphernalia, firearms, and explosive materials are also prohibited.

Host:	Steve Kleban, Cell (505) 362-1658, sdkleba@sandia.gov
Technical POC:	Jeff Tsao, Cell (505) 480-4267, jytsao@sandia.gov
Logistics POC:	**Sondra Spence, Cell: (505) 238-0227, sspence@sandia.gov**

Wednesday, 22 June 2016

0800-0900 **Registration and Continental Breakfast - - CGSC 1155/1154**

0830-0900 **Plenary: Welcome, Overview/Vision/Goals, Day One Process**
- *Carol* Adkins   Director, Energy Technologies and System Solutions

Center 6100

- Richard Griffith   Senior Manager, Nuclear Energy Safety Technologies

Group 6230

- Tim Trucano   Computing Research Center 1400
- Jeff Tsao   Semiconductor and Optical Sciences Group 1120

0900-1030 **Plenary: Framing of the Problem Space w/ Discussion – Part I**
- Laura Swiler   Optimization and UQ Dept`1441
  VVUQ Best Practices in Computational Science/Engineering Problems with some thoughts about extensions/limits to Complex Systems Models
- Stephen Verzi   Systems Research, Analysis and Applications Dept 6132
  Model/Use Characteristics that can Challenge Current VVUQ: Building Credible Models

1030-1045 **Refreshment Break**

1045-1130 **Plenary: Framing of the Problem Space w/ Discussion – Part II**
- Matt Lave   Photovoltaic and Distributed Systems Dept 6112
  Electric Grid Models: Current and Future Applications and Challenges to Those Applications

1130-1230 **Onsite Lunch**

1230-0200 **Plenary: Framing of the Problem Space w/ Discussion – Part III**
- Kasimir Gabert   Cyber Initiatives Dept 5638
  Cyber Models: Current and Future Applications and Challenges to Those Applications
- Curtis Johnson   Analytics and Cryptography Dept 5635
  Adaptive Complexity Revealed, with Challenges for Modeling and VVUQ of those Models

0200-0215 **Refreshment Break**

0215-0415 **Share "Seed" Research Ideas – Breakouts**
Workshop participants make brief presentation of research ideas, followed by brief discussions
Group 1: New VVUQ Theory (Mark Smith, facilitator)
Group 2: Concepts in Complexity (Asmeret Naugle, facilitator)
Group 3: Decision Making & Applications (Tatiana Flanagan, facilitator)

0415-0430 **Generate "Second-Pass" Research Ideas**
Workshop participants give ~3 "second-pass" research ideas to Sondra before adjourning

0430   **Adjourn**

0500-0700   **Offsite Dinner (Optional) (Day 2 Facilitators & Scribes Huddle)**

Thursday, 23 June 2016

0800-0915   **Continental Breakfast - - CGSC 1155/1154**

0830-0915   **Plenary: Articulation of Potential Research Areas ("Second-Pass" Research Ideas binned); Day Two Process**
- Tim Trucano   Computing Research Center 1400
- Jeff Tsao   Semiconductor and Optical Sciences Group 1120

0915-1030   **Discuss/Add/Organize "Second Pass" Research Ideas – (Five Breakout Groups – Part I)**
Discussion Leads: Alfred Hubler, Wei Chen, Bill Oberkampf, Daniel DeLaurentis, Lori Parrott
Scribes: Laura Swiler, Curtis Johnson, Matt Lave, Kasimir Gabert, Steve Verzi

1030-1045   **Refreshment Break**

1045-1115   **Plenary: Interim Report Out from Breakout Groups (Should Potential Research Areas be Re-Organized?)**

1115-1215   **Discuss/Add/Organize Research Ideas – (Five Breakout Groups – Part II)**

1215-0115   **Onsite Working Lunch**

1245-0145   **Plenary: Report Out from Breakout Groups (Articulate Potential Research Areas, Sub-Areas, and Key Research Ideas); Thanks; and Next Steps**
- Jeff Tsao   Semiconductor and Optical Sciences Group 1120

0145-0230   **What Did We Miss, and Who Can Help Write? -- (Five Breakout Groups – Part III)**

0230   **Main Workshop Adjourns**

0300-0500   **Sandia + Academic Alliance Exploration of Collaboration Opportunities: Brief Academic Alliance Presentations w/ Discussion**

0530-0700   **Offsite Dinner (On Own)**

Friday, 24 June 2016

0800-0830   **Continental Breakfast - - CGSC 1155/1154**
0830-0900   **Exploration of Collaboration Opportunities**
0900-1000   **Breakout Discussions around Specific Joint Research Opportunities**
1000-1015   *Refreshment Break*
1015-1100   **Breakout Discussions around Specific Joint Research Opportunities**
1100-1200   **Report Outs from Breakout Discussions**
1200   **Adjourn**

2. DAY ONE PLENARY TALKS

2.0   URLs of Video Recordings of Plenary Talks

Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.01 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 8:30 AM MDT
Duration: 00:30:20
Link: http://digitalmedia.sandia.gov/Mediasite/Play/574d7a30de734c5b9db82c4ccc63651a1d
Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.02 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 8:30 AM MDT
Duration: 00:40:24
Link: http://digitalmedia.sandia.gov/Mediasite/Play/06b990fe083b4f3c81a8ebb19b0447c41d
Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.03 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 8:30 AM MDT
Duration: 00:50:58
Link: http://digitalmedia.sandia.gov/Mediasite/Play/9af57fa33f6243eb89ab69b0f8594f851d
Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.04 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 10:30 AM MDT
Duration: 00:47:36
Link: http://digitalmedia.sandia.gov/Mediasite/Play/9b90b876dbb54bdb816e8e5d8a5895971d
Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.05 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 12:30 PM MDT
Duration: 00:38:00
Link: http://digitalmedia.sandia.gov/Mediasite/Play/57d4724445a54e5dbd58d9a6e55fa0291d
Presentation Details:
Title: (06/22 PC/VC/CC CGSC/1155) 001.06 Complex Systems and Their Models: Towards a New Science of Verification, Validation and Uncertainty Quantification - Jun. 2016
Date: Wednesday, June 22, 2016
Time: 12:30 PM MDT
Duration: 00:56:13
Link: http://digitalmedia.sandia.gov/Mediasite/Play/c4570546b2564e8e999f4aa394d4f4fd1d

2.1   Carol Adkins and Richard Griffith: Welcome

Welcome

Complex Systems and Their Models: Towards a New Science of Verification, Validation & Uncertainty Quantification

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Complex Systems Models and Their Applications:
Towards a New Science of Verification, Validation and Uncertainty Quantification

Carol Adkins
Director, Energy Technologies and System Solutions Center 6100

Complex Systems Models and Their Applications:
Towards a New Science of Verification, Validation and Uncertainty Quantification

Richard Griffith
Senior Manager & Resiliency in Complex System Research Challenge Lead

"Verification & Validation (V&V) is a difficult topic of significant importance, and Sandia is making good progress on establishing a framework… will require concerted study for at least a decade or more… Sandia would be the right sponsor for a workshop on this…commends the Complex Systems team for partnering with V&V experts from the "traditional" weapons engineering organization within Sandia" - 2015 Complex Systems External Advisory Board
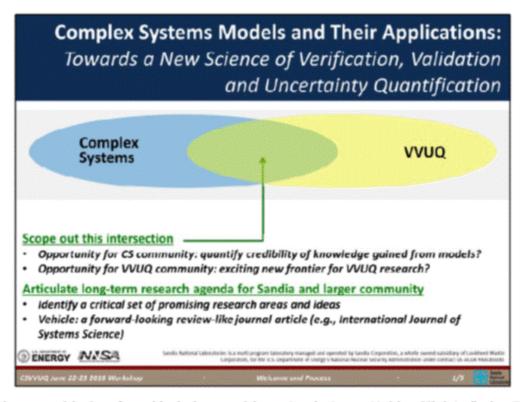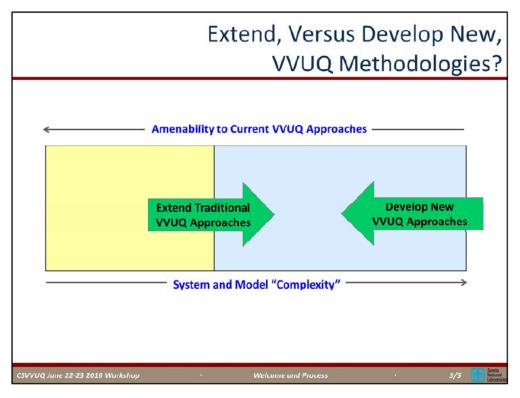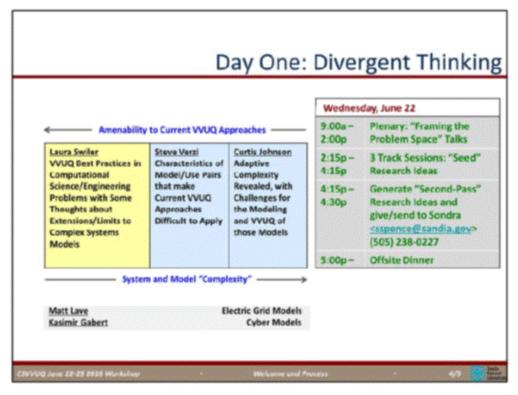
2.2   Jeff Tsao: Workshop Overview & Goals

- Welcome to, and thank-you for participating in, our workshop on Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification. What we're trying to do in this workshop, and in the pre- and post-work surrounding this workshop, is bring together two communities – the complex systems modeling community, and the VVUQ community – to scope out what they might do together at their intersection. Our working assumption is that there *is* opportunity for both communities at this intersection.
  - For the complex systems modeling community, how can they quantify the credibility of knowledge gained from their models, knowledge that is often incomplete and interim, but will nonetheless be used, sometimes in real-time, by decision makers?
  - For the VVUQ community, could this be an important next research frontier for them to tackle, with all sorts of exciting and interesting mathematical, statistical, maybe even cognitive-science, questions to answer?
- But we also don't just want to scope out this intersection. We want to articulate a long-term research agenda for Sandia and the larger community at this intersection. So our goal at this workshop is to identify a critical set of promising research areas and ideas at this intersection. And, if we get far enough with this, then, post-workshop, we hope to flesh out and articulate those research areas and ideas into a forward-looking journal article.

Extend, Versus Develop New, VVUQ Methodologies?

Amenability to Current VVUQ Approaches

Extend Traditional VVUQ Approaches

Develop New VVUQ Approaches

System and Model "Complexity"

CSVVUQ June 22-23 2016 Workshop · Welcome and Process · 3/5 — Sandia National Laboratories

- Now, whether the answer is to extend traditional VVUQ methodologies from the yellow zone to the blue zone, or whether the answer is to develop brand new VVUQ methodologies, or whether the answer is some of both, at this point doesn't seem so obvious. But the critical research areas and research ideas that might lead to filling in this blue zone is what I hope we will have made progress towards identifying by the end of this workshop.

Day One: Divergent Thinking

← Amenability to Current VVUQ Approaches →

Laura Swiler VVUQ Best Practices in Computational Science/Engineering Problems with Some Thoughts about Extensions/Limits to Complex Systems Models	**Steve Verzi** Characteristics of Model/Use Pairs that make Current VVUQ Approaches Difficult to Apply	**Curtis Johnson** Adaptive Complexity Revealed, with Challenges for the Modeling and VVUQ of those Models

← System and Model "Complexity" →

Matt Lave — Electric Grid Models
Kasimir Gabert — Cyber Models

Wednesday, June 22

9:00a – 2:00p	Plenary: "Framing the Problem Space" Talks
2:15p – 4:15p	3 Track Sessions: "Seed" Research Ideas
4:15p – 4:30p	Generate "Second-Pass" Research Ideas and give/send to Sondra <sspence@sandia.gov> (505) 238-0227
5:00p –	Offsite Dinner

- So now let me get into process a little bit. And let me start by saying that Sondra Spence has been our master chef for the workshop – Sondra, wave your hand -- so if anything comes up as this day and a half unfolds, she's our go-to person, along with her helper, Lauren Manzanares. But also if you can find Steve Kleban or me, we have been diligent sous chefs.
- Today, Day One, is all about Divergent Thinking.
- The rest of this morning and the early part of this afternoon, until 2p, we've prepared a series of plenary talks.
  - Laura Swiler will be talking about the yellow quadrant, traditional VVUQ approaches, with thoughts about how this might be extended out into the blue zone.
  - Steve Verzi will be talking about models for which current VVUQ approaches are more difficult.
  - To make all this a bit more concrete, Matt Lave and Kasimir Gabert will be giving examples of models in two domain areas of special interest to Sandia: the electric grid, and cyber.
  - Then, Curtis Johnson will be talking about the extreme of adaptive complexity, and some of the special challenges but also possibly special advantages, that those systems might bring.
- I want to emphasize that what we are trying to do in this workshop is very hard. So these talks will ask many more questions than they will answer. Our intent is to frame the problem space as best as we can, but there will be plenty of open-endedness, and we have allowed for plenty of time for Q&A after each talk.
- Then we'll break into three parallel track sessions, and this is when we get to hear the "seed" research ideas that you all submitted and that are in the handouts. Remember, these are informal and meant to stimulate thought, so don't feel you have to stay on script. If there is something you heard that stimulates you to change directions completely, go ahead and change directions. Let me thank in advance Mark Smith, Eric Vugrin and Asmeret Naugle, who will be facilitating those track sessions.
- Then at the end of the afternoon, and this is super-important, we'd like all of you to write down three or four "second-pass" Research Ideas that you think are promising. They could include the Research Idea you came in with. Or they could be totally different Research Ideas that were stimulated by the talks and discussions you heard during the day. We call them "second-pass" Research Ideas because they will be the result of a pretty intense mixing between the "seed" Research Ideas, the framing of the problem talks, and all of the discussion around these. They are super-important because they are what our Day Two will be structured around. So, at the end of today, no later than 4:30p, please write your three or four second-pass Research Ideas down, along with your name, and give them to Sondra. Paper is fine, a text message is fine, an email is also fine. Sondra does also have templates in the back for you to fill in if you'd like.
- Finally, we'll be going to dinner offsite in groups, Sondra should have those groups at the registration desk.

24

Day Two: Convergent Thinking

Thursday, June 23	
8:30a – 9:15a	Plenary: Articulation of Potential Research Areas (Tim Trucano)
9:15a – 10:30a	4 or 5 Breakout Sessions: Discuss/Add/Organize "Second Pass" Research Ideas
10:45a – 11:15p	Plenary: Interim Report Out (Should Potential Research Areas be Re-Organized?)
11:15p – 12:15p	4 or 5 Breakout Sessions (continued): Discuss/Add/Organize Research Ideas
12:45p – 1:45p	Plenary: Final Report Out (Articulate Potential Research Areas, Sub-Areas, and Key Research Ideas)
1:45p – 2:30p	Breakout Sessions: What Did We Miss, and Who Would Like to Help Write?
2:30p	Main workshop adjourns, but Academic Alliance agenda continues

- Then, while the rest of us are having dinner, Tim Trucano will be busy binning these second-pass Research Ideas, and Day Two, which is mostly about Convergent Thinking, will be organized around the Potential Research Areas that the Research Ideas binned into.

- We'll start the day with Tim articulating these Potential Research Areas.

- Then we'll break into sessions organized around these Potential Research Areas, where you'll discuss the various Research Ideas in those Research Areas, add new ones, and organize them into Research Sub-Areas. Here, let me thank in advance our Discussion Leads for these sessions: Wei Chen from Northwestern, Daniel DeLaurentis from Purdue, Alfred Hubler from University of Illinois, Bill Oberkampf, former Sandian, and Lori Parrott, manager of the Policy and Decision Analytics Department at Sandia. You can see we are emphasizing non-Sandians to be sure our workshop doesn't get too Sandia-centric.

- We might meet for a short interim plenary report-out, just for a sanity check on whether the Potential Research Areas are making sense or not. But we might not, if it looks like the Potential Research Areas *are* making sense.

- Then we'll break out again into Potential Research Areas, to continue with discussing, adding and organizing Research Ideas in those Areas.

- We'll do a final plenary report-out, where our Discussion Leads can articulate for the entire group their Potential Research Areas, the Sub-Areas they organized into, and some of the key Research Ideas in those Areas.

- And then, at the very end, we'll do one final breakout session after we've heard from all of the Potential Research Areas. This is where we'll have an opportunity to add what we now realize we've missed. This is also where we'll have an opportunity to volunteer to help write the journal article that we are hoping will emerge from this workshop.

- Now, before we actually dive into the workshop, any questions, comments, suggestions?

25

2.3 Laura Swiler: VVUQ Best Practices in Computational Science/Engineering Problems with Some Thoughts about Extensions/Limits to Complex Systems Models

Acknowledgements

With thanks to:

- Tim Trucano, Marty Pilch, Bill Oberkampf
- Brian Adams, Mike Eldred, Jim Stewart
- Laura McNamara
- Patrick Finley, Asmeret Bier
- Jim Kamm, Greg Weirs, Bill Rider
- ...and many others

Some thoughts on quality of models and assessments

- Do we have the courage to make sure that Sandia's modeling capabilities are not oversold?
- Expert judgment is not sufficient. Experts can be wrong and have their own biases.
- Just because someone pays you to do something doesn't mean it is valid.
- My job today is to provide some background on validation as it is performed in DOE's Advanced Simulation and Computing program and outline some problems with using this framework for complex systems models.
- I will emphasize reliance on data and benchmarks, performing uncertainty analyses and examining worst-case scenarios, and addressing the risk of using a model for a particular situation.

Do you know how good (or bad) your modeling and simulation is?

- "Perform due diligence and communicate frankly about assumptions, approximations, and limitations affecting simulation credibility.

Nothing against ostriches, of course
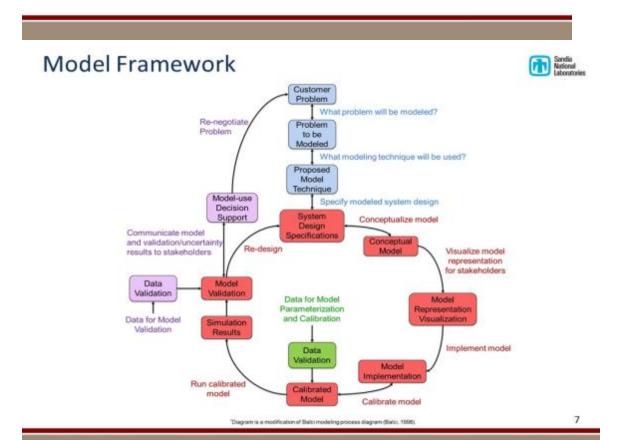
From the world of physics and engineering....

Prediction is hard:

- *Limited* physical data (observational or experimental)
- *Limited* simulations (high computational demands...)
- *Imperfect* computational models (missing physics, etc.)
- *Under-resolved* approximations or numerics
- *Unknown* model parameters and boundary conditions
- *Imperfect* humans
- We want to *extrapolate* to conditions beyond validation regime...

Model Framework

'Diagram is a modification of Balci modeling process diagram (Balci, 1998).

7

Verification: Are equations solved correctly?
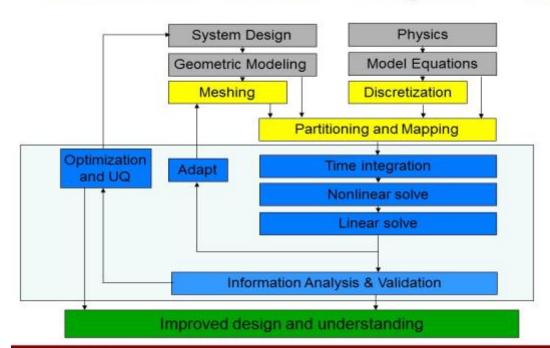
- Verification: a math and computer science issue
- Software quality is the bare minimum
- Software verification tools/concepts:
  - Requirements and software change and revision control
  - Unit and regression tests covering intended use and features, core and edge cases, not just line coverage (gcov/lcov)
  - Static and dynamic analysis (memory, performance, red flags)
  - Software quality engineering processes, design and code reviews
- Build in from the start!
- Numerical analysis tells us what properties an algorithm possesses — symmetry, stability, conservation, convergence, etc. — and under what conditions.
- In code verification, we test whether the implementation of the algorithm exhibits these properties with respect to known, exact solutions.

Best Practices: Verification Throughout

What is validation?

- Validation: Are we using the right model/equations for the intended application? A physics/engineering/science question
- Quantitative comparison between experimental data and numerical simulations.
- The purpose of validation is to determine the adequacy of a model for a particular application.
- Accounts for uncertainties and errors in both experimental data and simulations.
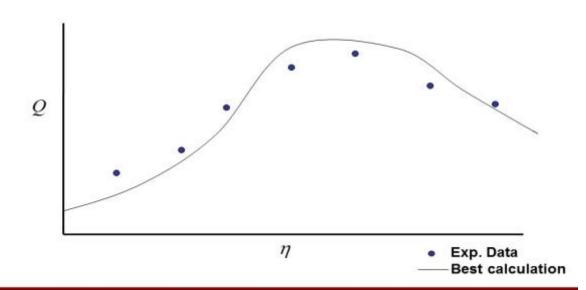
"If the test data are shown in blue and the simulation data are shown in yellow, then all I want to see is green."

(heard at Los Alamos, October 2005)

10
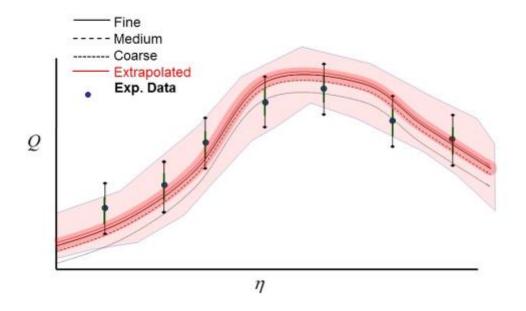
Is this model (black line) valid?

- Exp. Data
- Best calculation

31

Validation with experimental error, numerical and parametric uncertainty

Legend:
- —— Fine
- - - - - Medium
- ------ Coarse
- —— Extrapolated
- • Exp. Data

Q

η
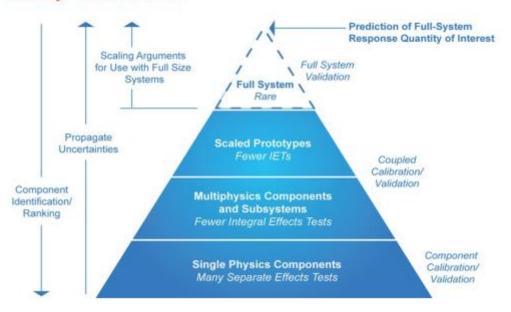
The *Validation Hierarchy*
A key construct

Prediction of Full-System Response Quantity of Interest

Scaling Arguments for Use with Full Size Systems

Full System Validation

Full System Rare

Propagate Uncertainties

Scaled Prototypes
Fewer IETs

Coupled Calibration/ Validation

Component Identification/ Ranking

Multiphysics Components and Subsystems
Fewer Integral Effects Tests

Single Physics Components
Many Separate Effects Tests

Component Calibration/ Validation

Example Validation Projects: Model to Use Case

- Validation of the radiative transfer equation (RTE) used in Sierra/Thermal-Fluid Dynamics calculations of pool-fire soot radiation.
  - Assess the validity of key model assumptions, which include: 1) gray soot radiation and 2) the size of the soot particles is much less than the wavelength of the radiation so that scattering is negligible.
  - Validation experiments performed with lightly-sooting liquid hydrocarbon fuels that yielded fully turbulent fires 2 m diameter. Radiation intensities were measured, with supplemental measurements of air flow and temperature, fuel temperature and burn rate, and flame surface emissive power, wall heat, and flame height and width provide a complete set of boundary condition data.

  - "Fire Intensity Data for Validation of the Radiative Transfer Equation." Thomas Blanchat and Dann Jernigan, SAND2016-0224.

14

Example Validation Projects

- Direct electron-beam-injection experiments for validation of air-chemistry models. SAND2016-2437C
- Development and Amputee Validation of Pressure and Shear Sensing Liner for Prosthetic Sockets. SAND2016-1580C
- Experimental data uncertainty calibration and validation of a viscoelastic potential energy clock model for inorganic sealing glasses. SAND2016-0856A
- Uncertainty Quantification Verification and Validation of a Thermal Simulation Tool for Molten Salt Batteries. SAND2016-3531C
- Development of a Benchmark Series of Cask Drops for Validation of Explicit Dynamic Finite Element Analyses. SAND2016-0757A
- Implementation and Validation of an Analytic Elastic-Plastic Contact Model with Strain Hardening in LAMMPS. SAND2016-2745C
- Mechanical Joints: Validation Assessment for Joint Problem Using an Energy Dissipation Model SAND2014-0864C, Validation of joint models in Salinas. SAND2004-3654C

15

Validation Metrics

- ASME V&V 20-2009 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer:
- https://www.asme.org/products/codes-standards/v-v-20-2009-standard-verification-validation

Term	Definition	Formula
T	Unknown truth data	
S	predicted value from a simulation	$S = T + \delta_S$
D	value determined from experimental data	$D = T + \delta_D$
E	Validation comparison error	$E = S - D = \delta_S - \delta_D$
δ	Error: Difference between measurement and true value	$\delta_S = \delta_{model} + \delta_{num} + \delta_{input}$
u	Uncertainty: Characterization of the dispersion of values	$u_{val} = \sqrt{u_{num}^2 + u_{input}^2 + u_D^2}$
δ_{model}	Model error due to modeling assumptions and approximations	$\delta_{model} = E - (\delta_{num} + \delta_{input} - \delta_D)$.

16

Validation Metrics:

We have an interval ($E\pm u_{val}$) in which δ_{model} falls:

$$\delta_{model} \in [E - u_{val}, E + u_{val}].$$

Case	Validation interpretation
$|E| \gg u_{val}$	It is likely that $\delta_{model} \approx E$. In this case, one would want to reduce the modeling error.
$|E| \le u_{val}$	The modeling error is within the noise level imposed by numerical, input, and experimental uncertainties and it will be harder to formulate model improvements.
Assume u_{val} is Gaussian	One can say that 95% of the population is covered within the interval ($E \pm 2*u_{val}$)

Performing validation

- Are we ready to estimate $\delta_{model} \in [E - u_{val}, E + u_{val}]$?
- What do we need? An estimate of numerical, parameter, and experimental uncertainties

$$u_{val} = \sqrt{u_{num}^2 + u_{input}^2 + u_D^2}$$

- Parameter uncertainty can be estimated in two ways:

$$u_{input}^2 = \sum_{p=1}^{m} \left(\frac{dS}{dx_p} u_{x_p}\right)^2 \qquad OR \qquad u_{input}^2 = \frac{1}{n}\sum_{i=1}^{N}(S_i - \mu_s)^2$$

- Multiple extensions to multiple responses (e.g. time-varying), various statistical methods for pooling experimental data across different conditions, thresholds for model adequacy, etc.

Other Validation Metrics
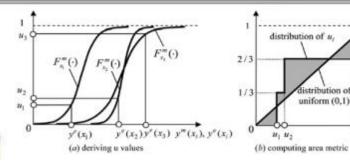
Sandia National Laboratories

"Toward a better understanding of model validation metrics." Y. Liu, W. Chen, P. Arendt, and H. Huang. Journal of Mechanical Design, July 2011, Vol. 133. DOI: 10.1115/1.4004223

Table 6   Summary of the main characteristics of the validation metrics

	Classical hypothesis	Bayes factor	Frequentist's metric	Area metric
Quantitative measure	Yes	Yes	Yes	Yes
Objective measure	Yes	No	Yes	Yes
Excludes any belief and criterion of accepting model	No	No	Yes	Yes
Includes all uncertainty sources	Yes	Yes	No	Yes
Feasible for multivariate case	Yes	Yes	No	No, but can be extended
Generalizes deterministic comparisons	No	No	Yes	Yes
Considers confidence level associated with amount of experimental data	Yes	Yes	Yes	No
Model not improved by widening the distribution of model parameter	Yes, if experimental data is more than one; otherwise No	Yes	No	Yes
Assesses global predictive capability	No	Yes, but needs improvement	Yes	Yes

(a) deriving u values

(b) computing area metric

19

Supporting capabilities

Sandia National Laboratories

- Sensitivity Analysis
  - Identify most important variables and their interactions
  - Understand code output variations as input factors vary
  - Often correlation coefficients, scatterplots, or variance-based indices
- Uncertainty Quantification
  - Determine the probability distribution of code outputs, given uncertainty in input factors
  - Assess the likelihood of typical or extreme outputs given input uncertainties: determine mean or median performance, assess variability in model responses, find probability of failure
  - Assess how close code predictions are to experimental data (validation) or performance limits (margins)
- Calibration
  - determine optimal parameter values that yield simulation results which "best match" the experimental data in some sense
  - Least-squares methods, Bayesian calibration methods

20

More on UQ/SA

- Advances in UQ methods have been tremendous BUT still severely limited with respect to dimensionality
- Most of our UQ/SA methods deal with continuous variables, not large numbers of discrete entities
- Sampling approaches are probably the only mainstream viable approach currently
- Surrogate models have also advanced tremendously the past 15 years, but again, limited to a handful of variables and responses, mainly continuous

Challenge:
general coupled multi-physics

- Can we efficiently propagate UQ across scales/disciplines?
- Naively wrapping multi-physics with UQ often too costly
- Can we invert loops and perform multi-physics analysis on UQ-enriched simulations (couple individual sampling loops based on scalar statistics, random fields, stochastic processes)?
- Embedded methods: examining GPUs, re-structuring of sampling loops (Eric Phipps, Sandia)
  - Instead of N separate samples where each requires a full solve of residuals/Jacobians, restructure the operations
  - Form a large block nonlinear problem
  - Krylov Basis Recycling
  - Increased vectorization and data locality
  - Implemented through Stokhos embedded UQ package

Lessons from the ASC Program

- DATA: Can I design and execute a validation experiment? What benchmarks are available? Difference between data used for calibrating the model vs. validation.
- Early emphasis on identifying and quantifying epistemic vs. aleatory uncertainty
- More recently, issues of model form uncertainty and model selection are of interest
- Bayesian methods are becoming more popular
- Limits of the validation hierarchy
- IT IS VERY EXPENSIVE!!!!
- Our validation statements are fairly limited
- Validation is a process

23

Frontiers of V&V/UQ

- Complexity of codes, multi-physics couplings
- Few examples of validation across the hierarchy
- UQ: Focus is on methods that minimize the number of function evaluations while maximizing accuracy in response statistics
- Calibration: People are tacking more complicated scenarios, want parameters estimated with uncertainties from a variety of experimental configurations
- People are doing experiment to simulation comparisons, but characterizing, propagating, and comparing uncertainties remains a big issue
- DATA, DATA, DATA
- How to deal with models that are "theory rich and data poor" vs. models that are "data rich and theory poor"?

24

Recommendations

- Focus on Data: what data was used to build the model, what data was used to calibrate the model, what data was used to evaluate the model, what data was used for validation? What benchmark data sets are available?
- Use of SA/UQ/Optimization in terms of understanding the behavior of the model, identifying worst case scenarios, understanding the spread of possible outcomes and their likelihood.
- Perform risk management on the use of the model: what are the risks associated with using this model?

Thoughts to consider

- Hans Meir's thoughts on validation of complex systems models:
  - Naïve practitioners in the softer sciences are pretending to "validate" their tools based on harder-science verification practices.
  - **We should be enforcing a course-correction that steers the managers and naïve practitioners away from that pseudo-science.**
  - Instead, we should recognize the deep uncertainty inherent in softer science challenges, stop pretending to gloss over that uncertainty with sophisticated modeling and validation tricks, and prevent the inappropriate use of models for predictive/forecasting purposes.

Thoughts to consider

- Models don't forecast....People Do. (Laura McNamara, SAND, NAS)

- The challenge of evaluating computational social modeling and simulation technologies *extends far beyond verification and validation*, and should include the relationship between a simulation technology and the people and organizations using it.
- This challenge of evaluation is not just one of usability and usefulness for technologies, but extends to the assessment of how new modeling and simulation technologies shape human and organizational judgment.
- The robust and systematic evaluation of organizational decision making processes, and the role of computational modeling and simulation technologies therein, is a critical problem for the organizations who promote, fund, develop, and seek to use computational social science tools, methods, and techniques in high-consequence decision making.

27

Thoughts to consider

- Sargent: three basic decision-making approaches for deciding whether a simulation model is valid
    - The model development team decides based on their testing
    - The user of the model decides, based on interaction with the model development team, ideally throughout all stages of development
    - An external third party decides, based on their own testing
- *Conceptual model validation* is defined as determining that the theories and assumptions underlying the conceptual model are
- *Computerized model verification* is defined as assuring that the computer programming and implementation of the conceptual model are correct.
- *Operational validation* is defined as determining that the model's output behavior has a satisfactory range of accuracy for the model's intended purpose over the domain of the model's intended applicability.
- *Data validity* is defined as ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct.

28
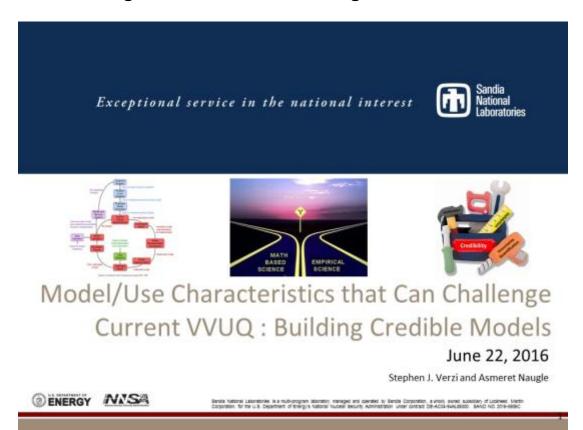
Thoughts to consider

- "When all models are wrong." Saltelli and Funtowicz, IST 2014
- Checklist for the responsible development and use of models.
  - Rule 1: Use models to clarify, not obscure
  - Rule 2: Adopt an "assumption hunting" attitude
  - Rule 3: Detect pseudo-science (the practice of ignoring or hiding the uncertainties in model inputs in order to ensure that model outputs can be linked to preferred policy choices. A common indicator of this kind of pseudoscience is spurious precision.)
  - Rule 4: Find sensitivity assumptions before they find you
  - Rule 5: Aim for transparency
  - Rule 6: Don't just "do the sums right" but "do the right sums."
  - Rule 7: Focus the analysis.

→SENSITIVITY AUDIT

2.4 Steve Verzi and Asmeret Naugle: Model/Use Characteristics that can Challenge Current VVUQ: Building Credible Models

Our technical team

- Jeffrey Tsao
- Asmeret Naugle
- Mark Smith
- Tatiana Flanagan
- Kasimir Gabert
- Matthew Lave
- Eric Vugrin
- Sasha Outkin
- Curtis Johnson
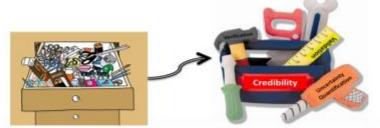
- Timothy Trucano
- Laura Swiler
- Laura McNamara
- George Backus
- Patrick Finley
- Dean Dobranich
- Timothy Wildey
- Brian Adams
- William Rider
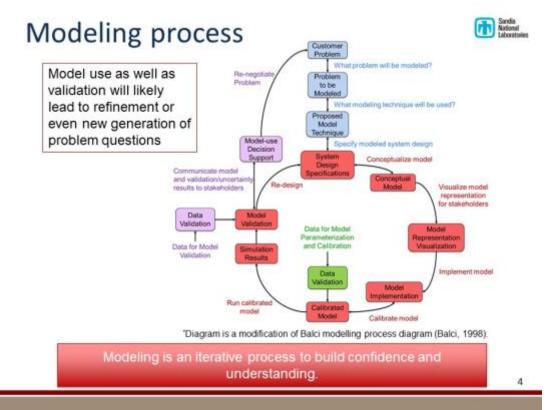
2

Outline

- Modeling process
- Characteristics of complex systems
- Purpose of VVUQ in modeling complex systems
- Model uses for complex systems
- When can modeling challenge current VVUQ?
- Opportunities for further research
- Summary

3

Modeling process

Model use as well as validation will likely lead to refinement or even new generation of problem questions

"Diagram is a modification of Balci modelling process diagram (Balci, 1998).

Modeling is an iterative process to build confidence and understanding.

4

Model uses
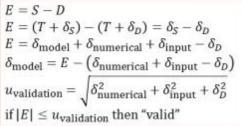
According to Shannon (Shannon, 1975)
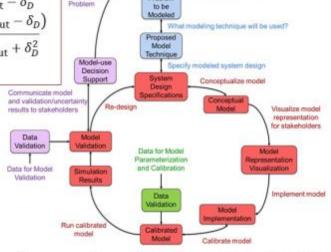
- Evaluation of system behavior
- Forecasting
- Comparison of different operating policies
- Optimization
- Sensitivity analysis
- Determination of functional relationships
- Training

5

Formal model validation

$$E = S - D$$
$$E = (T + \delta_S) - (T + \delta_D) = \delta_S - \delta_D$$
$$E = \delta_{\mathrm{model}} + \delta_{\mathrm{numerical}} + \delta_{\mathrm{input}} - \delta_D$$
$$\delta_{\mathrm{model}} = E - (\delta_{\mathrm{numerical}} + \delta_{\mathrm{input}} - \delta_D)$$
$$u_{\mathrm{validation}} = \sqrt{\delta_{\mathrm{numerical}}^2 + \delta_{\mathrm{input}}^2 + \delta_D^2}$$
$$\text{if } |E| \leq u_{\mathrm{validation}} \text{ then "valid"}$$

Formal validation: compare simulation results with experimental or observational data

*Diagram is a modification of Balci modelling process diagram (Balci, 1998).

Model error within noise level imposed by numerical, parameter input and experimental data uncertainties → "valid".

6

Characteristics of complex systems

- adaptive or self-organized behavior
- high throughput
- heterogeneity of subcomponents
- multi-scale interaction
- bifurcations and phase change
- cascading and/or emergent behavior
- synergistic components
- feedback loops
- non-linearity
- humans in the loop
- lack of established theory and/or unknown basic physical laws
- feedback from model to system
- inability to conduct experiments, lack of data and/or low signal-to-noise
- reliance upon soft quantity data

- out of equilibrium
- results and assessment focus on dynamics or dynamical behavior
- reductionist approach is inappropriate (irreducible)
- open system
- imbalanced information exchange
- exhibits power laws
- multi-objective behavior
- non-locality
- complementary quantities of interest
  - Heisenberg uncertainty principle
- inability to specify closed-form description but can be simulated
- potentially unpredictable
- social dynamics

List is incomplete, non-orthogonal and not definitive or universally agreed upon.

7

Purpose of VVUQ in modeling complex system

- Continue to use formal validation where possible
- Build confidence in modeling process where formal validation cannot be applied

Includes stakeholders and potential stakeholders

'Diagram is a modification of Balci modelling process diagram (Balci, 1998).

Understand and quantify credibility in the use of complex systems modeling approaches with respect to a particular use.

8

Model uses for complex systems

Key model/use factors to be addressed

- What is the question/quantity of interest (knowledge of interest)?
- Which model is appropriate for answering this question?

Potential model uses

- Prediction
- Policy exploration
- Risk analysis
- Hypothesis testing of system theory
- Real-time operation (feedback & control)

9

Model/use credibility
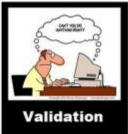
- Do we have the right model for our intended use?
  - Engineered systems – "Assurance that a product, service, or system meets the needs of customers and other identified stakeholders" ISO/IEC 15288
  - Software – "Process of evaluating software design and implementation to determine whether it satisfies specified requirements" IEEE SA 1012
- Are we asking the right question of our model?
  - Customers do not always know what they want
  - Customers need Y but ask for X
  - Have model Z that we want to apply
  - Want to research/build model W

Quote from Swiler, 2016.

Validation

"Perform due diligence and communicate frankly about assumptions, approximations, and limitations affecting simulation credibility."

10

Prediction is hard:

- *Limited* physical data (observational or experimental)
- *Limited* simulations (high computational demands…)
- *Imperfect* computational models (missing physics, etc.)
- *Under-resolved* approximations or numerics
- *Unknown* model parameters and boundary conditions
- *Imperfect* humans
- We want to *extrapolate* to conditions beyond validation regime…

Slide from Swiler, 2016.

11

When can modeling challenge current VVUQ?

- Limited (observational or experimental) data
  - Modeled system is unpredictable
- Additional data challenges
- Imperfect models
  - Computation and/or physics
  - Complex behavior dynamics
- Limited model simulations (time & cost)

Current approaches to assess model credibility

- Limited data
- Additional data challenges
- Imperfect models
- Extrapolating beyond validated (validate-able) regime

Limited data

- inability to conduct experiments
  - detonate "dirty bomb" in NYC
  - impact of a natural disaster
- imbalanced information
  - stealth
  - one-sided learning
- Modeled system is unpredictable

Need to understand potential effects we cannot create ourselves.

Additional data challenges

- Data quality
- Dimensionality
  - heterogeneity of subcomponents
- Too much data
  - multi-scale interactions
  - low signal-to-noise
- Continuous dynamic data
  - bifurcations and phase change
  - out of equilibrium
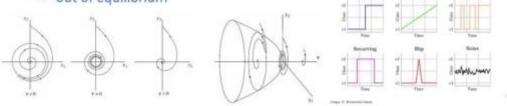
- Volume
- Velocity
- Variety
- Value
- Visibility

Types of concept drifts

14

Imperfect models

Computation and/or physics

- non-linearity
- synergistic components
- open system
- reductionist approach is inappropriate (irreducible)
- lack of established theory and/or unknown basic physical laws
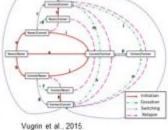
Exploratory modeling

Complex behavior dynamics

- adaptive or self-organized behavior
- cascading and/or emergent behavior
- multi-objective behavior
- social dynamics
- humans in the loop

Re-negotiate problem

Verzi et al., 2012.

Vugrin et al., 2015.

15
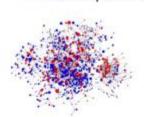
Limited model simulations

- high throughput
  - turbulent flow
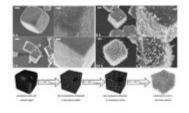
- complex behavior dynamics
  {cascading, emergent, self-organized, multi-objective, adaptive}
  - self-organizing materials
  - social dynamics
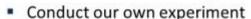
Need to understand potential effects that are too expensive to simulate completely (a large number of times).

16

Current approaches to assess model credibility

Limited data

- Find surrogate source
- Simulate it
  - Use known data to design generative model
  - Use micro scale model to generate macro scale data
- Conduct our own experiment
  - Small sample size – biased?
- Survey subject matter experts (SMEs)
- Model without it
  - Use "best guess"

WEEKDAYS
8 AM to 8 PM JUL 1-30

The NYPD will release small amounts of **harmless**, **colorless** **gas** in 5 boroughs and 21 subway lines. This **airborne** **non-toxic test** is part of NYPD's precaution against terror attacks. MTA Customers are advised to **stay calm**.

17
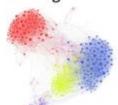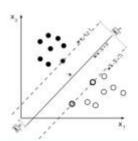
Current approaches to assess model credibility

Additional data issues

- Data quality V&V
- Dimension reduction
  - Feature extraction – deep learning
  - Principle components analysis
  - Compression
- Data clustering
- Machine learning

Data validation

18
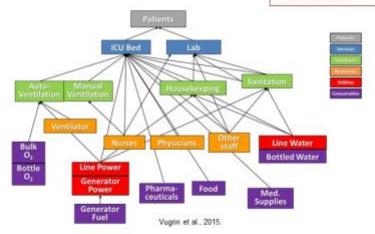
Current approaches to build model credibility with stakeholders

- Adaptive resilience in hospital ICU

Model representation visualization

Vugrin et al., 2015.
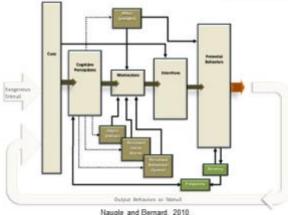
Work with customers to derive insights from model for adaptive substitution and to build credibility by recreation of historical event.

19

Current approaches to build model credibility with stakeholders

- Complex socio-cognitive behavior

Model representation visualization

Naugle and Bernard, 2010.
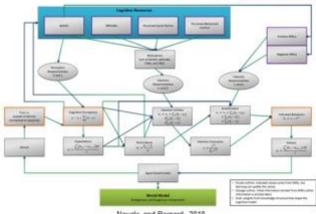
Work with customers/stakeholders to derive insights from modeling process and to understand modeled behavior equations.

20

Current approaches to build model credibility with scientific community

- Complex socio-cognitive behavior
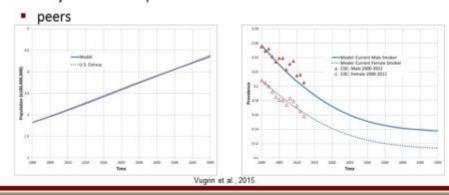
Model representation visualization

Naugle and Bernard, 2010.

Work with customers/stakeholders to derive insights from modeling process and to understand modeled behavior equations.

21

Current approaches to assess model credibility

Extrapolating beyond validated (validate-able) regime

- Model-to-model comparison
  - docking
  - co-validation (model verification)
- Independent review
  - subject matter experts
  - peers

Vugrin et al., 2015

22

Opportunities for further research
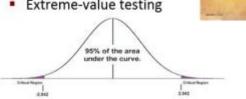
- Lack of data
  - Use analogous, synthetic or expert-elicited data
  - Reproduction of
    - system theory
    - phenomena
    - behavior

- Extending model credibility
  - Face validation
  - Turing test
  - Extreme-value testing

95% of the area under the curve.

23

Summary

- Modeling process
- Complex systems modeling characteristics
- Goal: understanding and quantifying model credibility
- VVUQ gaps
- Current approaches
- Opportunities

24

References

- Backus, G., Bernard, M., Verzi, S., Bier, A. and Glickman, M. 2010. "Foundations to the Unified Psycho-Cognitive Engine." *Sandia National Laboratories Technical Report*, SAND2010-6974.
- Balci, O. 1998. "Verification, Validation and Testing." (Chapter 10) In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, Banks, J. (Ed.), (Wiley Online Library) John Wiley & Sons, Inc., Hoboken, NJ. doi: 10.1002/9780470172445.ch10.
- Johnson, C.M., Backus, G.A., Brown, T.J., Colbaugh, R., Jones, K.A. and Tsao, J.Y. 2012. "A Case for Sandia Investment in Complex Adaptive Systems Science and Technology." *Sandia National Laboratories Technical Report*, SAND2012-3320.
- Naugle, A.B. and Bernard, M.L. 2014. "DYMATICA Modeling and Assessment Dynamic Multi-Scale Assessment Tool for Integrated Cognitive Behavioral Actions." *Sandia National Laboratories Technical Presentation*, SAND2014-15338PE.
- Rider, W.J., Kamm, J.R. and Weirs, V.G. 2010. "Verification, Validation and Uncertainty Quantification Workflow in CASL." *Sandia National Laboratories Technical Report*, SAND2010-234P.
- Sargent, R.G. 2011. "Verification and Validation of Simulation Models." *Proceedings of the 2011 Winter Simulation Conference*, Jain, S., Creasey, R.R., Himmelspach, J., White, K.P. and Fu, M. (Eds.): 183-198. http://dl.acm.org/citation.cfm?id=2431538.
- Shannon, R.E. 1975. *Systems Simulation: the Art and Science*, Prentice-Hall, Englewood Cliffs, NJ. doi: 10.1109/TSMC.1976.4309432.
- Sornette, D., Davis, A.B., Vixie, K.R., Pisarenko, V. and Kamm, J.R. 2007. "Algorithm for model validation: Theory and applications." PNAS, 104(16): 6562-6567. doi: 10.1073/pnas.0611677104.
- Verzi, S.J., Apelberg, B., Rostron, B. Brodsky, N., Brown, T.J., Husten, C. and Glass, R.J. 2012. "An Agent-Based Approach for Modeling Population Behavior and Health with Application to Tobacco Use." *Sandia National Laboratories Technical Report*, SAND2012-6898.
- Vugrin, E.D., Rostron, B.L., Verzi, S.J., Brodsky, N.S., Brown, T.J., Choiniere, C.J., Coleman, B.N., Paredes, A. and Apelberg, B.J. 2015. "Modeling the Potential Effects of New Tobacco Products and Policies: A Dynamic Population Model for Multiple Product Use and Harm." PLOS ONE, 10(3): e0121008. doi:10.1371/journal.pone.0121008.
- Vugrin E, Verzi S, Finley P, Turnquist M, Griffin A, Ricci K, Wyte-Lake T. 2015. "Modeling hospitals' adaptive capacity during infrastructure disruptions." *Journal of Healthcare Engineering*, 6(1): 85-120. PM ID: 25708379. doi: 10.1260/2040-2295.6.1.85.

25

Additional References

- Allen, E.H., Tallant, G.S. and Elliot, M.A. 2014. "Computer systems and methods for quantum verification and validation." US Patent 8832165, *Google Patents*. US 8832165.
- Koulouriotis, D.E., Diakoulakis, I.E., Emiris, D.M. and Zopounidis, C.D. 2005. "Development of dynamic cognitive networks as complex systems approximators: validation in financial time series." *Applied Soft Computing*, 5: 157-179.
- McElhaney, J. 2011. "No Laughing Matter: Failing the Giggle Test Might Leave You Crying." ABA Journal. http://www.abajournal.com/magazine/article/the_giggle_test.

References

Chronology of V & V for Complex Systems and Complex Adaptive Systems

- Wise, J.A., Hopkin, V.D. and Stager, P. (Eds.). 1993. *Verification and Validation of Complex Systems: Human Factors Issues, NATO ASI Series*, Springer, Heidelberg, Germany.
- Van Dijkum, C., De Tombe, D. and Van Kuijk, E. 1999. *Validation of simulation models*. Siswo Publication, Amsterdam.
- Kelton, W.D. and Law, A.M. 2000. *Simulation and modeling analysis*. McGraw Hill, Boston, MA.
- Stevenson, D.E. 2002. "Verification and Validation of Complex Systems." *Intelligent Engineering Systems through Artificial Neural Networks (2002)*, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.6.8571&rep=rep1&type=pdf.
- Küppers, G. and Lenhard, J. 2005. "Validation of simulation: Patterns in the social and natural sciences." *Journal of Artificial Societies and Social Simulation*, 8(4): 3. http://jasss.soc.surrey.ac.uk/8/4/3.html.
- Macal, C.M. and North, M.J. 2005. "Validation of an Agent-based Model of Deregulated Electric Power Markets." *Proceedings of the North American Computation Social Organization Science (NAACSOS) 2005 Conference*. doi: 10.1.1.72.6692.
- North, M.J., Hower, T.R., Collier, N.T. and Vos, J.R. 2007. "A declarative model assembly infrastructure for verification and validation." *Advancing social simulation: the first world congress*, Springer, Japan. 129-140. doi: 10.1007/978-4-431-73167-2_13.
- Windrum, P., Fagiolo, G. and Moneta, A. 2007. "Empirical validation of agent-based models: Alternatives and prospects." *Journal of Artificial Societies and Social Simulations*, 10(2): 8. http://jasss.soc.surrey.ac.uk/10/2/8.html.
- Omerod, P. and Rosewell, B. 2009. "Validation and Verification of Agent-Based Models in the Social Sciences." In *Epistemological Aspects of Computer Simulation in the Social Sciences: Second International Workshop, EPOS 2006*, Springer, Heidelberg, Germany: 130-140. doi:10.1007/978-3-642-01109-2_10.
- Bruch, E. and Atwell, J. 2013. "Agent-based models in empirical social research." *Sociological Methods & Research*, 44(2): 186-221. doi: 10.1177/0049124113506405.
- Luna, S, Lopes, A., Tao, H.Y.S, Zapata, F. and Pineda, R. 2013. "Integration, Verification, Validation, Test, and Evaluation (IVVT&E) Framework for Systems of Systems (SoS)." *Procedia Computer Science, 20*: 298-305. doi:10.1016/j.procs.2013.09.276.
- Bert, F.E., Rovere, S.L., Macal, C.M., North, M.J. and Podestá, G.P. 2014. "Lessons from a comprehensive validation of an agent-based model: The experience of the Pampas Model of Argentinean agricultural systems." *Ecological modeling*, 273: 284-298. doi: 10.1016/j.ecolmodel.2013.11.024.

Thank you

- Questions?

2.5 Matt Lave: Complex Systems and the Electric Grid

Transcription Grid

Separate Systems
The U.S. has three big regional power grids. But technical obstacles mean the grids have limited connections between them, making it hard for them to help each other in emergencies.

WESTERN

EASTERN

HIGHEST CAPACITY POWER LINES (kilovolts)
— 345 to 450
— 500 to 525
— 735 to 765
— Direct current

TEXAS

Note: Smaller-capacity lines and some direct-current lines are omitted for clarity.

Source: Platts
The Wall Street Journal
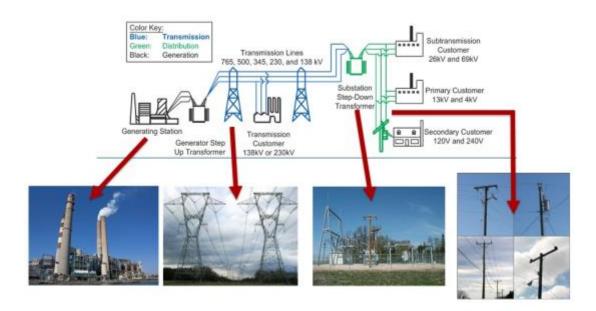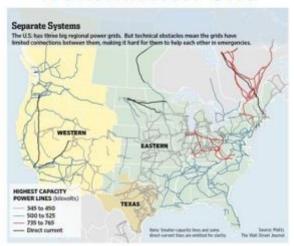
- Interconnected network with subsections overseen by different human operators
- Adaptive system: power flow magnitude and direction will change;
- Significant effort spent on outage (low-probability event) planning: "n-1" contingencies

http://www.energyandpolicy.org/value-of-solar-versus-fossil-fuels-part-three

3

Distribution Feeders

Fixed Cap
900 kVAr

3.36 miles

LDC Control Point

Switched Cap1
1200 kVAr

Switched Cap2
1200 kVAr

Substation

PEAK SOLAR

MORNING DEMAND   EVENING DEMAND

Now
- Little/no monitoring along feeder
- Photovoltaic (PV) installations inject variable power at locations along feeder
- Incomplete information about PV locations, performance

Upcoming
- Ubiquitous sensing (e.g., smart meters)
- Customer-owned storage
- Demand incentives: customer reactions
- Customer-owned PV inverters: grid support

4

Electric Grid Model/Decision Pairs

- **Traditional**
  - Rigorously derived from well-established physics and mathematics.
  - Example: Power flow with known loads and generation.
  - Decision: To accommodate a given load, what line rating do I need (how much current must it be able handle)?
- **Non-Traditional**
  - Typically involve layers of uncertainty and/or require large amounts of data to properly characterize.
  - Example: Stochastic Optimization
  - Decision: How to operate grid with uncertain solar power production (weather)?
- **Adaptively Complex**
  - Involving human activity and long timescales. Applications include determining likelihoods and consequences of rare events as well as subsequent contingency planning and recovery management.
  - Model: Coupled Simulations (weather, contingency analysis, cascading failures, agent-based models, sociological models, logistics, etc.).
  - Decision: How should the electric grid and associated infrastructure be designed and operated to maximize resilience?

Distribution and Transmission Models

- Distribution grid models
  - Often traditional
    - Straightforward decisions e.g. on line ratings
  - Becoming non-traditional
    - Addition of solar PV as generation: no longer radial flow from substation
  - Later, adaptively complex
    - PV inverters provide grid services – incentives
    - Additional devices such as programmable thermostats and home battery walls change consumption behaviors.
- Transmission grid models
  - Already non-traditional to adaptively complex
    - Stochastic analysis
    - Many human operators
    - Contingency n-1 and n-k analysis

Distribution Grid: Traditional
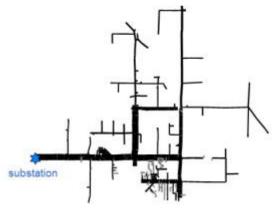
- **Planning line ratings**
  - Based expected effective resistance (load)
  - Essentially V = IR

- **Adjust voltage settings at substation**
  - Use basic electric equations to estimate voltage at end of line based on substation voltage and simple assumptions

substation

An example distribution grid. Thicker lines mean higher line ratings (can handle higher current).
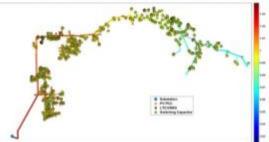
Distribution Grid: Non-Traditional

Keep safe voltage levels with PV

- Solar PV injects power and raises voltage along a distribution feeder
- May lead to voltage violations on the feeder
  - Fines, blackouts
- Uncertainty in PV production due to weather

Non-traditional because:

- Uncertainty in PV production due to weather
- PV system locations, sizes, performance are often unknown, can change
- Utility may have no local monitoring, or may on "see" net load: load minus PV generation
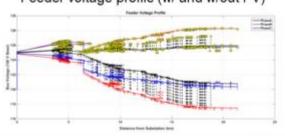
Feeder layout and PV locations

Feeder voltage profile (w/ and w/out PV)

Distribution Grid: Non-Traditional
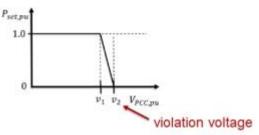
PV controls can help mitigate voltage violations.

- E.g., volt-watt control

$P_{set,pu}$

1.0

0

v_1 v_2 $V_{PCC,pu}$

violation voltage

Decision: What are best settings for PV controls to maximize PV energy?

- Due to interplay between different PV systems, difficult/impossible to "traditionally"[*] define v1 and v2.
- Instead, determined experimentally through model optimization.

[*] "traditionally" as in: v1 = fn(location, PV size)
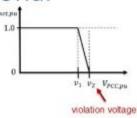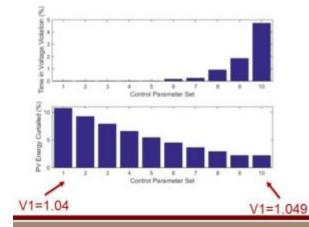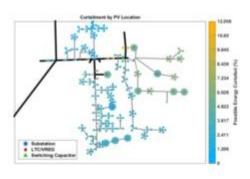
Distribution Grid: Non-Traditional

PV controls can help mitigate voltage violations.

$P_{set,pu}$

1.0

0

v_1 v_2 $V_{PCC,pu}$

- Tradeoff between curtailment of PV energy and voltage violations.
- Different results at each PV location.

violation voltage

Time in Voltage Violation (%)

Control Parameter Set

PV Energy Curtailed (%)

Control Parameter Set

Curtailment by PV Location

- Substation
- LTC/VREG
- Switching Capacitor

V1=1.04

V1=1.049
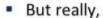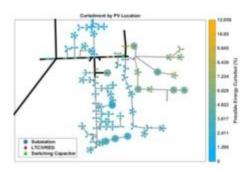
Distribution Grid: Adaptively Complex

Determine the impact of incentives/what are best policies

- How will customers react to price signals, incentives, etc.?

- Previous slide results:
  - Based on load from one historical year
  - Based on weather from one historical year
  - Based on static, hypothetical PV scenario
- But really,
  - All of these change (adapt) due to incentives, social acceptance, technology, etc.
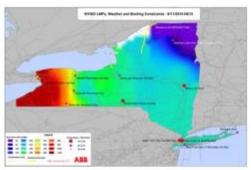
Transmission Grid Complexity

Transmission grid models used for:

- Load/generation balancing
  - Import/export

- Locational marginal pricing
  - Incentivize generation at certain nodes

- "n-1" contingency analysis
  - Plan for loss of generator or line

- "n-k" contingency analysis
  - Threat and vulnerability analysis

Locational marginal pricing showing high prices in Western New York due to unexpectedly high temperatures and hence high AC loads.

Image from: http://energymarketintel.com/wp-content/uploads/2015/05/NYISO-LMP-and-Weather-Heat-Map-5.11.15-HE15_v2-copy.png

Transcription Grid Complexity

Many different human operators, each with their own overlapping but not fully aligned incentives

- Avoid blackouts in own territory
- Manage generation you can control
- Communicate with neighbors for assets you don't have control over

NY Times: *On the Front Lines of the Power Grid, Oct. 25, 2011:*

"QUICK! You are on duty in a secret control room in a nondescript, windowless building. ...your task is crucial: you are matching the ever-changing power needs of tens of millions of electricity customers with supply coming from hundreds of electricity generators, deciding which units will run and which ones will be idle, and making quick adjustments for the generators you can't schedule, like the wind machines and solar panels.

Hardly anybody will ever know you are here, unless you mess up. All is going smoothly until you get a message from a neighboring electrical entity requesting emergency assistance. A quick glance at your computer screen tells you that you have sufficient spare capacity to help."

What do you do?

Transmission Grid Complexity Example: San Diego, September 8th, 2011

Hot day in San Diego

- Demand greater than local supply
  - September considered shoulder season: fewer generators online than summer
  - But, this alone is not a cause for concern

- generation being imported from the north (Los Angeles) and from the east (Arizona through El Centro)

- power flowing north from El Centro to Imperial Valley

Transmission Grid Complexity Example: San Diego, September 8th, 2011

Three human events lead to blackout

1) Worker in western Arizona accidentally misses a step in maintenance on a capacitor: Southwest Powerlink crossing AZ/CA border trips off. Increased power flow from Los Angeles, and reversal of flow to be in from Imperial Valley.

Transmission Grid Complexity Example: San Diego, September 8th, 2011

Three human events lead to blackout

1) Worker in western Arizona accidentally misses a step in maintenance on a capacitor: Southwest Powerlink crossing AZ/CA border trips off. Increased power flow from Los Angeles, and reversal of flow to be in from Imperial Valley.

2) Human operators accepted unrealistic contingency plan in Imperial Valley, line trips off.

Transmission Grid Complexity Example: San Diego, September 8th, 2011

Three human events lead to blackout
1) Worker in western Arizona accidentally misses a step in maintenance on a capacitor: Southwest Powerlink crossing AZ/CA border trips off. Increased power flow from Los Angeles, and reversal of flow to be in from Imperial Valley.

2) Human operators accepted unrealistic contingency plan in Imperial Valley, line trips off.

3) Human operators unaware of 8000 amp limit on line from Los Angeles, line exceeds that (operators know and think it is ok) and disconnects. This San Diego island does not have enough local generation to meet demand. **Blackout!**

17
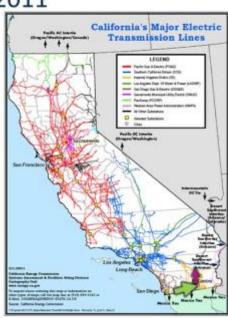
Transmission Grid Complexity Example: San Diego, September 8th, 2011

Blackout caused by accident, poor planning, and unknown grid limit.
- All very low probability events.
- No blackout if any of the 3 events didn't happen.

"Traditional" works have looked at vulnerable nodes/lines: what happens if this line fails. Limited stochastic analysis.

Challenge is to model (and validate!) human cognition and decisions coupled with low probability events.

18

Challenges in Electric Grid Modeling

Grid operator actions

- Many individual operators, incentives not always aligned
- Competition/interaction with other grid operators
- Shift to renewable, variable generation (wind/solar) introduces generation uncertainty

- Customer actions
  - Influenced by policy, but how do customers respond?
  - Customers acting different from expected, trying to game the system
  - Shifts in consumption behavior (electric vehicles, power walls, etc.)

- Low probability events
  - Blackouts
  - Operator errors/Incorrect information
  - Attacks (physical and cyber)
  - Extreme weather (Hurricane Sandy)

- Complex interactions
  - As distributed PV increases, power may flow from distribution grids to transmission during some times

Validation and UQ

Need to validate electric grid models:

- Quantitative
  - Accurately represent PV production/customer loads
  - Accurately determine voltage/current violations

- Qualitative:
  - **Convince grid operators that models are useful**
  - Ensure policies are working

Need to understand uncertainty:

- Penalties to utility for exceeding voltage/current limits; may lead to unsafe conditions, blackout, etc....how big is the risk?
- Suggested optimized state plus confidence: operator may choose to go against suggestion if confidence is low
- Threat analysis

Quantities/Questions of Interest

Quantity of Interest	Reason	Challenges
Electrical quantities: voltage, current, power	Ensure operation of electric grid is within allowable limits	Limited validation data, extreme events (e.g., power outages), evolving grid layout (upgrades, new PV)
Customer behavior: typical electric use, response to price incentives	Optimize operation of electric grid	Complicated human decisions, evolving behaviors
Resiliency of electric grid	Understand impact of threats	Resiliency difficult to quantify, extreme events (never happened previously), modeling recovery of the grid
Impact of renewable energy generation	Maximize amount of energy from renewable generation while maintain a safe, reliable, resilient grid.	Renewable generation is variable (weather), coupled human adoption/electrical impact, grid support from renewables (control and $)

Questions of Interest

What is the value of increased sensors/data on electric grids? – e.g., Does additional data reduce uncertainty and hence enable more renewable energy installations?

How does customer generation and storage (plus control schemes) impact electric grid resiliency and reliability?

What are best policies to safely increase renewable generation on the electric grid (incentivize installations, grid support, etc.)?

Electric Grid and Cybersecurity

Ukraine, December 23rd, 2015

From *Wired*, "Inside the cunning, unprecedented hack of Ukraine's power grid"

- "Inside the Prykarpattyaoblenergo control center,…as one worker was organizing papers at his desk that day, the cursor on his computer suddenly skirted across the screen of its own accord…He watched as it navigated purposefully towards buttons controlling the circuit breakers at a substation in the region then clicked on a box to open the breakers and take the substation offline."

- "Somewhere in a region outside the city he knew that thousands of residents had just lost their lights and heaters…All he could do was stare helplessly at his screen while the ghosts in the machine clicked open one breaker after another, eventually taking about 30 substations offline."

- "Ukrainian and US computer security experts involved in the investigation say the attackers overwrote firmware on critical devices at 16 of the substations, leaving them unresponsive to any remote commands from operators. The power is on, but workers still have to control the breakers manually."

- "That's actually a better outcome than what might occur in the US, experts say, since many power grid control systems here don't have manual backup functionality…"

https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/
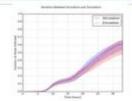
2.6   Kasimir Gabert: Cybersecurity Models
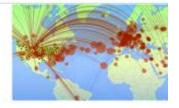
Cybersecurity Models

Complex Systems VVUQ Workshop 2016
Kasimir Gabert, Sandia National Labs

Outline

- **The Problem Space**
- Modeling Techniques
- Example Problems
- Verification and Validation Thoughts

Scope of Cybersecurity

- (Ambitious) Goal: eliminate surprise from our computers
- It is a large field, spanning the whole computer system space: from subtle software / hardware bugs to the motivation of cyber criminals to unintended radiation from physical devices

- Luckily, in order to be useful these systems are necessarily engineered to reduce surprise

3

Past (Surprising) Events

- January 2010 Operation Aurora
  - A series of advanced attacks first reported by Google and aimed at dozens of American companies. These attacks took advantage of previously unknown vulnerabilities in Internet Explorer and both exfiltrated intellectual property and accessed email of targeted users
- April 2011 Amazon Outage
  - A configuration error during an upgrade disconnected a large number of storage machines; once the error was fixed they all automatically tried to replicate, causing a re-mirroring storm, thread starvation, and a large system failure
- April 2014 Heartbleed
  - A vulnerability in the OpenSSL library was discovered allowing for arbitrary remote memory to be read, including private keys. This impacted a massive number of servers on the Internet.

4

Purpose of Cybersecurity Models

- Similar to cybersecurity: eliminate cyber surprise (using models!)
- We are asked to determine:
  - How a network or system design change might help or hurt (performance, security, usability)
  - How well a network can withstand or protect against an attack
  - Whether a new product will make a set of systems "more secure"
  - What an optimal deployment or design of a system change might be
- Some additional uses:
  - Help system developers prototype as they build
  - Help train individuals or teams
  - Dynamically explore a system with open-ended research questions

5

Outline
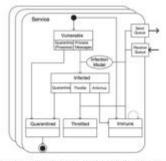
- The Problem Space
- **Modeling Techniques**
- Example Problems
- Verification and Validation Thoughts
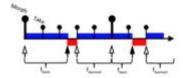
Differential Equation/Agent-based/ Game Theoretic Models

- Critically, these seem to require a thorough understanding of assumptions and important parts of the system
  - Can this claim be strengthened through validation?

"Agent-based modeling of malware dynamics in heterogeneous environments", Bose et al. 2011

"Evaluating Moving Target Defense with PLADD", Jones et al. 2015
PLADD: Probabilistic Learning Attacker, Dynamic Defender

$$\frac{dI(t)}{dt} = \beta I'(t)S'(t) - \frac{dR(t)}{dt}$$

"Modeling Botnet Propagation Using Time Zones", Dagon et al. 2006

7

Emulation-based Models

- Because we model computers using computers, the line between the "real world" and a "model" is blurry
  - A "real world" production virtual machine is a "model" element after being copied
- Emulation-based models ("Emulytics" at Sandia) take advantage of this blurry line by building models that consist mostly of virtual or physical machines running software pulled from the real world
- This approach can be augmented with simulation at a packet level or with the use of another (non-emulated) model

8

Model Assumptions / Parameters

- Numerous assumptions – known/unknown, implicit/explicit
  - For every device in the network (from computers to network cables):
    - The operating system or firmware, from the version to every specific configuration option in every running service
    - The hardware the device is running on, all the way from rough specifications (amount of memory, speed of processors) down to the specific motherboard, all of the integrated circuits on it, and their initial RAM states
    - The software running on top of the system and all of its options and compilation / configuration settings
    - The users' interactions with the device
  - Networks can have thousands of devices in them
  - Model-specific software, for example a tool that mimics a user or a change that artificially speeds up a download to save time
- Unlike atoms, a different parameter can cause a state change

9

Running an Emulytics Model

- The models are run (at clock rate) on computer systems
- We use virtualization platforms and numerous physical computers, each with their own complete set of assumptions and parameters
- A model may have significantly different output if it is oversubscribing the underlying hardware, experiences contention with other devices, or simply is not scheduled well across the physical cluster

- This adds a degree of non-determinism that does not seem to exist with many other models

10

Firewheel

- A platform developed at Sandia that eases the process of building, running, and studying these models
- It is another large code base, filled with numerous assumptions, many parameters, and plenty of bugs

FIREWHEEL

Outline

- The Problem Space
- Modeling Techniques
- **Example Problems**
- Verification and Validation Thoughts

First Example: Shadscale

- Project: Assist a customer in an annual tabletop cyber exercise by putting evidence behind discussed solutions
- Goal of the model: Given a concrete (invented) piece of destructive malware and a customer network, evaluate the different protections and mitigations the team comes up with

The Model

- Some considerations from the real world:
    - There are tens of thousands of computers on the network
    - Each has an operating system, physical hardware in some condition, various software installed, numerous real employee created data, etc.
    - There are many routers with unknown configurations, connected in a mostly unknown topology
    - There are thousands of employees
- We identified a single metric to evaluate mitigations with: Fraction of infected hosts over time
- To guide the model construction, we focused on 1) making sure we can output the metric and 2) making sure we can build each of the proposed mitigations

The Model Components
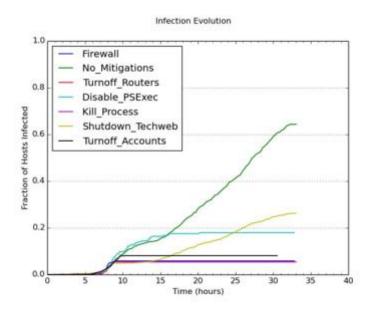
- Required by a mitigation:
  - Windows domain
  - Routers that can isolate subnets
  - Firewalls on network boundaries
- Required for the output:
  - Graphical interface automation in the Windows clients (the malware needs to be in the correct Windows security context)
  - Infection server replica
  - Required network infrastructure and servers to deliver and propagate malware
  - Simple user model (access per real log files, 20% click "Run")

Results

Infection Evolution

Shadscale Artifacts
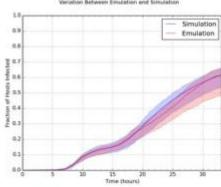
Sensitivity to Parameters
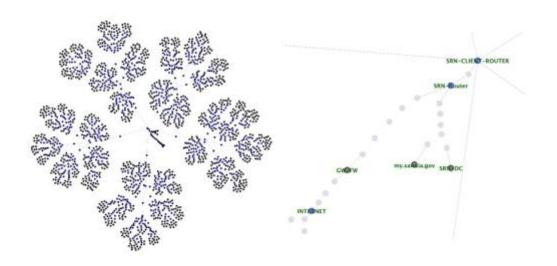
- A model this simple may be a good candidate for an initial validation pulling from traditional validation techniques
- The model (given that the artificial malware is so simple, even though it is peer-to-peer) seems to really only change given a different user model
- We built an agent-based model, erasing most parameters

Variation Between Emulation and Simulation
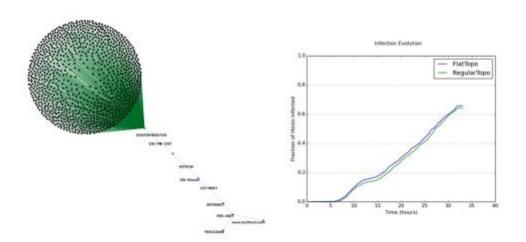
Topology Sensitivity
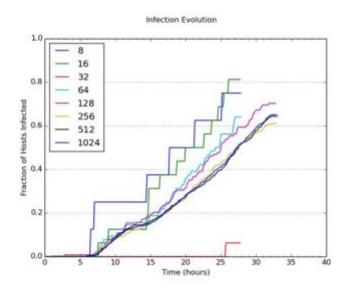
Topology Sensitivity
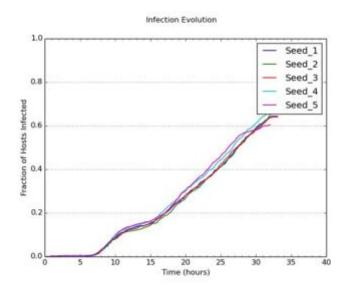
Topology Scale Sensitivity

Infection Evolution

Topology Scale Sensitivity

Infection Evolution
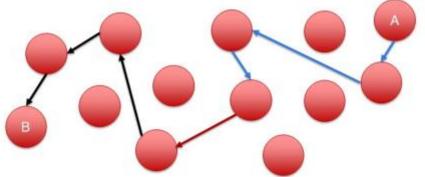
Second Example: I2P

- I2P (Invisible Internet Project) is a peer-to-peer overlay network designed to provide anonymous hosting
- Malware authors have moved communication servers into it—collaborating with Georgia Tech, we want to learn how to stop such malware

23

I2P Model Goals

1. Suppose one can temporarily turn off I2P traffic to an autonomous system on the Internet. How much would this have to happen to increase the tunnel ratios?
2. The I2P developers may be willing to update the code to ban the malware. Will all users who update isolate themselves from I2P?
3. It appears that a poisoned router entry could break I2P. (Testing this in the model showed that it is not the case.)
4. Given a small testing infrastructure, can we perform a larger population estimation by varying unknown model parameters?
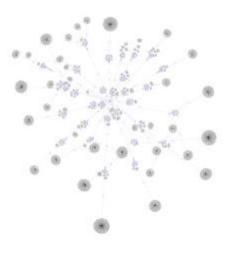
24

I2P Model

- I2P software installed on Ubuntu 14.04 desktop clients
- I2P bootstrapping using our own keys
- Network from reversing a research paper
- User behavior: an infinite loop browsing one website

25

Are These Valid?

- For Shadscale, how do we know that the mitigations would be as effective as the results say?
- For I2P, do we know that a poisoned router entry will do nothing?
- How can we know that our resulting population estimate bounds are reasonable?
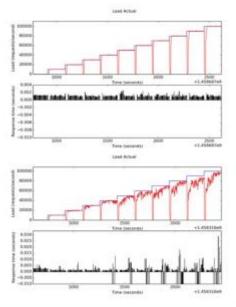
26

Outline

- The Problem Space
- Modeling Techniques
- Example Problems
- **Verification and Validation Thoughts**

Visive Verdict LDRD

- Work in progress
- Identified problems: massive number of parameters and difficult metrics / quantities of interest
- We are building small laboratory models of key components – hierarchical may work

28

Possible Validation Distinguisher

- As a general validation metric, use a Turing machine as a distinguisher between real and emulated
- D is a probabilistic Turing machine *distinguisher*, E is an oracle for the emulator, and R is an oracle for the real world
- Difficulty is in choosing the environment for D (the probability space and allowed accesses to the oracles)

- $Adv(D) = |\Pr[D(E) = 1] - \Pr[D(R) = 1]|$
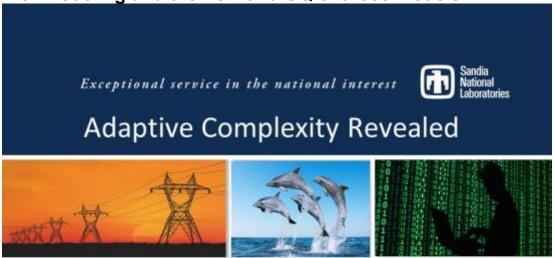
Possible Verification

- Software quality assurance seems to translate well
- For solution verification, can we compare against a "ground truth" oracle, one that responds with protocol descriptions or intended behavior (marketing literature?)?
- We may be able to build these verification checks into the models and allow them to run dynamically for each experiment
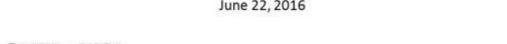
2.7 Curtis Johnson: Adaptive Complexity Revealed, with challenges for modeling and the V&V and UQ of those models

Goal of this Talk

- Lay out fundamental elements and dynamics of adaptive systems
- Illuminate modeling and V&V and UQ challenges with them

The Natural Selection Algorithm

1. New Candidates	Sexual Recombination and Mutation	Random Walk from Prior Successful Candidates
2. Test Candidates	Do they survive to reproduce? How many offspring do they have?	Those better adapted to current conditions will, on average, have more offspring
3. Repeat with New/Surviving Population		

Starting Point

- Stable, homogeneous environment
- Asexual reproduction
- Individuals are born mature and fertile, and stay fertile
- Unlimited resources
- No predation, no food chain
- No illness or accidental death
- Extremely abstract and unrealistic

4

Unbounded Growth

- Green population is immortal.
- Blue population lives to the birth of their great grandchildren

Populations are a simple function of fecundity and longevity. While green is growing faster than blue, they are not competing and are both unbounded.

5

Adding Copying Errors
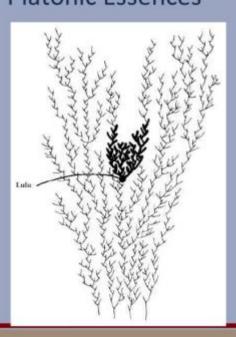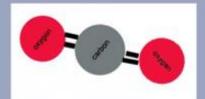
6

Types in Adaptive Complexity Are Not Platonic Essences

Lulu

oxygen carbon oxygen
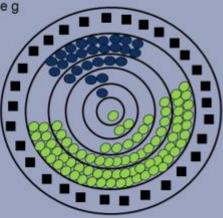
- Most species definitions have arbitrary components
- Because natural selection works by tinkering, individuals within a type can differ greatly
- More generally, the natural selection process rarely produces discrete types, but rather continuous variation.

7

Adding Scarce Resources

Assume the g-3 generation dies just before the g generation is produced.
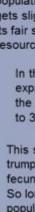
The scarce resource might be food, but it could also be other things. Let's imagine it's nesting sites.
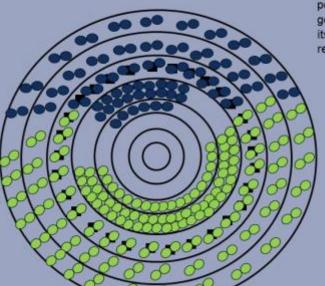
Assume an individual that acquires a resource will reproduce at its fecundity rate; otherwise it will not reproduce that generation.

8

Adding Scarce Resources
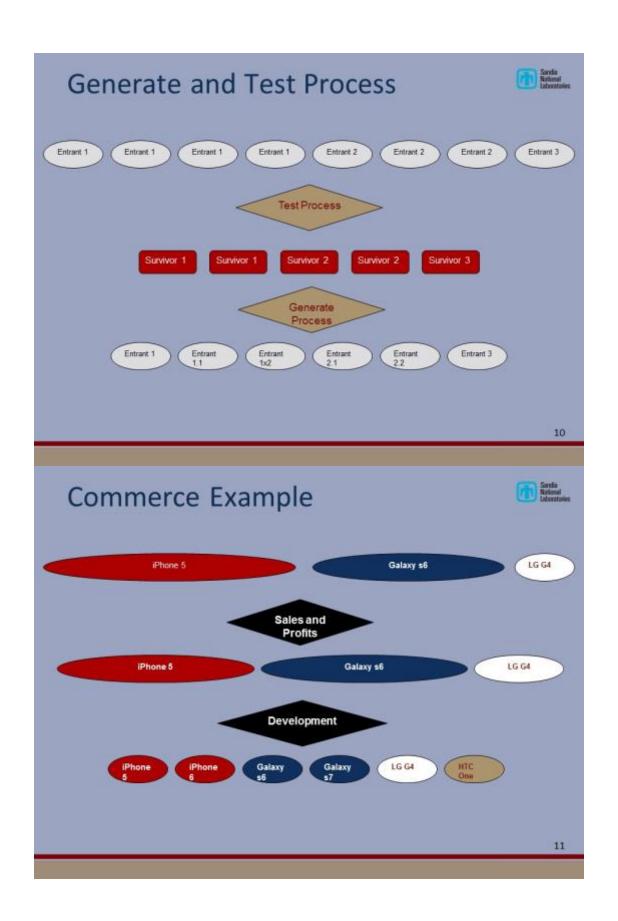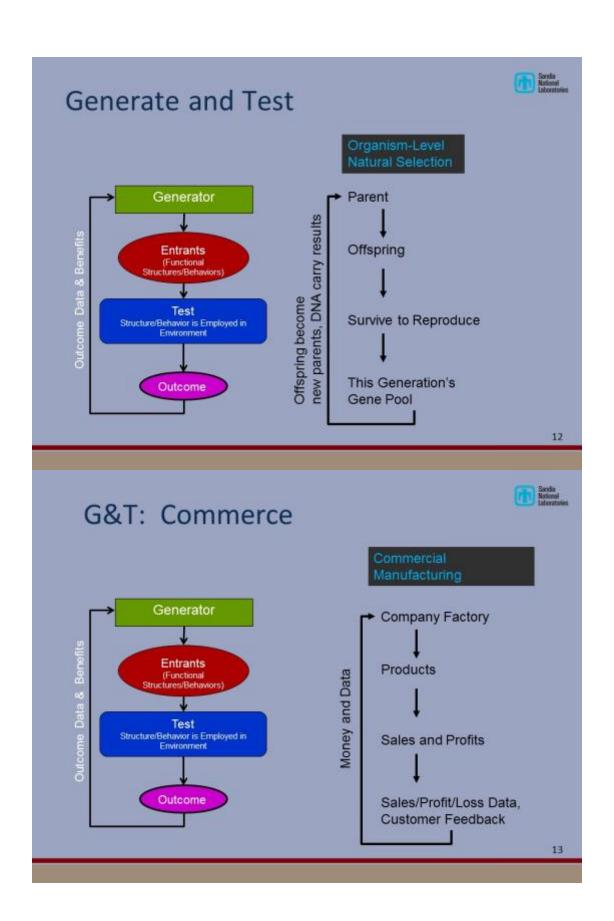
Assume the blue population consistently gets slightly more than its fair share of resources.

In this example, blue expands from 25% of the population in g5 to 38% by g8.

This slight advantage trumps start date, fecundity, and longevity. So long as neither population can produce more offspring per resource (and green cannot shift to another resource), green will go extinct.

9

Functionality is Born
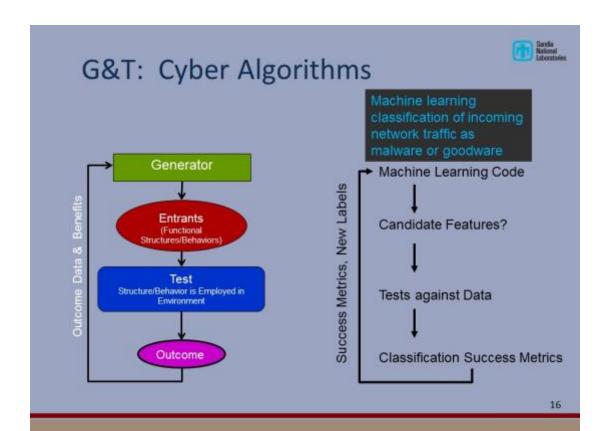
Generate-and-test feedback loops create and sustain functionality

- These systems are/appear goal directed in that they use flexible means to achieve an end
    - Functionality—means and ends, purposes—is not produced outside adaptive complexity
    - There are hierarchies of ends and means in adaptive systems
    - All but the ultimate (survive to reproduce) ends can change
        - e.g., when we added resource scarcity
- Generate and test cycle benefits include:
    - Novelty/Innovation/Creativity
    - Adaptedness (over time)
    - A life of their own: these systems work and 'strive' and adapt without an outside push
        - But also cancer
- Implications for modeling: optimization, etc.

14

G&T: Investment

Sandia National Laboratories

Investment

Outcome Data & Benefits

Generator

Entrants
(Functional Structures/Behaviors)

Test
Structure/Behavior is Employed in Environment

Outcome

Profits, Losses, Learning

Investor

Investments

Price changes

Profits and losses

15

G&T: Cyber Algorithms

Machine learning classification of incoming network traffic as malware or goodware

Generator → Entrants (Functional Structures/Behaviors) → Test: Structure/Behavior is Employed in Environment → Outcome

Outcome Data & Benefits

Success Metrics, New Labels

Machine Learning Code → Candidate Features? → Tests against Data → Classification Success Metrics

Adding a Variable Environment

- Over time and space
- Variation in the environment, including the biological environment, is what drives and preserves diversity and robustness/resilience.

Mutual Adaptedness of structure, function and environment (past-present)

- function
- structure
- Environment

- E.g., Controlled variables reveal function
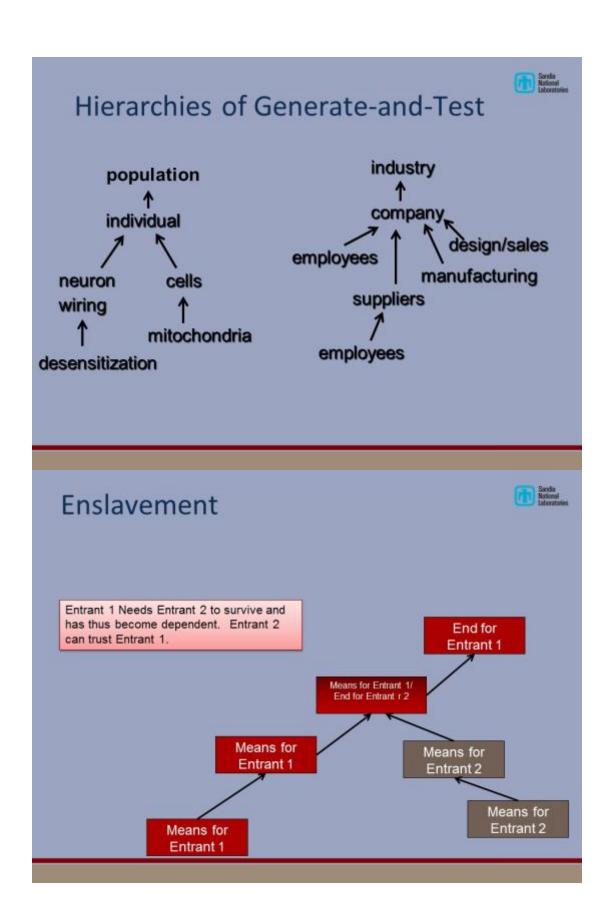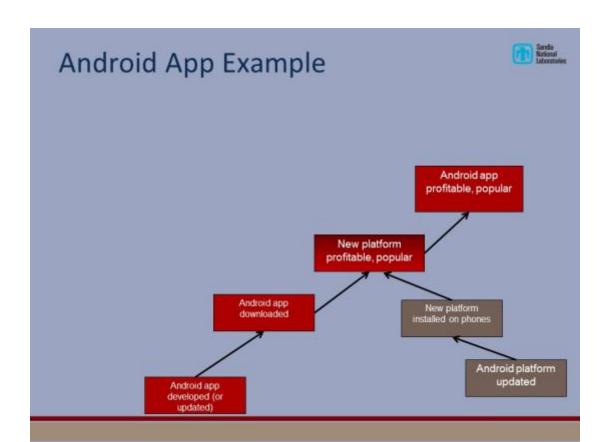
Latent Information and Functionality

- Adaptedness is only evident when it is needed
- I carry all the genetic material for a female, but my female traits aren't exposed in my body, and thus not acted on by selection while inside of me
- (But they popped back up in my daughter's body and lifespan)
- Homeostasis often works by mysterious means, and hides out-of-bounds behaviors

Hierarchies of Generate-and-Test

Enslavement

Android App Example

New platform profitable, popular
Android app profitable, popular
Android app downloaded
New platform installed on phones
Android app developed (or updated)
Android platform updated
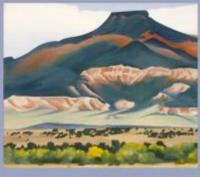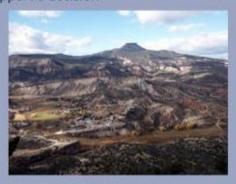
Models

Anything that 'computes' and is used to support a decision

Nothing is less real than realism. Details are confusing. It is only by selection, by elimination, by emphasis, that we get at the real meaning of things. --Georgia O'Keeffe

Minimalist Models	Representational Models
• Linear classifier	• Meshed physics simulation
• Clustering algorithm	• Agent-based model
• F=MA (as a complete model)	• System dynamics model
	• Emulation
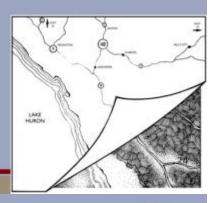
Challenges: What to Leave Out?

- This is a problem in all modeling, but compounded in adaptive systems:
  - Most types are heterogeneous (the boxes in the cyber model)
  - There are vast stores of latent information
  - There are vast stores of local information (each G&T cycle has its own)
  - There are important 'distant' connections
  - Threshold effects are common, so small changes often have outsized effects, and vice versa.
- Thus Butterfly Effects
- How can we improve validation establishing that nothing important for the use case was left out of the model?
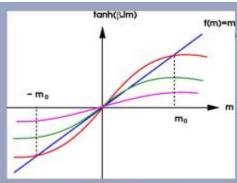
Challenges: Capture Critical Functionality

- Model homeostatic mechanisms or 'assume' them?
- Economic and social models focus on goals and functions
  - e.g., most stock market models don't model the actual trading system
  - Would you model teaching or parenting with physiologically accurate humans?
  - But when there is a complex engineered system, modelers often focus on representing it accurately, assuming the functionality is adequately captured
- What's missed?:
  - Power line owners are in business for profit—not to keep the grid up
  - Where is the malware purveyor's objective in the cyber model?
  - Where is the role of embarrassment or fear of punishment?
- How do we validate that a model pursues a goal, or robustly (but inconsistently) performs a function like its real-world counterpart?

25

'Mean Field' Approaches are Risky

- In non-adaptive models, it is frequently a good assumption that messy behavior is approximately random, or randomly distributed
- In adaptive models, slight tendencies can be critical (e.g., scarce resources example)
- Latent information is adaptedness—lying in wait for a triggering event. The change implemented is 'designed' to sustain functionality in the environment presumed to be signaled by the triggering event
- Nonetheless, 'mean field' approaches can be effective in adaptive complexity (e.g., Chicago school economic models)
- How do we know when we can safely assume a mean-field distribution?

26
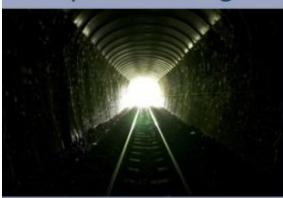
Basic Approaches

- Relatively fixed, detailed, descriptive model (e.g., system dynamics or agent-based)
- Radically generalizable heuristics (e.g., fight or flight, utility theory/satisficing, supply & demand)
- Emulate generate and test
    - Machine learning, etc.
- Multiple models at multiple scales, and/or with divergent methods. Can you get similar answers? Or ask the right question to the right model, feed that answer into the other, etc.?
- More effort to determine limits within which models are valid
- Ask different questions

27

Hope: Advantages for the Modeler

- Adaptedness:
  - Structure, function, and the environment are highly correlated.
  - There is good reason to think the past can predict the future
- Heterogeneity is limited by common functionality.
- Homeostasis,' or dynamic stability, can often be assumed

- Goals can often be known, or inferred confidently
- Goal/function-driven models, where appropriate, can often allow one to eliminate massive amounts of physical detail
- Goal/function-drive models can capture critical simplifying constraints (e.g., enslavement of apps by Android)

28

97

3. DAY ONE SEED RESEARCH IDEAS

3.1 Track Bins

Track 1: New VVUQ Theory	Track 2: Concepts in Complexity	Track 3: Decision Making & Applications
Facilitator: Mark A. Smith	Facilitator: Asmeret Naugle	Facilitator: Eric Vugrin
Daniel Appelo George Barr Matthew Denman Alfred Hubler Mark A. Smith Vicente Romero	Brian Adams Wei Chen Daniel DeLaurentis Jackson Mayo Bill Oberkampf Jessica Turnley David Stracuzzi	George Backus Dennis Engi Katherine Jones Elizabeth Keller Roshanak Nateghi Shreyas Sundaram

3.2 Seed Research Ideas and Presenter Short Bios

BRIAN ADAMS

Simple UQ for Complex Systems

Research Idea: Instead of demanding precise numeric output for VVUQ, can we instead take advantage of qualitative or subjective knowledge-of-interest type metrics to do more efficient uncertainty quantification and risk analysis? Do complex systems models' usage modes give rise to ensemble-free techniques for probabilistic analysis? How does that change our definition of a "parameter"? Yes, we need to model and understand complex systems, but when is a complex model required? Do we accept that "complex systems aren't amenable to reduction"? The systems themselves may not be, but risk analysis-relevant models might be. Can we purpose-build hierarchical reduced-order models that predict aggregate or integral knowledge-of-interest for uncertainty analysis, instead of performing detailed simulation? Need these be probabilistic to be effective?

Bio: Brian M. Adams is a principal member of technical staff in the optimization and uncertainty quantification department at Sandia National Laboratories. He holds a PhD in Computational and Applied Mathematics from North Carolina State University. Brian develops, implements,

and applies algorithms for optimization and uncertainty quantification on computational models. He leads the Dakota software project, managing software development and deployment to ensure impact on the span of Sandia science and engineering problems.

DANIEL APPELO
The Multi-Level Monte Carlo (MLMC) method

Research Idea: The Multi-Level Monte Carlo (MLMC) method has emerged as one of the most efficient and popular methods for forward propagation of uncertainty in many areas of scientific computing. The levels of MLMC have almost exclusively referred to geometrical refinement of some grid used to discretize the problem at hand but the MLMC framework allows for many other interpretations. We propose to explore, by examples, experiments and analysis, extensions of MLMC to more complex models where "the levels" as well as their accuracy and cost must be learned as a part of the computation.

Bio: Daniel Appelö is an Assistant Professor in applied mathematics at the University of New Mexico with expertise in numerical analysis and development of computational techniques for simulation of wave propagation problems. Prior to joining UNM Daniel was a postdoctoral researcher in applied mathematics and mechanical engineering at Caltech and LLNL. Daniel also holds a PhD in numerical analysis and a MS in electrical engineering from the Royal Institute of Technology in Stockholm Sweden.

GEORGE BACKUS
Risk: Uncertainty or Confidence?

Research Idea: Why do we make complex models and why do we study them? There is the important scientific/academic side, but, for Sandia, the focus is on a security problem and its mitigation, that is, the intervention options. By definition, for a complex model, there is, at best, a probabilistic statement of future dynamics. Therefore, the uncertainty on the predicted state of the model is more problematic and much less important than is the confidence that an intervention will prevent an undesirable state. This confidence is transitory as the system changes its evolution due to feedback from the intervention. Therefore, the intervention approach must also be a complex, feedback process. The metrics for risk management may be much more amenable to VVUQ methods than the metrics for describing the overall complex system.

Bio: George Backus (Sandia) focuses on the modeling/simulation of national security risks and the VVUQ of both human-behavioral models and highly-variable physical models, such as those used for climate change assessments. He utilizes many methods of analysis, including econometrics, system dynamics, and agent based models -- all of which look at the same world through different lenses, exposing different aspects of causality and interpretation.

GEORGE BARR
The 'standard' approach to Ecological, Epidemiological and Financial problems

Research Idea: The 'standard" approach to Ecological, Epidemiological and Financial problems which involve populations (integer units) is to assume that the continuum approximation can be applied (large enough population). This results in a first order differential equation (DE), which is then approximated using an explicit representation (e.g.Euler) to produce a set of recursion relations – the difference equations (diffE), along with time step constraints for stability.   [(dx/dt= f(x,t), dx/dt~ (x[i]-x[i-1])/□t)]

Some authors skip the differential equation step and simply write down *ad hoc* a set of recursion relations (then you have no time step constraints to worry after!)

Although the problems are integer or mixed integer/real number problems, only real number solutions are produced. ***There seem to be no published attempts to extract integer solutions from these integer problems***.

In fact, published solutions may ignore time step requirements for stability for DEs and ignore implicit time steps in recursion relations.

Question:
How can one estimate validation for a model describing integer populations when the analyses only produce real number values as solutions?

Proposal:
Pursue reformulation of "continuum approximation" analyses to produce integer solutions to these integer problems.

Possible Procedure:
  1) Use difference equation expansions as proper recursion relation models, but honor implicit time step constraints associated with actual problems.
      For example, in the recursion relation derived from the Logistic Eq'n, there is a mortality coefficient and there is a reproduction coefficient. For the dependent variable (number of critters) to change there is an implicit time step consistent with the smaller of the two coefficients(say mortality), but reproduction can not occur until there are a number of time steps consistent with the repro rate. Recursion equations, to be consistent with these implicit time steps need to be rewritten in modular form (See Modular Arithmetic --Wikipedia). And there is an additional constraint, namely changes to the dependent variable must be integer.
  2) Ad Hoc recursion relations (usually recognizable by the absence of the dependent variable at the previous time step), require similar analysis, but are a bit harder to unravel.

Result:
  1) Alternative mrthods of solution for comparison to Agent-Based and Cellular Automata models,
  2) Library of usable reference problems,
  3) Determination of the possibility of integer chaos
  4) Application of secondary techniques (from DE and diffEs) to constrain the analyses (e.g. Lypanov exponents, Attractors and Attractor basins, etc)

WEI CHEN
Designing VVUQ activities as an information seeking process

Research Idea: Understanding the coupling and the emerging behavior of a complex system is one top challenge in simulation based science and engineering. Complex systems research calls for rational approaches for decoupling (decomposition). While simplifications cannot be avoided, the challenge remains to assess whether a simplification is "safe," and the relative benefits of alternative simplifications (multifidelity models) may bring to engineering decision making or design. Validation, uncertainty quantification, and uncertainty management, can be viewed as an information seeking process. The complexity of an engineering design problem will be escalated to designing both the design artifact and the information seeking activities. This is a very complex decision making problem due to its dynamic nature, i.e., decisions made in an early phase will have a direct impact on the subsequent phases. Methods are therefore needed to manage such complexity.

Bio: Dr. Wei Chen is the Wilson-Cook Chair Professor in Engineering Design at Northwestern University. Her current research involves issues such as simulation-based design under uncertainty, model validation, stochastic multiscale analysis and design, robust shape and topology optimization, multidisciplinary optimization, consumer choice modeling and enterprise-driven decision-based design. Dr. Chen received her Ph.D. from the Georgia Institute of Technology in 1995. She is a Fellow of American Society of Mechanical Engineers (ASME) and an Associate Fellow of American Institute of Aeronautics and Astronautics (AIAA). She was the Chair of the ASME Design Engineering Division (DED) and is currently serving as a review editor of *Structural and Multidisciplinary Optimization*, the Associate Editor of SIAM/ASA *Journal on Uncertainty Quantification* (JUQ), and the Department Editor for the *IIE Transactions*. She served as a member of National Academy Committee on Mathematical Foundations of Validation, Verification, and Uncertainty Quantification.

MATTHEW DENMAN
Current UQ methodologies

Research Idea: Current UQ methodologies typically examine the statistical impact of input parameter uncertainties through computer models; often with Monte Carlo sampling. These methods do not produce uncertainty distributions which include uncertainties associated with:
- the fact that the model does not exactly recreate the validation data,
- scaling relationships used to apply models slightly outside of the validation data,
- extrapolations well outside the of the validation and scaling relationships.

Can the UQ community develop an integrated uncertainty measure that incorporates all of these sources of uncertainty?

Bio: Dr. Matthew Denman is an expert in simulation informed risk assessment, uncertainty quantification, and licensing concerns for nuclear power plants. He has contributed to the SOARCA and Dynamic Programing uncertainty quantification efforts using the MELCOR computer code. In FY15, he managed the nuclear power related efforts in the MIPM effort to develop and demonstrate the Integrated Cyber Physical Impact Analysis (ICPIA) Framework. He is the PI for dynamic human/system modeling research for the Nuclear Regulatory Commission

and is the lead modeler for the Fukushima Unit 2 accident reconstruction analysis and is currently investigating Bayesian calibration of input parameters to gain insights on Unit 2's accident progression.

DENNIS ENGI
Exploring the spectrum of solution perspectives in a quest for understanding the relative value of these divergent imperatives

Research Idea: At the national level, issues that are vital to the quality of life of our citizenry impact a variety of dimensions including human health, human rights, economic well-being, cultural heritage preservation, environmental quality, availability of services, and faith that our institutions will continue to serve us well. The solution space typically involves combinations of regulations; fiscal incentives; information, education, and outreach; technology development and deployment; inter/intra institutional relations; and enforcement. Opposing ideologues invariably argue that the "best" solution can be found and implemented most effectively by individual agents at one extreme or by a central authority at the other. In practice, the solutions that are actually implemented are drawn from both extremes as well as the myriad possibilities in between. No one -- not the opposing ideologues nor those in between -- can convincingly argue that the solutions that are implemented in practice are "optimal" or even close to optimal. The research suggestion is to explore the spectrum of solution perspectives in a quest for understanding the relative value of these divergent imperatives.

Bio: Dennis Engi has had a career as Senior Scientist at Sandia National Laboratories and as Professor and Head of the School of Industrial Engineering at Purdue University. He is currently College Professor of Engineering at New Mexico State University where he teaches a graduate level class in Complex, Adaptive, Intelligent System of Systems Engineering. His current research interests are in identifying, understanding, and managing issues that have significant impacts on our quality of life.

ALFRED HUBLER
The Role of Outliers in Verifying Models of Chaotic Systems and other Complex Systems

Research Idea: Most data compression tools in natural science are integral methods that describe averaged properties of the data set and ignore extreme values and outliers. Data compression with candle stick bars does the opposite, it focuses on the extremes: the high and the low are the extremes in amplitude and the open and close are the extremes in time. Only the extremes are kept, whereas the rest of the time series is ignored. Why is that? Is there a distinct difference between financial time series and time series from natural science that justifies using different approaches? Is it just the simplicity of the method that makes it so appealing, or are there deeper reasons that make the description of a time series with candle stick bars superior to data fitting? [1]
[1] A. Hubler, Are Candle Stick Bars a Good Tool for Data Compression in Natural Science? Complexity, Vol. 17, No. 1, 5-8 (2011)
Bio: Alfred W. Hubler received his *diplom* in 1983 and Ph.D. in 1987, *summa cum laude*, from the Department of Physics, Technical University of Munich, Germany. After a postdoctoral

fellowship at the University of Stuttgart, Germany, he came to the University of Illinois as a visiting assistant professor in 1989, and became assistant professor in 1990. Later that year, he also became the associate director of the Center for Complex Systems Research at Illinois, of which he is now the director. Professor Hubler served as a Toshiba Chair Professor at Keio University, Tokyo, in 1993-94.

Since beginning his thesis research, Professor Hubler has worked on nonlinear dynamics and has investigated a broad range of nonlinear phenomena. He is primarily a theorist, but he is also experienced in and capable of guiding both experimental and computational work. He has made solid contributions to the study of the chaotic dynamics in classical systems, both in idealized physical models and in engineering systems. He has been a pioneer in several important recent developments in nonlinear science research, including the control of chaos, the resonant coupling of nonlinear oscillators, and resonant stimulation and novel spectroscopies in nonlinear systems. Professor Hubler was among the very first to recognize that seemingly erratic, random motions associated with deterministic chaos could, in fact, be controlled, and that "chaotic" systems could be more "flexible" than systems undergoing more regular motion.

A skillful and committed teacher, Professor Hubler has also creatively applied the principles of nonlinear resonance to develop an intuitive, interactive web-based software package used to teach a variety of university science courses, at Illinois and around the world. Dubbed "CyberProf," the software analyzes student homework problems in real time and provides meaningful, intelligent, individualized feedback to each student.

KATHERINE JONES
Role of decision theory in VV/UQ of complex systems models

Research Idea: The representation of individual or group decisions made by humans in complex systems models can be particularly difficult to validate. These representations can be rooted in theory, statistical analysis, or subject matter expert elicitation. There are challenges to each of these approaches; for example, research around irrationality or bounded rationality reveals that traditional utility theory may not always accurately predict decisions, statistical approaches which draw priors from group behavior and experience can be limited by the effect of subjective probabilities on an actual individual decision maker, and subject matter experts often have trouble articulating or even understanding how and why they choose one action over another under a given set of circumstances. Behavioral economics (e.g. Prospect Theory) and decision sciences offer insights that can add realism to complex systems models. Since testing behavior under the exact conditions represented in the model is often not possible for national security applications and human studies often must be limited in scope, finding appropriate comparable experimental scenarios is critical. Perhaps there is a "right" amount of sub-optimal decision making that should be inserted into the model based on certain attributes (such as the number of possible state spaces/outcomes of a decision, the availability of supporting statistical data, or the amount and accuracy of information an agent/individual receives on which they must base their decision). As part of the validation process, targeted experiments should be conducted to test (separately) hypotheses about both group and individual behavior at decision points identified as critical via sensitivity analysis. Hopefully, some of these experiments could be broadly applicable to national security related complex systems models.

Bio: Katherine Jones is Principal Research & Development staff in the Operations Research and Computational Analysis department at Sandia National Laboratories. She has a Master's degree in Systems Engineering from the University of Virginia and an MBA from Duke University. Her research at Sandia focuses on risk analysis and decision support modeling, and has spanned a number of different domains, including nuclear surety, infrastructure analysis, defense, physical security, and energy. In addition to her eleven years at Sandia, Katherine worked at the Federal Trade Commission and McNeil Consumer Healthcare, a Johnson & Johnson Company.

ELIZABETH KELLER
Beyond the validation of UQ methods

Research Idea: Beyond the validation of UQ methods themselves, how can we improve the communication of uncertainty to policy/decision makers and stakeholders? What are strategies and tools for effectively depicting uncertainty and how do they vary depending on the audience and the timing and nature of the task/decision? How might the integration/collaboration of behavioral and decision science/scientists with data and simulation science/scientists provide useful insights?

Bio: Born and raised in New Mexico, Elizabeth Kistin Keller received her B.A. in Political Science and Latin American Studies as a Morehead Scholar at the University of North Carolina-Chapel Hill and her Masters and PhD in International Development Studies (political science, economics and anthropology) as a Rhodes Scholar at Oxford University. Before joining Sandia, Elizabeth spent several years working on economic development and natural resource governance in North America, Southern Africa and South and South East Asia. She currently works in Systems Analysis (Org. 159) where she focuses on global security studies, policy analytics and strategic foresight. She also serves as an adjunct professor in UNM's Department of Geography and Environmental Studies and an affiliate of the Utton Transboundary Resources Center.

JACKSON MAYO
Enabling V&V for Engineered Complex Systems via Resilient Design

Research Idea: To establish model-based confidence in system behavior under a larger set of conditions than we can exhaustively test or simulate, V&V and UQ rely on ideas of smoothness and bounded error propagation. This works for many physical systems because their behavior is (at least statistically) *reducible*: A sufficiently small change in input (model parameters, initial conditions, or forcing) leads to a small change in output. Therefore, we can use sampling, interpolation, etc., to understand the range of behaviors including scenarios not tested or simulated. This makes it feasible to analyze a physical model for its correspondence to the real system and for its conformance to design requirements such as safety and security. Complex systems, on the other hand, involve behaviors of interest that are generically *irreducible*. For scalably large information networks, no generic bounds on behavior are possible according to Turing's halting problem. Yet, "evolved" complex systems do show a degree of smoothness, predictability, and resiliency in their key behaviors, reflecting their adaptive origins. Resiliency is closely related to V&V and UQ because errors and uncertainties in a simulation can have effects similar to perturbations affecting the real system. Hence, we propose focusing on the

resiliency properties (either existing or desired) of complex system models to ground their V&V. This offers a particular impact on complex systems whose engineering does not tend to proceed adaptively, such as digital systems – including controllers in cyber-physical infrastructures. Here, standard programming techniques yield an unverifiable result, difficult to predict and highly susceptible to failures and attacks (a single bit change can radically alter behavior). Techniques promoting error damping and limited information propagation, including formal-methods-based design, can help build in resiliency for such infrastructures. Designing systems that permit meaningful bounds on behavior is a precondition for the goals of V&V.

Bio: Jackson Mayo received a Ph.D. in physics from Princeton University in 2005. He is a Principal R&D Scientist in the Scalable Modeling & Analysis Systems Department at Sandia National Laboratories in Livermore, California, and specializes in modeling of complex, nonlinear, and statistical phenomena. Particular interests include complex systems analysis, formal methods, and the resiliency and security of computing systems.

ROSHANAK NATEGHI
Ensuring the resilience of power infrastructure systems

Research Idea: Ensuring the resilience of power infrastructure systems is critical, since disruptions to these systems can bring about large scale outages leading to huge societal losses. Severe weather has been the leading cause of unplanned outages in the US (DOE, 2014). Sever-weather induced outages may increase in the future due to climate change (Resiliency Report 2014). However, the link between climate change and extreme events is still beset with uncertainties and historical records are poor predictors of climate extremes. The General Circulation Models (GCMs) are currently the most sophisticated tools for simulating the response of the global climate to increased concentrations of GHG in the atmosphere. However, there are still a lot of uncertainties associated with the GCM scenarios. Moreover, the coarse spatial resolutions of these models, diminish the utility of these projections for local planning. Several dynamical and statistical downscaling of GCM models are currently available; however, downscaling of these model outputs will further compounds the uncertainties associated with them. These uncertainties complicate long-term investment decisions to better plan for climate change at a local scale. There is therefore a critical need in developing sophisticated stochastic methods to better capture the link between climate change and extreme events and adequately account for the uncertainties. Anticipating the likelihood of potential future climate extremes will help identify proactive adaptation and mitigation strategies. It can also help update design criteria for more resilient power systems; which is important since historically the design of these systems have been based on standardized engineering practices; finely tuned to historical weather patterns that may not be appropriate for the changing climate.

Bio: Roshi Nateghi is an Assistant Professor with joint appointments in the departments of Industrial Engineering and Environmental and Ecological Engineering at Purdue University. Her areas of research include modeling sustainability and resilience of energy infrastructure systems and performance analysis of critical infrastructures under climate change.
She received her MEng degree in Mechanical Engineering from Imperial College London in 2006. She then completed her MSE and PhD degrees in the Department of Geography and Environmental Engineering (DoGEE) at Johns Hopkins University. She was a National Science

Foundation Science, Engineering and Education for Sustainability (NSF SEES) fellow from 2012 to 2015; holding joint appointments at Johns Hopkins and Resources for the Future.

VICENTE ROMERO

Data Science Test Case for Calibration and Validation of Models of Complex Adaptive System Resilience

Hypothesis

It seems possible that formalized calibration and validation of models of Complex Adaptive Systems (CAS) could be accomplished to a useful degree if "non-traditional" tools are employed such as Big Data mining and analytics (BDMA).

By 'non-traditional tools' I speak from my realm of familiarity with computational physics applications of engineered systems for which adequately revealing experiments can be designed and conducted, and where relatively well-developed approaches for measuring experimental input conditions and system responses are available for effective model calibration and validation.

Although there is continuing need to develop better measurement and inference systems for more fully characterizing experimental inputs and outputs in this realm, and affordability of experiments and diagnostics limits what can be done practically in any given situation, the methodologies and infrastructure for empirically based model calibration and validation in this realm are much further along than for CAS type systems (e.g. natural, cyber, electric grid, battlefield) that can't be tested (or adequately tested) under controlled conditions or the conditions cannot presently be adequately characterized.

But these CAS systems operate in environmental conditions that could potentially be characterized to a useful degree by Big Data mining and analytics. Likewise, BDMA could potentially extract direct output measures and/or indirect signals of system behavior under identified environmental drivers that could enable useful calibration and validation of CAS models.

Motivating Illustrative Example

This example also features a Resiliency element of Complex Adaptive Systems that Sandia wants to explore.

A few years ago, a deep three-year drought in New Mexico required the Middle Rio Grande Conservancy District (MRGCD) to borrow significant water from the Bernalillo County Water Authority (BCWA) to enable middle Rio Grande farmers to irrigate their crops. In recent hearings over large-scale planned community developments on Albuquerque's west side, BCWA stated there is ample water for the planned developments. These statements have been hotly disputed by various groups, and the questions arise: If BCWA fully allocates its water for future development and has no reserve to buffer MRGCD against future instances of deep drought, might permanent damage occur to the middle Rio Grande agricultural economy, and to what extent, and what would be the implication to the larger NM economy and the socio-economic stability in the middle Rio Grande corridor?

Defensible analysis of the dynamics involved, including perhaps projected trends and uncertainties from climate change, could advise state-wide policy and planning on these important and consequential issues. Defensibility could perhaps be enhanced significantly by calibrating or validating models of the involved dynamics against the recent four-year drought agro-economic experience in California's central valley. This would require focused Big Data mining and analytics and other model calibration, validation, and UQ methodology and infrastructure developments on a challenging scale.

Bio: Vicente Romero has been with Sandia National Laboratories for 29 years. He is in the UQ, V&V, and Simulation Credibility group in the Engineering Sciences Directorate. He has a modeling background in optical, thermal, fluid, and structural systems, specializing in complex coupled systems and applications where statistical or stochastic behavior is important. Dr. Romero also has extensive experience in developing and applying optimization and uncertainty quantification techniques for model calibration, validation, and risk assessment and reduction in nuclear weapon systems subjected to stressing thermal-mechanical-electrical-radiation environments.

MARK A SMITH
Pluralities as Truth Surrogates

Research Idea: A common problem when trying to validate a complex systems model is the lack of real-world results to compare it with. Real data may depend on the natural, random occurrence of uncertain, rare events, or it may not be possible to perform the necessary experiments due to lack of money or time or because the experiments would be dangerous, unethical, or illegal. One way around this problem is to use the results of multiple independently developed models as cross-surrogates for real data. The research task then is to develop and formalize processes whereby a plurality is formed between several models addressing the same problem and somehow use that plurality as a surrogate for ground truth for validation purposes. This is reasonable because each modeling team will have *some* intuition and tacit information about the real system in question, and good assumptions are likely to be positively correlated while bad assumptions are hopefully uncorrelated. Thus, increasing the number of independent models (or independent submodels or independent raw assumptions) will increase the signal to noise ratio of the results. Seeing a mode of agreement between model results would go a long way to giving confidence to both the modelers and model users. Of course there is always a possibility of making a mistake when model users make decisions, but generally speaking, would they be better off if they didn't do any modeling at all?

Bio: Mark A. Smith has been working in mathematical analysis and computer modeling of complex systems for over 25 years. He has a Ph.D. in Physics from MIT for research performed in the Information Mechanics Group in the MIT Laboratory for Computer Science using special-purpose cellular automata machines to create abstract models of physics. Since then, he has worked as a consultant on system dynamics modeling of public universities, bioinformatics of the *Drosophila* genome, and computational techniques for modeling social systems in central Asia. He came to Sandia as a security vulnerability analyst in the Physical Security department where he performed modeling and simulation to evaluate protection strategies for high-value assets against a variety of threats. He is currently in the Complex Systems for National Security

Group at Sandia where he works with a broad spectrum of customers on decision support involving reliability engineering, system-of-systems analysis, energy security, and operations research.

DAVID STRACUZZI

Exploring how analysis of the data produced by complex systems, including the associated uncertainty and value of information, can be used to improve underlying system models and performance optimization methods

Research Idea: Quantification and propagation of uncertainty through the equations that govern physical and engineered systems has long been a staple of Sandia's work in complex systems. Recent work in the data sciences considers a related problem: quantification of uncertainty in data analytics. Unlike methodologies such as QMU, uncertainty in data analytics has no partial differential equations to describe interactions. Instead, the goal is to quantify how well the data speaks to questions of interest, such as "did event X occur at location Y?" Put another way, uncertainty quantification in this context determines the *value of information* (VoI) with respect to a target question. In turn, VoI provides a metric for evaluating and optimizing complex systems with respect to the *utility* of the system's output. For example, one could optimize a sensor constellation to collect data that minimizes decision uncertainty instead of focusing solely on factors such as sensor scheduling constraints and collection priorities. The research idea is then to explore how analysis of the data produced by complex systems, including the associated uncertainty and value of information, can be used to improve underlying system models and performance optimization methods.

Bio: Dr. David Stracuzzi has a background in artificial intelligence and machine learning with most of his recent work focused on analyzing the uncertainty in sensor data. He earned his Ph.D. in Computer Science from the University of Massachusetts at Amherst in 2006, followed by a postdoc at Stanford University, and three years as research faculty at Arizona State University before coming to Sandia. He is currently leading both LDRD and customer-funded projects in uncertainty quantification for data analytics.

Shreyas Sundaram

Understanding the Impact of Human Decision-Making on Robustness and Security of Complex Interdependent Systems

Research Idea: The functionality, reliability, and security of engineered systems depends critically on the way that they are perceived and used by humans. In order to design resilient systems, it is therefore necessary to understand the "human element." The language of game-theory offers a mathematical framework for reasoning about decision-making in shared systems, but existing approaches focus almost exclusively on modeling the players using classical models from "Expected Utility Theory." In contrast, studies in behavioral psychology and economics over the past few decades have shown that humans consistently deviate from such classical models of decision-making. Therefore, new theory is required in order to understand the implications of behavioral decision-making on the robustness and security of systems. In this Research Idea, we propose to develop a mathematical framework based on "Prospect Theory" to rigorously analyze the implications of human decision-making on security investments in

systems. Prospect Theory, which won the 2002 Nobel Prize in economics, suggests that humans perceive probabilities in a skewed manner, typically underweighting the probability of highly likely events and overweighting the probability of highly unlikely events. Humans also perceive gains and losses differently: a loss of $1 may be viewed as a loss of $k, where k is some number that varies between individuals. Such misperceptions of probabilities and values will have a clear impact on the way humans decide to use systems and invest in security precautions. We propose to characterize this impact by studying a class of interdependent security games, where each node in a network is controlled by a separate decision maker who is choosing security measures for that node. The attack risk faced at each node depends on the security level at that node and at other interdependent nodes (capturing the ability of attackers to tunnel into other connected systems once they have infiltrated a given system). We will identify the landscape of security investments that occur under prospect-theoretic decision making, and propose approaches to design networked systems that mitigate and account for security vulnerabilities that arise due to behavioral risk perceptions. Along these lines, we also propose to investigate whether the framework of Prospect Theory can be used to reason about how humans trust (or mistrust) models of complex systems, and the corresponding implications for human utilization of those systems.

Bio: Shreyas Sundaram is an Assistant Professor in the School of Electrical and Computer Engineering at Purdue University. He received his Ph.D. in Electrical Engineering from the University of Illinois at Urbana-Champaign in 2009, and was a Postdoctoral Researcher at the University of Pennsylvania from 2009 to 2010. He was an Assistant Professor at the University of Waterloo, Canada, from 2010 to 2014.

He received a 2016 Air Force Research Laboratory Summer Faculty Fellowship, and the Ruth and Joel Spira Outstanding Teacher Award at Purdue. At Waterloo, he received the Department of Electrical and Computer Engineering Research Award in 2014, the University of Waterloo Outstanding Performance Award in 2013 and the Faculty of Engineering Distinguished Performance Award in 2012. He received the M. E. Van Valkenburg Graduate Research Award and the Robert T. Chien Memorial Award from the University of Illinois, both for excellence in graduate research, and he was a finalist for the Best Student Paper Award at the 2007 and 2008 American Control Conferences. His research interests include network science, fault-tolerant and secure control of large-scale dynamical systems, game theory, linear system and estimation theory, and the application of algebraic graph theory to system analysis.

JESSICA TURNLEY
Replicability of models and the impacts of confidence in and interpretation of results

Research Idea: A model does not include everything because then, of course, it wouldn't be the model but would be the thing itself. How do we justify the bounding of the problem such that we pick some processes and things from the real world to include but not others? Sensitivity analyses can tell us what happens if we eliminate something from the model – but if it wasn't included in the first place, we can't do a sensitivity analysis on it. How does the composition of the modeling team (by discipline, personality, experience, etc.)/project sponsor/etc. affect this selection? This begs the question of the replicability of the model, which impacts confidence in and interpretation of results.

Bio: Dr. Jessica Glicken Turnley is President of Galisteo Consulting Group, Inc., a consulting firm in Albuquerque, NM and a Senior Fellow at the Joint Special Operations University, US Special Operations Command (USSOCOM).   As a cultural anthropologist, Dr. Turnley brings a unique perspective to her work in national security, policy analysis, organizational development and culture change, and community-based environmental decision-making and economic development. Her technical work includes the development of computational social simulations, analyses of the development and use environments of national security technologies, explorations of the socio-cultural dimensions of security and policy questions, and analyses of military structures and research institutions. She has served both as a consultant to and as an on-staff technical manager at Sandia National Laboratories, and worked directly with many elements of the Department of Energy and other members of the national security community. She has worked on various projects for USSOCOM, for other military services, and directly for various parts of DOD. Dr. Turnley has worked with the intelligence community, including service on the Defense Intelligence Agency's Advisory Board. Dr. Turnley also works in community-based environmental decision-making, supporting the EPA and others. She also works in technology-based and microeconomic development. Dr. Turnley holds a Ph.D. in Cultural Anthropology with an emphasis in cultural linguistics from Cornell, a M.A. in Social Anthropology from the University of Michigan, Ann Arbor, and a B.A. in Anthropology and English Literature from the University of California, Santa Cruz, and has served as a Fulbright Scholar in Indonesia.

DANIEL DELAURENTIS

Test and validation of autonomous, learning human-machine teams

Research Idea: The challenge is to provide means for verification and validation of a complex system of systems of humans and machines (e.g., aviation system of systems ...vehicles plus operators plus ATM function) massive in its heterogeneity (vehicle type and development timescale, autonomy level and sophistication, operating mode) and scale (number of systems, number of operating modes, number of communication transactions). Such heterogeneity makes a purely formal methods approach, and a purely simulation-based approach, completely impossible. A new paradigm is needed to verify and validate are key system of systems properties that relate to safety, reachability in system performance, and degree of controllability.

Bio: Dr. Dan DeLaurentis was appointed President's Fellow for Defense Initiatives in June 2015, serving the EVPRP Office by aligning, preparing, and following through on partnership opportunities with the Department of Defense and related agencies. He also retains his roles as Professor in Purdue's School of Aeronautics & Astronautics and Director of Purdue's Center for Integrated Systems in Aerospace (CISA). Dr. DeLaurentis leads CISA's active research activity with the Missile Defense Agency's Enhanced C2BMC program developing agent-based modeling and simulation for development of advanced battle management architectures. He is also principal investigator on a cooperative agreement research grant with NASA's Systems-wide Safety and Assurance Technologies (SSAT) project focused on assessing safety impacts for distributed separation assurance approaches. His primary research interests are in the areas of problem formulation, architecture modeling and robust design and control methods for aerospace systems and systems-of-systems (SoS). This includes agent-based modeling, network theory,

optimization, complex system analytic methods, and aerospace vehicle modeling. His research is conducted under current grants and contracts from NASA (aeronautics and space organizations), FAA, Navy, the DoD Systems Engineering Research Center UARC, Missile Defense Agency, and several industrial firms. Dr. DeLaurentis has over 150 refereed journal/conference publications papers and presentations at/in major national and international venues. Dr. DeLaurentis is a two-time recipient of Purdue's Seeds for Success Award, awarded for executing research contracts in excess of $1M during a single academic year.

Dr. DeLaurentis is the co-lead of the Enterprises as Systems and System of Systems Thrust Area in the DoD's Systems Engineering Research Center (SERC) where he leads a project to develop an Analytical Workbench for system of system (SoS) architecture analysis, design and evolution. He also is a member of the SERC's Research Council. Dr. DeLaurentis is an Associate Fellow of the American Institute of Aeronautics and Astronautics (AIAA), past Chair of AIAA's Air Transportation Systems (ATS) Technical Committee, and presently serves as Deputy Director, Aircraft and Atmospheric Systems Group, AIAA Technical Activities Committee (TAC). He is also a member of the IEEE and INCOSE and active in "systems circles" of those societies. In 2012, Dr. DeLaurentis was awarded the CT Sun School of Aeronautics and Astronautics Excellence in Research Award, the highest research honor of the School. In 2012, he also was elected Faculty Fellow of CERIAS (Purdue's Center for Education and Research in Information Assurance and Security). He was a 2014 National Security Forum Fellow, Air War College, Maxwell Air Force Base and in 2014 appointed to the Complex Systems Capability External Advisory Panel (CSEAP) of Sandia National Laboratories.

3.3   Notes from Track Session 1 on New VVUQ Theory (Facilitator, Mark Smith)

Participants: Mark Smith, Vicente Romero, Alfred Hubler, George Barr, Nathaniel Brown, Kasimir Gabert, Daniel Appelo, Laura Swiler, Steve Kleban, Walt Witkowski, Jeff Tsao, Linas Mockus, Byron someone—an intern
Note taker: Emma Johnson

Intros:

Hubler: complex systems = systems under large stress or large throughput
Laura: Would like to see validation when we have no data or very limited data, cases where we don't' know what we are modeling—can't assign likelihoods to outcomes.
Jeff: Interested in how science evolves, how is credibility built—how can we build on this parallel to VVUQ

Mark Smith Idea:

Problem: Many cases where we don't have controlled experimental data to validate CS models—rare random events, can't afford experiments, don't have time to do them, experiments unethical, otherwise impractical.
Idea: Given problem, teams to develop models independently, then use those models/ideas as surrogates for each other.
Each team knows something—tacit knowledge, intuition—and will make some good assumptions. Hope good assumptions are positively correlated, bad assumptions uncorrelated, increase signal to noise ratio.
Goal: Use a measure of the plurality of results as ground truth—use as surrogate for experimental results.
Thing to do: Develop theory to formalize processes to use this plurality of models.
Seeing agreement would build confidence in models (for users)
Risk: Users could make mistakes—probably still better than not modeling at all.
There are different approaches to same problem (disciplines) are all reasonable lines of attack (ie physical security), hopefully you would start seeing commonality between these approaches.
Could imagine models as different levels of abstraction.
How would you go about this research? Understand work on wisdom of crowds. Mathematically or otherwise capture benefits of diverse

Comments:

Alfred: Reminds him of grading test in high level course. Truth of students' solution looks better. Mr. Sony snr always had several groups working separately on same problem—very successful. For his research, has students work independently on research—don't share.

Laura: How do you account for all models being biased? ie global temperature. Even 20 models can all be missing one thing—they are all wrong. All these different models are not truly independent.
Downside: expensive to do this—lots of models.

Vicente: Would shy away from claim that you can establish ground truth this way. Agrees that two teams are always better than one.

The man who has one watch knows what time it is, but the man who has two watches doesn't know what time it is—divergent answers are actually really valuable information as well. But yes—they could all be wrong also. Systematic error a barrier to ground truth.

George: Crowd sourcing—make model a game. Thousands of folks would be working on problem, you can introduce the constraints. This is happening at Stanford—proteins. The initial coding is online—you just have to constrain problems, figure out how to do that.

Linas: Could you couple the multiple models, feedback between models could help develop more accurate models.

Vicente Romero Idea:

His background: complex physics, edge of understanding in terms of environments, risk analysis. Large amount of evolution of system from beginning to end of event (heat, chemical composition in nuclear subsystems), coupling between thermal, mechanical, and fluids
Current Process: Model, try to assess uncertainties, make predictions, look at thresholds for performance of safety. So this is a bottom-up approach. Material and phenomena level leads to component level to system level. You develop models for the fire, for thermal and chemical composition, etc.

Idea: Could think of top-down approach. Start from system-level observables and understanding of working of components you could propose a structure for models and then try to optimize or represent those parameters in model
Most models have some combination of top-down and bottom-up. But today we have heard more about top-down complex systems. That's often all you can use.
Question: How can computational physics modeling and V&V methodology for engineered systems help with VVUQ. Parallel of fire boundary condition for weapon and drought boundary condition for agroeconomic system. Could talk about degradation of system, failure modes, failure probability (resilience of system). What are the similarities between these two types of systems? What strategies can we bring over from the physics systems to approach other types of complex systems? What is the transferability between systems (of fairly concrete effects)? In constituent models, there is a fairly direct transference.
Where do we draw the boundaries of the analogy that everything in life is an evolving experiment.

Comments:

Mark: This seems to be a common trend in complex systems to try to find exemplar systems.
Kasimir: Reminiscent of Curtis' talk: trying to find unified theory of complex system.
Laura: Bottom-up will work in many cases—Kasimir's for example.
Kasimir: Top-down data is impossible to come by in cyber—no one will tell you.

Daniel Appelo Idea

Multi-level Monte Carlo: Computing an expected value by Monte Carlo sampling. Don't know how to calculate quantity of interest, use surrogate.
Compute sample average many times.
$E[Q\_exp] = 1/N \sum\_1^N \{Q\_EXP^i\}$
Now choose cheap model.
$= E[Q\_EXP + Q\_CHEAP - Q\_CHEAP] = E[Q\_CHEAP] + E[-Q\_CHEAP + Q\_EXP]$
If difference between cheap and expensive is small, then you can pay less. Then if you have a fixed budget, all you have to do is optimize the number of course samples and the number of finer expensive method.
This relies on knowing accuracy and cost of models a priori—could you do some sort of machine learning scheme or something where you had a test period and extrapolated from there.

Comments:

Laura: They have been trying this for physics model: Thinking of low and high as two different resolutions, and low has to be pretty accurate
Daniel: We need data which suggests next data to reduce uncertainty.
Linas: The data thing is a lot to ask… Simulation as verification of this
Vicente: Is this useful for tail probabilities?
Laura: There has been some work on variance.

George Barr Idea:

We approximate problems using first order Des. We have a long history and bias in continuum problems. These problems involve integers & units. We output real numbers, not integers.
Validation: How good is your analysis? How do you know it is that good? The answer to the first question is we don't know.
4 study problems:
1. Logistics equation
2. Predator prey equation
3. Ebola model
4. Tribolium model—beetle used to eat huge amounts of wheat flour.

If you aren't getting integers solutions, how do you know that your results have any meaning—ie. Chaotic behavior.

With time-step constraint in logistics equation, you don't get chaotic behavior. Answers people are getting are recursion relation unrelated.

Implicit time-step in recursion relation: example: owls eat 5 lemmings a week. Implicit time-step=week. Another implicit time-step: gestation period for owl and lemming. So this is a modular arithmetic problem here.

Need to introduce the timestep when you have DE.

Idea: Examine a number of these kinds of problems (ecological, epidimeological, etc), rewrite the difference equations respecting the different time-steps and construct a library of reference problems with inputs and outputs that are integers. Then you can start to see if there is integer chaos that shows ups. When you have DE model you can ask how good can analysis be

Comments:
Linas: Other ways to discretize?--could be handled by orthogonal something
Kasimir: Emulation models: give up on trying to understand how something works and run it as is. George says that two methods have to agree.

Linas: Extrapolation in V&V?
Vicente: That's a very active area of research. Development of test problems and variety of test problems is what we need.
Laura: Blur between interpolation and extrapolation, in joints problem, calibrate under sinusoidal loads and then validate in damped condition.

Alfred Hubler idea:

Most of his students go into predicting stock market.
Finance world: candlestick data compression: low, open, close, high—throw everything else out. Only use the extremes!
Model human heart—try to predict what the doctor would do. What is the maximum deviation between norm and the patient—this was most useful in predicting the doctor. Small deviations don't have much information. Matches candlestick bar.
Which kind of data compression should I do? You only have to keep extreme points.
Characteristic for chaotic systems that you have to keep extremes. Maybe this is something fundamental.
But Q: Which measure should we use to do verification of the model?

Comments:

Vicente: Validates on 2.5 and 97.5 percentile—mean doesn't tell you much.
Jeff: If you have a new theory you test it on some weird case, not on the normal theory.
Laura: How are you defining chaotic system? This presupposes chaos, right?
Alfred: Yes chaos. And extremes seem to be good. Could be good because of distribution functions

Alfred has other comments based on talks:
Bound something extreme in space. How do we do a control? Control the extremes. Another research idea: Predicting and control social and antisocial system.
Antisocial=asocial (don't care about neighbors, behavior determined by internal state)

F=ma for rigid bodies. Bouncing ball has vibrational mode—best model a difference equation. When top and bottom have same timestep then you get chaos. F=ma is mapping function.

Doesn't apply for soft objects. So chaos is reality. Time and space discrete models are the best predictors for macroscopic systems (continuous ones). (Averaging at level of vibrational mode)

Daniel: It's impossible to discretize here. You can't do that many computations per second.
Alfred: the true physics is at this discretization.

Alfred: You use models to control. Run simulation real time—synchronize real system and model through (weak) bidirectional communication. This way you can control real system. Example: We synchronize with our virtual images. Resonance works between linear systems and works even better between nonlinear system. Two mechanical clocks on a wall synchronize.

3.4 Notes from Track Session 2 on Concepts in Complexity (Facilitator, Asmeret Naugle)

Participants: Asmeret Naugle, Marcy Hoover, Jessica Turnley, Jackson Mayo, Curtis Johnson, Dean Jones, Anne Lilje, David Stracuzzi, Wei Chen, Bill Oberkampf, Steve Verzi, Daniel Appello, Jim Stewart, Brian Adams, Pat Finley

Note taker: Thomas Kajder

Brian Adams:
- How are these complex models used? -> How will be decisions be made
- How UQ will help decision makers?
- How can we embrace UQ in the interest of "soft" decisions or fuzzy decisions?
- How certain do we need UQ because of inherent uncertainty in decision?
- Can we simplify the UQ to +- some quantity for efficiency's sake?
- For the purpose of risk analysis we may be able to use simple models – the model may not be reducible but the error might be.
- When can we decompose a complex model without losing accuracy (within limits) and replace complex simulators with simpler ones?

Jessica: There is a difference between reductionism and abstraction -> abstraction is a more accurate way we can simplify systems

Bill: On validation hierarchy – when you claim you can predict something it is based on extrapolating simpler systems to higher order interactions. How do you propagate errors up through systems? Failing is in the developer's reduction of complex systems being incorrect – physics that you have never tested in your model. We must go top->down, not bottom->up.

Dean: What are we after with customers? Confidence – despite data holes – with proper math and the results are appropriate. There are systems we can be more precise. We can do too much – how do we put quantitative metrics in a proper framework. How confident are you -> go into a geeky UQ -> customer looking for real confidence

Jessica: Customers aren't geeks (generally). Geeky UQ is not the language they speak.
Curtis: Daily battle on triage analytics for all of internet – we have systems that hand human analysts too much – we want to increase our strategic advantage. Doing better is really what is important for our customers. Dean will build model for a year that lasts years – we have short term model but still want VVUQ.

Dean: My problem is lack of data to Curtis lack of time. How do we VV in real-time and VV in acquisition of new data.

Wei Chen:
- View VVUQ as information seeking process that can be designed
- Perhaps not as interesting for social, but for physical
- The uncertainty keeps reducing in physical model over time

- Aleatory: due to natural uncertainty
- Epistemic: data/knowledge uncertainty
- We need methods to account for all uncertainties
- Requiring analysis in multiple disciplines – how to allocate physical/virtual resource to reduce cost of VVUQ

The coupling of disciplines can be feed-forward or feedback (oneway/twoway)
- Challenge 1 : coupling in analysis/UQ
- Challenge 2 : dynamic decision making in resource allocation
- How do we optimize these challenges over time?

Challenge 3 : heterogeneous info from different sources (multifidelity simulations/experiments)
The "fidelity" of the model can range from notebook to multimillion
1. Combining multiple fidelities
2. Managing the coupling
3. Reduce epistemic uncertainty
4.
Kreegan method dealing with low quantity of data – often time we need to rely on private data so we can use Gaussian process to do UQ

Decompose process into three steps : where in the input space of multidiscipline do we allocate resources, use sensitivity response to tell which input has most effect, which type of resource shall we allocate (experiment or simulation)

I want to reduce uncertainty to 10% -> how shall I do this using the steps

Brian: Glad you mentioned multifidelity – we use multiple rigorously and judiciously

Wei: If we have resources available find the best way to use them all together

Curtis: People are awash in data are scavengers – look for interesting things and react accordingly. Only need to give "good" advice – our problems are reducing uncertainty telling people what to watch.

Pat: Changing the paradigm from bigger computers always to judicious UQ is important.
Jim: It's not necessary to have an absolute truth, reducing risk is important.

Jackson Mayo:
- Constraints from math/compsci for some VV techniques
- Complex systems need to be simulated by computers – another complex system.
- The undecidable prediction of future systems is fundamental to early compsci.
- We cannot VVUQ any arbitrary system – they need special characteristics.
- Why can we predict these complex systems?
- Resilient systems have smoothness – small changes lead to small results leading to predictable, resilient and evolvable systems.
- Ability to bound the effect of perturbations is crucial for VV and UQ

- Beware of applying techniques assuming smoothness to non-smooth problems such as complex digital systems (some of these systems) – attacks are upon the not naturally resilient nature of the systems (small change in input -> big bug)
- Securing cyber and VV of cyber are equivalent problems

Wei: What about physical systems that respond in a small change -> big response?

Jackson: Some highly sensitive systems such as phase transitions are difficulties in physics models, but some compsci is all "phase changes"

Bill: How do we keep with categories of these types of complex systems? No structure -> useless. As we better categorize better chance of VVUQ, important path forward.

Curtis: Perturbations are homeostasis on a big scale – on cyber another factor is computers have closest analogue of replication to biology. Part of the large failures in digital is not adaptability – is the test/change cycle so powerful.

Jackson: Computing systems have tremendous potential to be organic, but is not necessity. Typically a human a priori what will work – mostly human input vs autonomous.

Curtis: All code was not written against autonomous code.

Bill Oberkampf:
- Important distinction between fixed underlying mathematical structure and not
- We can VVUQ Bayesian nets, neural nets, ai, threat detection, etc
- Can we use mathematical structure on uncertain parameters changing as the software learns and structural constraints or an adaptive structure
- Opponents can throw curveballs to lie of their intents
- If mathematical structure is static we can use VVUQ
- Although trees can fool parasites, the intent of humans is beyond the degree of freedom of other things.
-

Curtis: we have a hierarchy of controller on top of options and instead we can do controller on top of controllers to account for the "dynamic systems" for the time/resource scale we have chosen

In spite of evolutionary algorithms a first binning cut is the static vs dynamic

We can still do VVUQ on code verification of a Bayesian net because we can put exact solution in with specific parameters and since structure is fixed we could do it.

David: mathematical theory underpinning bayes net then we can verify performance of the network – we reduce to code doing what math says it does.
If we can determine structure and the structure does not change then VVUQ is possible
David: Put enough time and money – image a persons brain – the structure is there but the underpinning is not there.

We are on meso-scale macro-scale we need to look at models on a practical level
Philosophical knowability of the extended prediction of "known" theoretical underpinnings
Proposed new change: If we can test at multiple evolution points then it can be VVUQ

In response to open questions:

David: Are you arguing self-aware system that is impossible due to math indescribable or…
We cannot know things perfectly, our adversaries, so we cannot VVUQ them

Jackson: Reconcile the two ideas. What is structure in one level is parameters in another level –
some subsystem that is fixed – the level that you model at can be undecidable. Some higher level
structure than the Turing machine is decidable.

Structure is the representation of the model – the physical side descends to atoms (too deep).
What level of adaptability is the main point.

Jessica Turnley:
- Question of how we choose what we include in the model – affects validity, credibility
- Social science side – everything does not go down to physics – a problem in where the
  source of the model comes from

- We pick from the world what we model – which is theory.

- The fundamental structure of the model can change – what is then the replicability of the
  model?
- Customer can drive what is included in the model – is the model we chose credible vs
  another teams model

- Million ways to talk about social drivers – which one is true or accurate? The models
  validity is in question.

Steve: Where models agree can give us confidence.
The process can give as much value as the artifact. We make explicit what is generally implicit –
bringing what is important for everyone onto the table
What are the rewards for specific behaviors?

Jim: Related to unconscious bias – red flags that arise on bias – detect when that might occur.
Multiple experts brought together for models.

Daniel: Do you think it is better for the modelers to use the model to come up the answers? Or is
it the customer who should engage with the model?

Engaging with model is incredible learning process. Do you want a number or be in the process?
Demands for quick results vs ability to decompose to a number.

Question of where you bound the "model" – if you can't socially give results then so what – is the gui/talks part of the model. Credibility comes from the whole package

Asmeret: credibility vs confidence. How good the model is vs how confident the people are in it. Two interplaying pieces.

Bill: "We have internal view of credible vs external credibility for surgeons. We need to be self confident to sell models." Strategic issue of overconfidence.

Brian: Mathematical model selection is established, how is social?
Social models seem to be less established in structure, but little footnotes etc. Why is it structured this way? Because it made sense to me. Self policing is lax.

David Stracuzzi:
- Data is reserved for VV – from a data science perspective all we have is data – no model.
- We are trying to find behavioral patterns – models can't track the infinities.
- Variables in variable environments is difficult to model.
- Target could be trying to hide in cyber world.
- How well does the data speak to the question you are asking? How much does each variable affect the question? A flip of sensitivity analysis.
- With optimization we can try to find which data we want to gather/store/use? If trying to optimize sensor constellation – logistic questions now – but we could back it out and use specific things for your questions.
- Question in social of which data should we use and look at in the deluge.
- We could look at individual uncertainties on variables and combine instead of top UQ back towards variables.

Curtis: Does value of information take into account moving towards correct/incorrect answers? Information is valuable to the extent to the effect on sensitivity/uncertainty of answer.

Wei: Shannon's entropy is related to uncertainty. Context of purpose redefinition of optimizing performance in financial sector value is actual dollars.

Wei: It depends on intended use of the model.

Bill: Difference between UQ and risk assessment can be hard to quantify in real dollars etc. Do you ask question of impact of positive/negative consequence of model prediction?

Daniel DeLaurentis:
- System of systems exhibit features of distributed entities with variable independence
- How do we verify system of systems?
- Formal methods have made strides in software systems – we can only reach so far is the idea that we have limits in the computing power
- Is there some marriage of formals methods based approach with simulation such that the SOS will be learning the whole way
Jackson: Constraints on local analysis can inform the SOS

How do you compose models such that the components are all correct?

3.5 Notes from Track Session 3 on Decision Making (Facilitator, Eric Vugrin

<u>Participants:</u> Roshi Nateghi, Katherine Jones, Tammy Brown, Elizabeth Keller, Matthew Lave, George Backus, Dennis Engi, Shreyas Sundaram, Dan Pless, Chris Fraizer, Dan Pless, John Sirola, Keenan Breik, Linda Delafuente
<u>Note taker:</u> Jacob Caswell

<u>Notes:</u>
Introduction: Heard a lot of talks in the plenary session about VVUQ done at Sandia. Must now brainstorm to generate research ideas. Goal is to throw out ideas, outline research seed ideas. Start out with those who originally sent in research seed, have them outline their ideas, intro it, discuss it, see where it goes. Give everyone a chance, but will move on if either spent too much time, or if conversation drying up.

Confidence and Uncertainty
Always worried about UQ of models. But at Sandia, the focus is on solving problems. Despite uncertainty in a model, there may be high confidence that an intervention will yield a good outcome. Discussion focuses on framing UQ on how to eliminate bad outcomes. For instance, concerning ISIS, doing more research not always an option, a decision must be made. The decision hopes to yield the optimal solution. All of the information that is necessary is in the uncertainty. Uncertainty is not an enemy, but a friend. That's where the information is. Maximize the value of the uncertainty. A deeper understanding comes from knowing what you don't know, and may be essential in yielding a good outcome. Focus on solution centric approach, rather than model centric.

The goals of the models are essential for decision making, and so the uncertainty not only describes the system, but the outcome.

Our customers have problems, but often do not give enough time for a very complicated problem with limited data. How does Sandia best approach this? How good are our models, how confident are we in their outcomes? This is a struggle for all Sandians.

We must do a lot of work on behavioral assessment. What will a group do? Requests come in quickly, the consequences are significant. How do we approach it? What uncertainty doesn't matter? Assume tremendous uncertainty, potentially use optimization, or run sensitivity analysis.

A simple model, though, can quickly help. Always use the "crap to gold" model. A bad initial model, functional, but at least understands the problem. Appreciate how much you don't know. Functionality through emphasis on the intervention and unknown.

Laura gave the example of a project working on validation, but not all projects work for long periods of time. Is this a generalized modeling challenge?
- Typically, customers are in risk domain.

- Project in question was delivering results to weapons program. Fast response vs. good response.

Reasonably common in control theory for model operating in line. Making adjustments continually. Very fast response required vs. decision support. Could be from hours to years. Range occurs in different domains

Be careful in fast to good response, especially in role of policy maker. Occasionally SNL must respond quickly to obtain confidence in work, and may extend to longer projects
ASC used to do co-development before DOD project. There is no dead right answer in speed when responding to needs when dealing with national security. No one expects those in an urgent crisis to wait 8 years. But resulting wake of "was the decision correct" persists.

Agree with the outcome vs. the model. But don't you need model of uncertainty that needs to occur to understand what doesn't matter?

Yes, build a model, be explicit in assumptions, potentially use simple model. Always embrace the uncertainty of the model to improve the certainty in your intervention. Control theory: If I have a lot of uncertainty, given what we work in at SNL, a lot of energy must be expended to reduce the uncertainty.

What I think about your suggestion is a way to formalize the question. What question are we trying to answer? Asked that way is very open, not technically specified. What decision are you specifying, how do the outcomes matter. "what is the question" becomes a technical specification for the intervention.

Forget about what you want. The customer may want to make money, but really doesn't want to go broke. Refines what the customer thinks about the situation, helps make the model.
It resonates with a relatively small amount of studies on policymakers in making decisions. The studies care most about uncertainty in the decisions. Requires understanding of assumptions in decision. Model for uncertainty in the study went to appendix, not to the forefront of studies.

One perspective: for long projects (10 years), one year VV is approached is a way to assess capability as a distinct process as using VV to assess credibility. For a 10-year project about capability development, four different generations of code, running on different computers, VV projects are understanding capability as it develops rather than a quick turnaround project for uncertainty. Additionally, more decision focused rather than model focused: DoD viewpoint agrees uses term use risk. Focuses on systems engineering "requirements on model: if met, you have this much risk."

Many times in 1500, when you build a model, it is interpretive: you argue the validity in a boundary condition. When are the procedures or physics valid? It corresponds to a range of values. In social science models are extrapolative. The model has never been tested in this way, never been used that way, little data.

Thought that model must answer question. Maybe too limited. One should accept compromise on that for complex systems to make the model be a subject to VV relative to the data you might

be able to get. For example, if you build an electric systems model, you might want to try to look at past outage reports to predict that kind of information even if it doesn't exactly answer the question people have. Trying to match the questions that can be asked. For example, when H1N1 broke out, (a timeline of modeling matters), one of the folks said that modeling did not have a good name, it always came after decisions and contradicted decisions. Doing modeling separately to that, when however, many people got ill, but when the data was collected in real time was the number of counties in real time. Focusing on questions might be too limited, instead, focus on what you might be able to get.

Good perspective. Tendency to say that the output to the model must be the answer to the question. But a good model of 75% of the information where the decision maker must make a choice may also be useful.

One of the benefits for models is helping a decision maker understand the complexity of the problem. For instance, if we try to generate their understanding of the interactions, then modeling only a subset of the data fails.

It is a balance, you can get information about things you can't directly measure with models. It's a balancing act.

The thing is: these discussions are interesting but relevant. A problem in mind: in the past, models made for power outages were developed, but a long term model of climate change has very much uncertainty with downscaling, model, longtime horizons--A lot of uncertainties. There are a lot of investments needed for the planners to have confidence in what you are telling them. If their decisions require major investments, how can they be confident in making their decision?

We must be careful when honestly talking about uncertainty with the customer.

Credibility comes from who is delivering the message as much as the message. If they don't know you and you give them bad news, then it will not work.

The point of this case study could be to build the model and get some result, but the real knowledge comes from building the model. If done right, you don't need the model. It's the thinking through the process. Sometimes the goal is not to predict a number, but understand the factors. So customers can think about factors.

You usually think about simple or complex models. A balancing between knowing more about more or less about more. Pick the uncertainty. What are you trying to avoid as the customer? For instance, if you can answer more honestly and credibly with a simple model.

Pedigree charts to visualize multidimensionality of uncertainties. Done the most experiments on uptake of policymakers.

Another question regarding utilities. In the event of a natural disaster, the response is active, they must spend billions immediately to build walls that hamper response, or simply need to be rebuilt

later, and cost another billion. It is hard to give directions with simple model, but we may know that customer's decision may be wrong.

With utilities and resilience, part of the process is to add new concepts or metrics for consideration into the decision process. The decision process doesn't account for that yet.

One question: what is the responsibility of modelers to understand the decision context? When is it critical to know the nuances of the nested decision making process. We must know about incentives. Where do we bound models where the complexities of the decision making process is essential for model builders, where must they be built into the models.

If we assume the omnipotent decision maker, there is no need to worry. But realistically there are no omnipotent or forced decisions. There's going to be some of the powers in one place and some in other places. Depends on who the customer is.

The scale of what you are modeling must be considered. The policymaker, whoever it is, may be focused on the micro problem. Macroscale, however, may be a safer place to model. Sometimes they may not ask the question for the right space. They may not be grasping a bigger context. If they're asking you for a micro-analysis over a huge area, that may not be feasible. Giving a bigger context may be easier.

Sometimes the reverse is done. They want us to give a specific scale model, without accounting for the bigger picture.

Regarding risk communication, scientists may be embarrassed by uncertainties. Given our understanding and uncertainties. Do you want a 1/10 chance? Or 5/10 chance? Such a response may miss the understanding of risk. Embrace uncertainty rather than be embarrassed by it.

But there are also risks we don't understand. We may not be provided with all risks that cannot be understood. It is upon the decision maker to give all relevant risks. Give customers all uncertainties, do not necessarily make the decision for them.

But, customers may want to be told what to do. They may not know what to do with the probabilistic forecast.

Fear dominates greed. Often when models are produced, when the model is built, new things that can't be measured may appear. This can be considered an insight. These things should be brought up. "Is this an issue?" If this happens, here is how much you will lose x amount of money. Do you want to avoid that or not? At least a way to approach the problem.

What if they go for the most conservative because of risk aversion?

They may want decisions, my job is to operate, yours is to analyze. So if being able to shape options and quantify both uncertainties and impacts, the potentially both could fulfill the request you mentioned. But focuses the level of uncertainty and provides a response.

Prospect Theory

How do we give information? Often those who receive information misunderstand it. Humans make decisions and affect resilience. Humans perceive values and losses in unique ways. They overweight low probability events and underweight high probability events. We may want to shape the uncertainties in different ways. Can we come up with a mathematical framework using prospect theory? Reason what the implications are for resiliency. One scenario, a complex system, with a certain probability of attack, which is viewed through a skewed lens by the decision maker. Having irrational decision makers may help the outcome. Homogeneous is better than heterogeneous. How do we bring human decision makers into the equation? We must understand what humans do with the information.

Model the humans with the uncertainty, adding that uncertainty would work better to provide a high confidence high certainty. Less uncertainty qualification issues when factoring predictably irrational people. There are irrationalities that are predictable.

Questions: What kind of model are we talking about? Descriptive models, predictive models, prescriptive models? For the solutions space of policymakers, must realize there are different categories of policy. These categories may impact technology deployment. What are they trying to accomplish? Working to advance the quality of life for the citizenship. How would that be articulated? How do you bring all these measures of the quality of life together? If you don't have the holistic perspective of all dimensions of quality of life, you won't get very far. Start thinking holistically on the solutions.

We may have access to their decisions, and may be able to model how "irrational" their decisions have been and take that into account. Data on that scale may not be digestible. The one or two things on a mind can be modeled and taken into account.

Those who try to account for those almost invariably fail in modeling. Must include all normative. But that could be incorporated into the model, and interpolated from data. The best approximation must be made, but must be examined holistically. But there remains the question of using prospect theory as descriptive vs. interpretive. There is also a difference between what is said and what is done. Rather than going on what is said, go with the descriptive model.

This is closely related to all three things. Slightly different step in modeling process. A lot of models in the national security domain cannot always find data for what people would do in the situations. People react differently in high stress situations. Incorporating rationality and bounded irrationality, in the places you would want descriptive representation, then you must develop a good representation without experimental data. Knowing that can't be done through elicitation, people don't have adequate information, or would be embarrassed to describe their true priorities. A lot of the traditional utility theories may not be adequate. Even a couple of individuals having a different prioritization structure could create different unpredicted behavior in model. Think about behavioral concepts to add realism in the steps of process for individuals in the process. Can't directly do the experiment, but with sensitivity analysis, some corollary to create an acceptable experiment could be possible.

Sometimes the model could be used for high level policymaker vs. low level implementer. Sometimes the model must be slightly manipulative to influence the outcome. Picking manipulations is significant, and context is essential.

Realistic outcomes may diverge dramatically from expected outcomes in social models. Ethical questions about manipulations, using prospect theory to better incorporate range in modeling agents. But also influence the communication of models. There are questions in deciding what the intended reaction of the policymaker is. Not about misrepresenting data, but quantizing the data so it falls into the range where it doesn't matter. Always trading off policies or requirements. How do we know that noise won't affect the outcomes in ways that cannot be well described? Has to do with heterogeneity and granularity. Depends on the model and the asymptotic similarity of the problem. Is it the noise or is it a signal?

We don't always know what the decision maker needs to know. We simply must give all information they could potentially use. You must give them as much information as possible. Sometimes they cannot always understand or comprehend the data. How do we convey all of that information?

Understanding the success of our models
Are you trusted or not? Sometimes "red, yellow, or green" is asked of us. Trust is earned in helping with what they need. Where is the line drawn? That's just part of modeling. How do we extend and go to the next space?

How do you assess the end result? The success may be measured as "we gave the information in some amount of time."

There are two pieces: Did they use our information? And what were the unintended consequences? What is embarrassing to scientists is different from what's embarrassing to a policymaker. If we are caught up in what is embarrassing, the overall first goal will not be met. How do we validate the value of our work? Are the decision makers using this information, is the situation better? So many levels of bureaucracy between customers and decision makers that cannot always be overcome. Must take lots of time to do the elicitation. Data was never acquired.

One answer could be, are you funded next year, and at what level?
ASC has a history of funding VV, but impact has not been measured. But the money keeps flowing in. Somehow, maybe without a direct answer, some signal that something right was done. Usually asked to make models to inform decisions, used, informed, decision is made, and the project is over. Not a lot of use cases for feedback loops. Tabletop simulations offer useful space for feedback. Can observe how information is used, and what information is taken, and factor that into the overall response. May look for data that isn't there. Certain value of a pre-model that allows us to test the assumptions of the full model. But the time is never available for this method.

Back to the crap to gold approach. The danger in a yearlong project is the progress can't be made until the model is perfected. Alternatively, you present the imperfect model at month three. The customer would understand more and be more willing to use it.

Building the buy inHelps them understand what can and cannot be accomplished. Regardless of what you tell them about the uncertainty.How to transform the output of this model to help give them the right information regarding their decision using manipulation? Transform the information into a lower more understandable form. Once you know someone they make the same decisions. Take analogous decisions to inform future decisions.

Proposed Research Questions
What's the research question?
All over the place. How do I quantify the measure of confidence in a probabilistic risk statement? It can't be done in the model, but in the optimization model. Here is a portfolio of policies that I am confident cannot happen. How to be done in a complex system? By incorporating interventions, and include the decision maker, and include the decision as part of the model. The model gives data. Information is needed to make the decision. Intervention and decision consequences, and add the human element of how phenomena may be used to quantify all consequences.
Analogues cannot be used to validate a model. How else can they be used?
As part of the model and contributor to the uncertainty.
If we have this complex system and want to validate that, how do we figure that out?
Formalisms are different in complex systems. Portfolio is transitory. New decisions must be made for the consequences of past decisions.
Progressive hedging must be expanded upon.

The role of understanding the decision making context, and the impact of models
How do we build understanding the decision and context, is that part of the process or the model? Prospect theory: how do we integrate understandings of decision maker's interpretations of uncertainty. Regarding feedback loops, there is no time to figure out impact, but for SNL as an institution, should there be analysis for the impact on national security decisions. How is that decided when going forward? What is our responsibility for institutional learning and hearing feedback?

Interesting take to do at an institutional level. Could only done at the project or program model. Hard without resources or connections.

Is there a business development case? We as an institution should put money on this. How to use a case to decide this?

Find multiple programs that all serve the same customer? Rather than a task's influence, but the whole cyber portfolio. A lot of potential feeds, and the cost benefit may be beneficial.

1500 has that opportunity, customer internal, dedicated funding scheme. Interesting, but unlikely. If a systems groups could join. (150).

It would be interesting for DoD side of things to have impact for nuclear weapons. Only way we know if it is being used: if funding continues. That is the only way a lasting impact is confirmed.

Is it business development? Or part of the process? Ask customers, do you like us? Why do they like us?

Would it be an informative thing for the lab to know? Would it be important for them to know our effectiveness? Could be used to show how effective Sandia is.
There are multiple uses for that data.
It could be a personal bragging point, and bragging point for the lab.
Unintended consequences piece. Objective is to illuminate those unintended consequences.

Discussion of stakeholder's priorities and our responsibilities
You wouldn't have to model each cell of the matrix. Total qualitative description. Identify things qualitatively due to particular regulation, and how it impacts other than intended ways. Whether or not it is modeled directly.

If they didn't come to you with a question of consequences, what is the responsibility?
Are most of the customers looking to minimize losses or take course of action with least uncertainty?

Often, it depends a lot on the particular application. For some, it's what's going to help them win a lawsuit if they get sued. Others are more forward thinking, give me the science, and they'll deal with the policy. Knowing that should help drive how the model is implemented. It varies a lot.

Depends on the context. Direct customer may interact with many agencies, and any developed analyses were multipurpose. No single question, but trying to inform a community. Informs what information is generated and how it is communicated.

Sometimes the model isn't the deliverable. Sometimes the model is used as a tool to develop the solution itself. The result of a model is a subset of the possible results.

Conversation about external sponsors
What are non Sandia sponsors interested in?
NSF and DOE regarding resilience of grid. One part is the resiliency of the grid under extreme climate. Understand pushing the boundaries of robust decision making. How do customers seek answers?

"Don't tell me what I'm not asking for" mentality from the customers occasionally.
Incentives may not be in place for minimizing consequences. Customer didn't like hearing input about increasing tree trimming, for instance. How do you think about, with changing grid topology, how do you project future demands. How to efficiently distribute and manage the grid.

Characteristics of models with heterogeneity
Have the characteristics of models where heterogeneity must be injected been identified? What are the indicators that say you must have some percentage of your actions, not because of uncertainty, but due to heterogeneity of human behavior?

If there is literature on it, probably not complete.

Caution against starting out with a model and assuming you have to do stuff. Understand the system and the question and intent, and build as needed. Quick and dirty, build, compare, and then retool as necessary. More of an art.

A better process could be developed to know enough vs. not enough. A lot of detailed papers, but none that generalize in any way, across heterogeneity and granularity. Helps define confidence in output.

Replace heterogeneity and nonlinearity? At what point are nonlinearities need to be characterized?

Certainly depends on the question, but what does it look like in the extreme, nonlinear, then put it into the model.

Never let the invalidity of nonlinear models affect their implementation.

It is good enough.

Does it predict? No. Well enough? Yes.

The way "invalid" is used is more colloquial vs. precise.

Worse than that for the grid. Cannot be run on nonlinear model. Directly implementing the linear model will result in blackouts.

Thoughts about decision making as a research topic
John's point earlier, turning this into a specific research idea is hard, since it often depends on the decision being made.

Under what conditions does x, y, and z work? Today, just throwing out ideas.

<div align="center">Discussion Ends: 4:07 PM</div>

4. DAY TWO SECOND-PASS RESEARCH IDEAS

4.1 Second-Pass Research Ideas

1 - Chris Frazier
1. Understanding better the boundary at which adding complexity to the model(s) no longer adds benefit. This might be because the eventual decision no longer changes, or that the adds sophistication is confusing or confounding or counterproductive (which may include raising costs or run times towards infeasibility) the goal of this is to "simplify" models. Which in theory, may be easier to run through a VVQU process.
2. Introduce the quality of model in terms of how it is used as a measure of validation. "Quality: might be: in the model actually used? Are the effects of the decisions based on the model "good"? is the model impactful? Did the model lead to mistakes or "bad" outcomes?
3. Can we incorporate modeling decision making into the complex systems models to deliver information which leads to "better" outcomes in the end?

2 - Dennis Engi:
1. Explore the use of "Argumentation Theory" in place of standard V+V & UQ techniques. This may be of real value for complex problems.
2. Use a "model" that has both qualitative as well as quantitative aspects of the holistic problem/solution space in order to identify unintended consequences. Always remember: you can't optimize a system by optimizing its systems.
3. Explore the spectrum of solution specification from individual agents to a cultural authority to learn the variety of impacts

3 - Katherine Jones:
1. What are the characteristics of models where heterogeneity in agent behavior (or human behavior in this case) needs to be inserted to ensure better representation of possible outcomes? Adaptation on (learning) ability or agents populations are?

4 - Dan Pless:
1. Design models to effective V&V from the start: don't focus exclusively on answering the question, but understand how to use the data that actually is available in the data poor domains that are common in complex systems.

5 - Laura Swiler:
1. High dimensional SA/UQ for complex systems with thousands to millions of parameters. (note: I don't have an approach for this except sampling combined with dimension reduction approaches, but it is a research need for VVUQ for complex systems)

6 - Jackson Mayo:
1. Better quantitative and qualitative methods to work with simpler models, avoiding needless/misleading detail and guarding against unforeseen fundamental errors (black-swans) e.g. ensembles of diverse but mutually cross-checking models

2. Constraining systems (e.g. engineered designs) to improve analyzability via stability/resiliency, and via modeling at an appropriate abstraction level that captures the structure that makes it predictable

7 - Wei Chen:
1. Different VVUQ methods need to be developed for diligent types of complex systems depending on the level of adaptability.
2. How to account for Human bias in the validation/credibility issue.
3. How do we define the value of information considering the intended use of the model?

8 - Bill Oberkampf:
1. An important point was made concerning the importance of the communication between model developers (i.e. Sandia) and customer concerning expected results, credibility, useful results etc.
2. The justification and documentation of credibility of a model must include VVUQ as well as limitations of the model and what it should NOT be used for.
3. On my slides I have argued that the complete systems group at Sandia must segregate complex systems models according to adaptive versus non-adaptive structures.

9 - Pat Finley:
1. Validation of networks. Many (most?) production CS decision marking informing models from Sandia use networks. Empirical networks available for validation (e.g. survey results) often are poor analogs for the systems network (e.g. anafluence nets).
2. Validation of Sandia models often hinges on whether the models networks "close enough" to the systems latent network. Rigorous methods to adapt similarity metrics approaches to networks.

10 - Vicente Romero:
1. What properties of other complex systems would allow them to be amenable to existing VVUQ methodology for computational physics models (based on bottom-up hierarchical approaches). What types of complex systems have these properties under what conditions?

11 - Alfred Hubler:
1. Accurate prediction of time/space continuous macroscopic systems with time and space discrete models.
2. Managing complex systems with controlled synchronized real time virtual systems.
3. Modelling the role of thresholds such as static friction with terminating decimals.
4. The role of universal differential equations in modeling complex systems.

12 - Steve Verzi:
1. Use a hierarchical, layered neural ? (can't read word) approach to understand & quantify uncertainty across multi-scale modeling related to Wei Chen's research idea.

13 - Jeff Tsao:
1. Building human confidence in interim knowledge: cost versus accuracy tradeoffs between "thinking fast" versus "thinking slow" approaches
2.   Comparing models against models versus comparing models against observation: moving from absolute to relative validation

3. Tradeoffs between causal and correlational models in sparse and data rich environments
4. Bounding models in time and hierarchical level: systematic approaches to optimizing to the model's use

14 - Dan DeLaurentis:

1. Does the answer to "who is the user" of a model for a complex system get enough attention, especially in the context of a "decision-centric" mentality. In other words, the state/features/organizational setting of a decision-maker should be accounted in the modeling process

15 - Kasimir Gabert:
1. Using (validated and expensive) mappings to the General Grand Complex Model, we validate a dirty bomb response in New York using experimental data from a controlled ant farm experiment"
2. "With Amazon Mechanical Turk we paid one thousand people to come up with independent models and the appropriately combined model is valid"
3. "We successfully built a distinguisher that can determine it is inside of the model using techniques that necessary break the usefulness of the model."

16 - Matt Lave
1. How to include operator decisions in models and how to best tailor models to help make decisions rather than simply produce data?
2. How to present uncertainty in an upfront what that is useful to operators rather than hiding the uncertainty because we are embarrassed by it.
3. How to display/model/communicate uncertainty of an adaptive system that will change in the future?

17 - Shreyas Sundaram
1. Embracing uncertainty: focus on achieving desirable outcomes as opposed to getting the model right
2. Bringing models of humans (and their decision making) into the overall model. Including ideas from behavioral economics/psychology. Shaping the uncertainty quantification/risk communication based on this understanding.
3. Understanding unintended consequences: optimizing a decision for a given stakeholder might lead to consequences for others. How do we reason about the holistic picture in complex systems?

18 - Drew Levin
1. Do customers actually care about uncertainty or do they simply want discrete answers?
2. How do we communicate uncertainty to customers who don't want to or don't know how to integrate it?
3. How do we evaluate how well our models properly reflect that true issues the customer is interested in? Can we facilitate ways to obtain usage feedback from customers after delivery and project completion?

19 - Steve Kleban
  1. Can the credibility of complex systems models be quantified?
  2. Can it be analytically determined when a system needs to be modeled with complex systems techniques?

20 - John Siirola
  1. Is there a bigger vision of validation that goes beyond model validation to explicitly include the use/decision making process? Can that be experimentally validated? Is there some way to validate without direct experimentation?
  2. Is there a bigger picture for "UQ" that goes beyond simple parameter sensitivity? Complex systems and complex in time. Similarly, I would assume the UQ would need to be with regards to time series parameters. How do we quantify/sample time series? (They don't appear to be defined by "simple" distributions).
  3. If VVUQ is always use-specific, how can we generalize complex systems/VVUQ to perform research that can be revised in a general approach?

21 - George Backus
  1. Are there procedures to determine minimal levels of granularity and heterogeneity needed to adequately carry out VVUQ analysis? Can this logic be extended to determine "cuts" that reduces the dimensionality of the problem and therefore make VVUQ more tractable?
  2. Is there legitimacy in applying VVUQ to the confidence in intervention rather than in model outputs?
  3. Does adding more uncertainty such as human behavior in the model actually add feedback and increase the confidence in intervention efficacy?

22 - Ken Hu
  1. Given a process that incorporates VAV, modeling, customers, and decision, what are the different ways that information flows around, and does that influence the outcomes of the decisions? I've seen connections made between small subsets of this process, but not the whole process.

23 - Elizabeth Kistin Keller
  1. Evaluate the effectiveness of Sandia modeling efforts in terms of impact on actual decision and examine role of VVUQ in enabling decision to be made based on info. Pilot internally (1500 with internal customers) and expand externally to impact on national security decisions.
  2. Evaluate the impact of different methods (textual, numeric, visual) for communication uncertainty regarding intervention options? How vary by audience, time frame, nature of problem.

24 - Roshi Nateghi
  1. Implementing holistic risk assessment, incorporating unintended consequences within models
  2. Projection of system behavior under deep uncertainty and effective communication of the associated uncertainties with the decision makers.
  3. Adequacy planning in the electric grid (considering the emerging changes in the topology of the grid).

25 - Brian Adams
1. (With credit to Wei Chen/seconding the idea). How can models of multiple fidelities be used with heterogeneous data in a rigorous way to reduce uncertainty or increase credibility/confidence of complex systems models?
2. How can we embrace/leverage qualitative/subjective/fuzzy knowledge of interest type metrics (instead of precise numeric QoIs) and intended use for decision making to do more efficient UQ and risk analysis? If model outputs are "soft" how can confidence be quantified?

26 - Jim Stewart
1. Hypothesis: complexity arises from relatively few and simple "rules" or actions. These actions happen as a result from being subjected to a myriad of environmental/external/internal conditions (the "parameters").
   Idea: Why not invert the modeling concept? Why not "parameterize" complex systems models in terms of the "rules". This could dramatically reduce the dimensionality. Sensitivity analysis would be performed with respect to the rules.
2. The rules become the parameters, such that inserting or removing rules does not change the fundamental model structure or form. Not only is dimensionality reduced; this could open the door for more rigorous V&V/UQ approaches.
   Key point: In traditional models, the parameters are nouns or objects and rules are verbs or actions. In the proposed new paradigm, the rules become the objects or essentially the dual problem.
3. Still unresolved: how to view the "parameters" in the new paradigm? Can they be recast as rules?

27 - Curtis Johnson
1. Value of information under different conditions, over time or with different data. Do VoI at regular intervals (or same sampling scheme) and aggregate VOI and correlate with features of that data.
2. Value of algorithms – same as above, but a la meta-learning, characterize the data features/decision etc. that make a situation amenable to a given algorithm or approach.
3. Leverage goals and protocols – model and or validate at the level of the brains/goals/decision of the system. Below this level, the details may not matter.

28 - Asmeret Naugle
1. Confidence is relative – how should we account for situations where decisions must be made and models are compared to or replacing very low confidence methods?
2. Can we classify systems or identify characteristics of systems that specifically make models of those systems good or bad candidates for reductionist modeling and/or VVUQ strategies? (This may also apply to utility of hierarchical V&V).

29 - David Stracuzzi
1. Repeat of submitted idea: explore how the analyses of the data produced by complex systems, including the associated uncertainty and value of information, can be used to improve underlying system models and performance optimization methods. The idea naturally extends into understanding the wish associated with the made decisions.
2. A human factor's point of view: how does the knowledge of uncertainty impact human decision making? What information and uncertainties should be realized to achieve the

desired outcomes? This level of inquiry also extends into the problem of effectively reading or otherwise communicating uncertainty and validity.

30 - Jessica Turnley
1. What is the boundary of the model? Is it limited to computation artifact – or does it include the communication to the user (the user interface, the comm. Package)? How would expansion of the definition affect a definition of model validity/credibility?
2. What about some kind of assessment that focuses on how decision would change thin the use of a model? This would turn attention to the role of the model in decision making.

31 - George Barr
1. Reexamine continuum analyses of ecological., epidemiological problems by rewriting the related recursion relations honoring the implicit time steps associated with reproduction, mortality etc.). to develop integer solutions what are integer problems. Develop a library of such solutions for comparison to other methods; examine other behaviors suggested by developments already in the "continuum analyses such as chaos.

32 - George Barr and Mark Smith
1. Tapping into the wisdom of the crowds for complex system model validation using crowd sourced gaming.

33 - Daniel Appelo
1. Machine learning meets multilevel, a framework for automatic surrogate selection and cost vs. error optimization.

4.2 Tim Trucano: Articulation of Binning of Second-Pass Research Ideas

Day 2 Breakouts:

I. Value of Information (VOI)
II. Strong Tests
III. Risk
IV. Model Design

Cross Cutting Themes:

"Error"
"Polarities"

I. VOI

Lead: DeLaurentis – Scribe: Johnson

- Formal Information management
- Formal Information Valuation
- Nature of "Evidence"

[27.1] "Value of Information under different circumstances"
[25.2] "How can we embrace/leverage qualitative/subjective/fuzzy knowledge?"

See: 1.1, 2.2, 2.3, 6.2, 7.3, 8.1, 8.2, 12.1 13.3, 16.2, 16.3, 18.1, 20.2, 22.1, 23.2, 25.2, 27.1, 27.2, 29.1
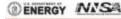
II. Strong Tests
Lead: Oberkampf – Scribe: Gabert

- Selection
- Metrics
- Qualitative Dimensions
- Nature of Evidence

[34.1] "Methods for creating more robust model validation…"
[5.1] "High dimensional SA/UQ for complex systems…"

See: 1.2, 2.1, 2.3, 5.1, 7.1, 7.2, 9.1, 9.2, 10.1, 13.2, 15.1, 15.2, 15.3, 20.1, 20.3, 21.1, 23.1, 32.1

III. Risk
Lead: Chen – Scribe: Lave

- What is risk?
- How is it managed?
- Nature of evidence
- EXTRAPOLATION

[8.1] "An important point was made concerning the importance of … [communication]…"
[17.3] "Understanding unintended consequences…"

See: 1.3, 5.1, 8.1, 8.2, 13.1, 13.3, 13.4, 14.1, 16.1, 16.2, 16.3, 17.1, 17.3, 18.1, 18.3, 19.1, 19.2, 20.1, 20.2, 21.2, 21.3, 23.1, 23.2, 24.1, 24.2, 24.3, 27.3, 28.1, 29.2, 30.1, 30.2

IV. Model Design (V+V, UQ Informed)
Lead: Hubler – Scribe: Swiler

[2.2] "Use a 'model' that has both qualitative as well as quantitative aspects…"
[4.1] "Design models [for] effective V+V from the start"
[25.1] "How can models of multiple fidelities be used?..."

See: 2.2, 2.3, 3.1, 4.1, 6.1, 6.2, 8.3, 11.1, 11.2, 11.3, 11.4, 12.1, 17.2, 19.2, 21.3, 25.1, 26.1, 26.2, 26.3, 29.1, 30.1, 31.1, 33.1

Cross-Cut: "Error"

Type I: Belief model is "wrong" when model is "right"

Type II: Belief model is "right" when model is "wrong:

Type III: Solved wrong problem

Type IV: Problem solution used incorrectly

Cross-Cut: "Polarities"

Qualitative v. Quantitative

Objective v. Subjective

Knowing what is wrong v. knowing what is right

Complex models v. models of complex systems

Fast response v. "Good" response

5. DAY TWO BREAKOUT GROUP A ON THE VALUE OF INFORMATION (VOI) (FACILITATOR, DAN DELAURENTIS; SCRIBE, CURTIS JOHNSON)

5.1 Session 1 Note-taker Notes
Track 1 – The Value of Information – Facilitator Dan DeLaurentis, Scribe Curtis Johnson

Breakout, Group 1, Session time 0915

High-level minutes

Topic: The Value of Information (VoI)

Introductions:

Daniel DeLaurentis: [intro]

Mark Smith: [intro]

David Stracuzzi: Interested in systems that analyze the world online and using their performance to analyze and improve their own performance.

Brian Adams: Interested in what are the outputs people care about

Drew: Interested in what clients care about. Clients don't care what goes in and out of the model. How do we hide all that and communicate uncertainty to clients. PhD in complex systems. Dissertation looked at distributed systems in biological search, for example how information is distributed in the immune system among T-cells.

Curtis Johnson: Believes information is ad hoc. All information should be curated and stuffed into a data base. But that's difficult. We want to reuse code and data, but requirements change so drastically that this becomes challenging. Thinking ahead about new data types and new features of the data can make information more valuable and continuously reusable.

Daniel: A model is a living thing that is continuously involving. A model can be driven to answer new questions. … We should not be too scared to explore topics not officially listed as discussion topics.

Drew: When does more information become a bad thing? More info can make communication harder and can actually degrade the model.

Daniel: Formal information valuation is critical. Without an objective function, this is difficult. Being able to prove credibility by demonstration would be good.

David: The purpose of valuating information is to change the understanding of the output of the model. Specifically, it is a mechanism for narrowing your focus to the things you should pay attention.

Daniel: In design of experiments, you start by screening to identify important parameters and variables to avoid too much information. This is relevant to model design.

Drew: Yes, such parameters can be analyzed quantitatively as well. Some parameters don't matter.

Curtis: There are moving targets. The value of information is more difficult when the data and environment change over time. For example, spammers are constantly changing their strategy. What features of a detector are important today, tomorrow.

David: Sensitivity analysis is a great tool, but in complex systems, it's not something you can really do. What might be better is how certain are you about the outputs of your model with respect to the input.

Curtis: We can sometimes shift machine learning to a learn about machine learning. Meta learning.

Daniel: Complex systems models suffer from lack of data. How can we use meta learning to prove credibility.

Curtis: We can at least give confidence that we are doing everything we can. That we are actively moving toward an optimum by meta learning. Effectiveness can be measured by looking at financial loss and gain due to the model

David: This analysis is long term and can be ineffective because it is a form of averaging.

Drew: The technique Curtis suggests is a weak signal. But there are other statistics used in sports that could be useful. Get an idea of what an average player looks like and see if replacing a player improves the performance.

Daniel: Complex systems models must not be used to generate insight.

Mark: Agreed. The models can barely help with predictions.

Drew: These models diverge extremely quickly. But specific predictions are difficult.

Curtis: But the model can at least narrow down our sight to what we want to look at with more energy.

Brian: Sometimes we want to look at the range of possibilities. A model can bound outcomes.

Curtis: At a distance, we can make predictions. For example, we can identify what might go viral with good probability.

Brian: How do you quantify the amount of gain in insight...

Daniel: This sounds like a good research target.

Curtis: One practical way is to see whether workers are more effective or efficient with the help of models in hand.

Daniel: One model tried to make predictions in what people do over 50 year period by looking at levels of regret and what we know about how regret affects decisions in people. – Let's look back at what you were saying about using a model to refine itself.

David: Suppose I put information into a mathematical model. If you understood the uncertainty In those predictions, you can know where in the math is that uncertainty coming from? Look at

the predictions the model makes. I have some expectations about what the output should look like. When the patterns I find don't match what I'm expecting or aren't what I'm looking for. This means the model is missing something I'm looking for. This is true mathematically as well. Perhaps there is a simple phenomenon that the mathematics is not capturing.

Brian: Compare the model to expert opinion. That's a weak or fuzzy but perhaps useful notion of validation. Quantifying expert opinion might be a research direction.

Curtis: Perhaps the dynamism of the situation and of the model could actually be and advantage instead of an obstacle. Sometimes motion can give us more information by giving us new perspectives.

Brian: Perhaps moving models and ensemble data sounds scary but can be harnessed.

Daniel: What about fuzzy, subjective, qualitative behavior from the decision maker. How do we include this in the model.

Drew: Yesterday, a group talked about ways to integrate the feedback from the client into the model design earlier. This helps show which features or outputs of the model were actually used by the decision maker.

Daniel: Consider this. What percentage of all the models developed by Sandia get used by the customer directly as opposed to Sandia developing the model and using it to collect outputs and report them to the customer.

Curtis: There's a lot of both. And models change users change models change users. You have to expect the questions to change as the client is served. By giving the client weak tools but that they can use to build their own models and iterate rapidly, sometimes they can get much more use and effect out of the models.

David: If you ask a question that the model is not ready to deal with, it should give you a low certainty answer.

Curtis: Users can add another layer of primitive VVUQ by quickly observing the effect of using the tool. And giving them the tools gets them using them and gives a leg up to credibility.

Brian: Is there a way to measure the value of the tools to the client?

Proposal/Themes: …

Effect: Produces learning insight among the user or decision maker and improve credibility. Advances the users understanding of either the model or the problem space and changes or improves the questions asked (of the model).

Quantify how well the model helps you make accurate predictions.

Lori: Models adapt to the environment of changing questions. That adaptedness gives some measure of validation of the model.

5.2 Session 1 VOI Scribe Notes

Mark: Measuring Diversity

David: How much info do we share from the model output?

Drew: Communication with the customer-helping customer make decisions

Curtis: What information and features to use for what question, changing over time

Design of experiments, screening process for what experiments to run

Sensitivity analysis

Can we do VOI over time? Across a changing environment?

How certain are model outputs relative to inputs?

What about meta-learning?

Sports data: +/-, WAR—value of substituting one for another

Redefining prediction as bounding possibilities, reducing uncertainty.

How do we quantify the value of 'bounding possibilities', triage, etc.?

Regrets?

Efficiency of the human analysts using the model?

IF you understand the uncertainty in the outputs of the model,

Then you can do some things to allocate responsibility for that that uncertainty

to different model/math components.

Possibility theory, fuzzy stuff, quantification of SME

Can we do something like SAR? Sample the model and data over time in such a way that the changes help with VV/UQ

What can we do with fuzzy inputs from the decision maker?

How do we put a feedback loop from customer reaction to the outputs to change the questions and the model?

Prediction vs. Insight?

Can changing state of both modeler and decision maker be a way to make this quantifiable and amenable to VVUQ?

Is there a way to measure the insight gained from diverse analysts? Can we leverage the 'unreliability' of the analysts? Can we quantify the gain from using the model?

Themes:

Formal information management approach for complex systems modeling and V&V/UQ

Learning and insight of users/customers, enhances credibility?

Advancing the user/customer's understanding of the model or the problem

Improves the model itself

Changes and improves the questions asked

Can we quantify these?

Accuracy of predictions

What data do we need to answer the question?

Red Queen syndrome and adaptedness: the model and the users and the data sources and the questions keeping up with the changing times…

5.3 Session 2 Note-taker Notes

Topic: The Value of Information (VoI)

Curtis: How do we treat VoI as a dynamical system? Can we have a canonical set of inputs and track how the outputs change as a proxy for the model changing?

David: We need some sort of metric. What does it mean to improve?

Brian: Solving this problem speaks directly to VoI.

David: There isn't much evidence to back this up, but VoI is most prominent in financial systems where value is easily bound to dollars. For a network of sensors, for example, we can see how much each sensor contributes to the final result.

Drew: This is sort of like sensitivity analysis.

David: Right, but backward.

Curtis: Maybe we can substitute machine learning features for sensors and have this carry over to machine learning.

David: VoI can be seen as a tool for driving down uncertainty. But it can also be seen as driving down the confidence. It's unclear whether this is actually true or whether the two are connected.

Curtis: We are going to find some way to track the model, the questions, the data, inputs and outputs, over time. Then we'll take existing static VoI approaches and do some Bayes net like

things, and we'll do a next VoI time step, and use that to drive how we change the model. We are building a model of the model, or a model of the system.

Daniel: That sounds like a typical VoI approach but adapted.

Curtis: One kind of change to the model is in its parameters.

Drew: Sensitivity analysis is about how much a parameter change affects the output. This is more like which parameter values give the best output.

David: The question is, whatever you actually care about at the end of the day, and we want to know which of these parameters have the greatest impact on the final decision.

Curtis: By looking at the output of the model I can refine my questions until I get to the point that I ask a question that I can actually answer it. Maybe we can measure the "impedence" or the match between the question and the model.

David: If you ask a question and can't get a high certainty, then either you are asking the wrong question or you have the wrong model.

Emma: Why is a metric so important?

David: The value of a having a metric, a numeric quantification, for the quality of the model, is that it allows us to optimize the model with respect to the parameters, the question, and so on. To do this we need to understand what the VoI metric should be. What are the things we are trying to relate using this metric? Is it really inputs and parameters to outputs? I'm making this up as I go along.

Daniel: The VoI is the core. In optimization language, the objective function is …

Curtis: The objective function is the adaptedness of the system. The object functions are the uncertainty. We can't make it the actual quality of the decisions, which is impossible to know.

David: Can we use confidence in the model as the objective? That wouldn't seem to make sense in this approach.

Daniel: Can we develop a test to understand the sensitivies of the model?

David: Yes, but unless you're bringing in outside data, you're not going to know how this affects the outside world.

Curtis: We can't get that much outside data, but maybe we can use what we get to define a sort of known region or safe region for the model.

David: Critical subpoints are what are the manipulable parts of the model and other system parts like the question being asked, what is the VoI metric, and for validation how to do we pull in external data and use it.

Curtis: Parameters can have various form including rule sets, real numbers.

5.4   Session 2 Scribe Notes and Report Out

Meta-model:
Over time, track

- Model parameters
- Data inputs
- Data outputs
- Model questions
- Real-world and experimental outcomes

~~~~~~

Adaptive Modeling: Using VOI to Minimize the Impedence between Model, User Question, and Data

- What are the manipulable parts of the model system:
  - Of the model itself:
    - Parameters
    - Data inputs
    - More structural changes
  - Of the decision system around the model:
    - Questions asked of the model
    - Risk calculus of decision maker
    - Credibility and Confidence
- Decide what VOI metric should be

Pulling in external data, using it for validation

6. DAY TWO BREAKOUT GROUP B ON STRONG TESTS (FACILITATOR, WILLIAM OBERKAMPF; SCRIBE, KASIMIR GABERT)

6.1 Session 1 Note-taker High Notes

<u>Breakout Group 2: Strong Tests – Thursday June 23, 2016</u>

- What are the subcategories? Perspective from Complex Systems
  - o Selection of Tests:
    - Where would you do the tests in this massive dimension world?
    - Selection of tests from the perspective of a model user versus a customer.
    - Must define a test exactly so it can be reused. Evaluation criteria. Logic about why it's a test worth doing.
    - Tim Trucano: "Specify the problem. Specify the Evaluation criteria and tell me why."
    - What principles are you using to choose a test?
    - Means of generating or discovering strong tests
    - If we want to build confidence/credibility, we break it down into subsystems. How can you develop tests that are simpler than the complete system? How can you do more simplified tests (not as strong as complete system, but you can't test the complete system anyway)? Specialized, simplified tests. Ask can our software pass this test? They hardly ever pass but you learn things and go fix it.
    - Try to use more rigorous modeling techniques. Using combinatorial optimization (a more rigorous approach, but more computationally complex). Can we write down mathematically a set of equations that represent this system? Then do we have a solution procedure that can implement this mathematics? Maybe you can write it down but you can't solve it. Let's say you go with system dynamics. You've introduced potential error. Is this solution procedure appropriate for the math? Let's use more rigorous solution procedure. Dakota has a whole lot of heuristics in it. Introduces error. How much error? Goes back to questions of how to quantitatively/qualitatively measure this. The desire is more rigor.
    - Can you construct a model at a lower level that avoids making a certain assumption at a higher level and so it is a more restrictive problem but then you can do a test on that lower piece and suppose you can come up with some analytical solution to it?
    - How can you take this problem apart piece by piece? How could you come up with a simpler model (example: who will Zika affect?) and how do those pieces form the hierarchy? And then you could test these individual pieces?
    - It's in the interaction of these pieces that the complexity comes out. Maybe you can validate each piece on their own <u>but then how do you validate their interaction?</u>
    - Issue is with assuming that we can do these hierarchical decomposition. Come up with a lot of ideas but we haven't shown that the emulation of a physical solution is sufficient. Metrics aren't strong for this model. How can I discover the correct test for the model?

- How well does your model do on the real strong test because you can't ever do it on the complete one?
  - Metrics
    - Represents a polarity.
  - Qualitative Dimensions
    - What happens when metrics, input, output, etc could be qualitative?
    - Is it possible to define a strong test in the face of qualitative information?
    - What does a comparison of the benchmark mean that is a comparison with a qualitative dimension?
    - How do you measure qualitative and how do you incorporate it with quantitative?
    - A lot of work already done for PDE based problems. What is the cost of evaluating this function? How much uncertainty does it reduce?
    - Example: How vulnerable is the DOD supply chain?
      - What does it mean to be vulnerable?
      - Subjective.
      - Should you drive the output to be strictly quantitative?
    -
  - Natural of Evidence
    - Why are you doing tests? To develop evidence of what?
    - Is the evidence quantitative?
  - Means and mechanisms of evidence collection
    - What are the boundary conditions on the new systems?
    - Can you potentially use big data and data analysis to peer into an existing system and assess whether or not that system has degraded or not, etc (perform an experiment)?
    - These issues still apply to data analytics also. How do you know these algorithms are good?
- V & V is about testing / rooted in testing. Most think about it from validation perspective: "show me that the model is right." Goal is to assess the model.
- Weakest part of complex systems. Needs attention.
- Get a base of material (physics based or whatnot) to learn from.
- A lot of knowledge about how to do this in the physics-based world but we haven't talked much about what does that really mean (to have the laws of physics on your side, can do physical experiments). How do we take that knowledge from that world and get anything out of it in a non-physics based world?
  - Maybe there are similarities and they may be greater than people expect but so what? You have a lot of things happening in physics modeling that are beyond what you need in other complex systems (?).
  - Even in the realm of models that pass fundamental tests (basic laws of physics) can still not be good models (they fail later).
  - Even if physics, you don't always have everything you need to benchmark against.
  - What is the test basis?
    - In code verification, it is a test of software and numerical methods that transform equations into discrete methods (method for mapping calculus

to arithmetic). Come up with special cases of analytical solutions that says if your code does not reproduce this, then your code is wrong.

- Solution verification: numerical error estimation of the fact that you went from continuous mathematics to discrete mathematics (arithmetic is only rough approximation to calculus – what is the error?).

o What do you do in absence of experiments?

- You have a mapping that has inputs that is associated with work units and that drives down an error. In a complex system, you have to have some notion of error (could be a weighted notion of error?) (?)
- You could have multiple models and determine which one is best… But best in what sense?

o How do you know if an approximation is good and relevant to answer the question?

- Where does physics and non-physics based diverge?
- What happens if you know data only AFTER the event? Movement of humans and goods after natural disaster, etc. Weapon testing if you can't do a full test (setting it off).
- Even if you suppose the hierarchy (triangle in Laura slides) is true, how do you do even a simple comparison? Even you have to compare a real world system and a cyber system, that's not a solvable problem. If we can build a framework that is a computer program to distinguish/compare to things. In crypto world, you build programs that cannot distinguish between two things (at least with a non-negligible probability).
- Want a framework for a complex system where a third party can come in and say "did you do these steps?" and be able to make an assessment without knowing all the ins and outs. Look at results and use those as a strong test of the work you've done.
- For complicated problems, there's no way of knowing a priori what the tests will be. Must be teased out of the problem.

- Subsets of complex systems to enable V&V testing: We need to break down complex models by structure into simpler models. These simpler models must have the ability to run experiments.
- Need taxonomy of complex systems by experimentation potential.
- Disciplines within complex systems:
- Leveraging computational physics V&V/UQ in complex models: What from computational physics V&V can be leveraged to V&V/UQ of complex systems? We need to understand the nature of these things so we can decompose the other systems and find the similarities.
- Technique for validating without running experiments? Trying to validate a new model against a previous model.
  o Sparse data: what is the uncertainty given that you only have one sample?

6.2   Session 2 Scribe Notes and Report Out

0. Applying data mining and analytics to discern observables for strong testing
1. Finding substructures of complex systems that enable strong testing
2. A taxonomy of complex systems' experimentation abilities
3. Disciplines within complex systems models
4. Leveraging computational physics V&V/UQ methodology in complex systems' models
5. Deep learning aggregation of the input/output behavior of heterogeneous models to assess complex systems' model
6. Building strong tests from the perspective of both the model builder and the model user
7. Design of strong tests based on observable proxy quantities
8. Development of both quantitative and qualitative evidence for strong testing

----

Possible titles

0. Title: Applying data mining and analytics to discern observables for strong tests
   Idea: How can data mining and data analytics help discern observables or find signatures of the input conditions and response

1. Title: Substructures of complex systems to enable V&V testing

   Idea: We need to break down complex models by structure into simpler models. These simpler models must have the ability to run experiments

2. Title: A taxonomy of complex systems' experimentation abilities
   Idea: Need a taxonomy of complex systems by experimentation potential (We can use exactly the real system at different scales, we can use a very good invented proxy, we can't think of a proxy, etc.)

3. Title: Disciplines within complex systems models
   Idea: Taxonomies of complex system models themselves

      Complex Systems
         - Natural
            - etc.
         - Engineered

      Old days:
         - Deterministic chaos

156

- Scaling exponents
- Using computers to discover interesting things about dynamical systems

New days:
   - Actually helping customers with an applied bent

Types of model techniques:
   - Simulation models
       - System Dynamics
       - Agent based
       - Time-step models
       - Discrete event simulations
   - Optimization models
       -

Different breakdown:
   - Physics models
       - No agents, no cognitive actors
       - There are principles of physics
   - Biological models
   - Behavioral models
   - Engineered non-physics-law-based models
       - (e.g. modeling the grid under normal behavior, growth)
       - (e.g. improve threat detection in a secure area)
       - Computer network models

4. Title: Leveraging computational physics V&V/UQ in complex models
   Idea: What from computational physics V&V can be leveraged to V&V/UQ of complex systems?
   We need to understand the nature of these things, so we need to decompose the other systems and find the similarities

   QASPR - Looked at 18 different validation/code verification metrics
       18 different ways of making the comparisons
       Some apply to certain QoIs, some others
       Note: code verification metrics likely do not need to include stochastic

   Note: Metrics frequently mean difference measures

5. Title: Deep learning aggregation of differing models' behavior to validate a complex model
   Idea: Technique for "validating" without running experiments
           -- This may still be a "Strong Test"!
    - Aggregate input/output of behavior of differing models, result

157

is defined to be valid
- This operates over sparse data. Do we need an uncertainty model?
- This is experimental data with uncertainty

Is there a difference between data from controlled experiments and uncontrolled observations?
- This is okay, difference is smaller
- This is removing your ability to completely specify the model boundary conditions (?) to perform a validation test
- You can choose the sensors, but not the event. How can you adapt a given test (where you can piggy back onto a physical experiment) in order to obtain data to help you validate your model?

6. Title: Building strong tests from the perspective of both the model builder and the model user
   Idea: How can we design strong tests from both the perspective of the model builder and the model user?

7. Title: Design of strong tests based on observable proxy quantities
   Idea: Related to quantities of interest by validating other quantities of interest
- How much nyquil/theraflu was purchased in an area may proxy how many H1N1 infections they are

8. Title: Development of both quantitative and qualitative evidence for strong testing
   Idea: Under category of nature of evidence

Note: This workshop may be too broad, methods are likely different for different goals and systems

Posit specific test problems that we could apply thought to
   Bomb in Chicago on Monday in the morning - requires you to know spread of radioactivity and its deposition, social activities of people, how first responders and local hospitals respond, etc. We could then show specifics for each of these
      -- can we separate the egos from the models?

Validation Goal: Assess the model through testing

We need a taxonomy of complex systems focused on experimentation potential.

Selection

- Select strong tests
  Think hierarchy
- Means of generating or discovering strong tests
- We may need to research into how we really can know that
  combinations are correct

Metrics
   - What is a benchmark in a qualitative dimension
       Some examples of qualitative questions:
           1. How vulnerable is the DoD supply chain?

Qualitative Dimensions

Nature of Evidence
   - Why are we doing tests? To develop evidence of x?

Means and mechanisms of evidence collection information
   - Use big data (where does the data come from??) to peer into existing
     system and perform an experiment
   - These should be simpler than the complete system
     Not as strong, but that is okay

From the ideas:
   1.2 -
   2.1 Explore Argumentation Theory
   2.3 Explore spectrum of solution specification
   5.1 SA/UQ with millions of parameters
   7.1 -
   7.2 Human bias must be accounted for
   9.1 Survey results are poor analogs
   9.2 -
   10.1 Proxy properties that are amenable to existing techniques
   13.2 Compare models to other models
   15.1 Validate through appropriate mappings in other domains
        Create a general model that we can validate to
   15.2 - 13.2
   15.3 Use a distinguisher to compare model and real world
   20.1 Validate with use/decision pair instead of direct experiment
   20.3 - 15.1
   21.1 Find minimal levels of granularity and heterogeneity
        Perform "cuts" to reduce dimensionality
   23.1 Evaluate end-to-end use, internally first then external
   32.1 Use crowd sourced gaming to validate

34.1 Aggregate input/output of behavior of differing models, result is
    defined to be valid

7. DAY TWO BREAKOUT GROUP C ON RISK (FACILITATOR, WEI CHEN; SCRIBE, MATT LAVE)

7.1 Session 1 Note-taker Notes

<u>Break out III: Risk Session 1</u>
June 23, 2016
<u>Lead:</u> Wei Chen
<u>Scribe:</u> Matthew Lave
Discussion begins: 9:30 AM

- What is risk?
  How do we define risk?
  o Classical definition: Probability x consequence
  o Risk perception, perception of a negative consequence. Specifically the anticipation of it, not just a calculation. Has to do with the example of living next to a reactor. Calculated and perceived are orthogonal
  o Risk is personal.
  o Risks we cannot control are perceived more intensely than risks we can control
  o For the decision maker, it isn't the risk, but the characterization of confidence.
    ▪ Confidence and dependability
    ▪ Dependability: Why we approached it as such
  o For a decision maker, they may have a variety of stakeholders, who may each have a different perception of the consequences.
  o Change stakeholders, get different ordering of risk
  o Unanticipated risk and consequences.
  o Danger and opportunity
  o Risk is perceived only as negative in the US
    ▪ Not quite, the stock market, for instance. High risk, high reward.
    ▪ Churchill: Never waste a good crisis
      • In times of turmoil, opportunity to create order
  o Quantitative analysis shows high risk
    ▪ Decision makers see possibility to lead to good outcomes.
  o Risk tolerance of every individual
    ▪ Some are more risk averse than others
  o You must express risk attitude. Risk attitude will influence the decision.
  o In a research challenge, there is qualitative and quantitative.
    ▪ Whether we think of risk in terms of the things we most want to avoid, at least in the quantitative sense, makes an intractable consequence.
    ▪ Quantitative could become nontractable. You have data, but no information.
  o In the environmental arena, environmental risk, the difference between risk as a point and risk as a probability brings a difficulty of expressing risk to different people. Binary "is it risky or not?"
    ▪ Communication of risk and uncertainty. We have no good methods for communicating uncertainty. Arguably even with experts.
    ▪ How to make useful information rather than data.

161

- Communication of risk
  - Communication helps understand "what is risk?" Communicating about risk.
  - Often there are two sides of probability and consequence. Cannot express to 17 place accuracy, maybe 1 place. At least do an ordinal ranking of likelihood and consequences.
    - Display graphically and ordinally
  - Communicating and calculating.
  - Need the method of characterizing risk.
    - Computational scientists have completely different view
  - Probability of failure more broadly.
  - How do we talk about risk in different fields.
  - Risk is such a big topic
    - Deals with how stakeholders see it
    - How they will use it
    - A lot more than just risk with the model being wrong, or not being able to communicate it.
  - Also about how they use it.
    - Do we know how they will use it?
    - Cannot tell them the risk of using the model
    - Only the risk of the prediction.
    - They must calculate the risk of action
  - Do we have a moral obligation to be part of that process?
    - If it is misinterpreted, is that partially our responsibility?
  - How do we communicate about the model?
    - To the customer
    - What assumptions?
    - Communicate it in a language they are comfortable with
  - Do we know what role the model will play in their decision making process?
    - Do we understand it enough?
  - Bad examples
    - Giving equations to those who aren't comfortable
    - They may understand it differently
  - Risk mitigation
    - If you are looking at a process kind of system engineering.
    - May want to improve the model or system.
  - For the classical definition
    - Include condition and current state of knowledge.
    - May not be possible. Includes the inclusion of knowledge, and the model.
    - Should be integrated into the model.
  - Where are we computing the expected value of the good news?
    - Done almost explicitly for negative risk.
    - Goes back to the point that risk is not JUST negative.
    - It depends. Risk can be something that is good.
    - Not included in the model
  - The way risk is framed:
    - Tend to be worried about how a model can screw me.

- Focused on how packaging, quantifying, risk of models being deployed.
- Not black and white.
- Boldness implies a higher level of risk
- Irreducible risk
  - Not do it at all
- What is the role of risk and risk management.
  - Where should risk management in complex systems go?
  - What is the role of VVUQ in achieving this end?
  - Philosophical understanding of risk and communication / management.
- A decision maker must understand the value of taking the risk
- Many decision makers work in environments where they are not necessarily rewarded for good things
- Bad things are recognized, considered, remembered some factor more than good things.
  - If one bad thing is done, may be remembered 10x more than a good thing.
- In the context of a company
  - Include a function of reputation
  - Risk must be taken from the eye of the beholder.
  - Expected value takes no account of this.
  - Prospect theory discusses the asymmetry of good vs. bad rewards.
- Why do any modeling?
  - Want to understand, and lower the risk around the decision.
  - The information gained from modeling is also laden with a variety of different risks.
  - How much confidence do I have in that?
- Confidence, and communication
- Has nothing to do with the model, just with the decision
  - Can we help make a decision that reduces the likelihood of a bad option.
  - Understanding risk vs. irreducibility.
- Validation has to do with your confidence that the next time the model is used, it will produce models of a certain type.
  - It has been sufficiently tested against experience
  - Validity is not in the model, but the person using the model.
- RESEARCH TOPICS:
  - How do we extend the relatively accepted approach to quantitative risk to include qualitative aspects like perception, or risk attitude or non-quantitative metrics?
  - How do we extend the concept of risk as an output of the model, as opposed to risk inherent in the model and how the model is used.
    - Risk of a model, relative type errors,
    - There is no risk until the model is used
    - How to extend the concept of risk from just the output of the model
  - Communication of risk
- How can we know what uses the model will be applied for.
  - What if a model must be used for multiple applications?
- Modeling systems vs. modeling problems.

- Always wrong to take a model and apply it. (Type IV)
- How is it managed?
  - Do nothing at all
    - UNACCEPTABLE
  - Uncertainty reduction
  - Characterization of uncertainty
    - Uncertainty of many different aspects, model, numerical
  - Built in control strategy for any unanticipated environment.
    - Included in process or model?
    - Uncertainty management of the model
    - Acknowledge uncertainty of model. Strategy to build into system to react to uncertainty
    - Physical system, like aircraft. Must account for the probability of it buckling or failing. However, aircraft could also be changed in design to have flexible wings.
  - Have mitigation strategies. Assume the model is wrong.
    - Epistemic: model reduction
    - Editorial uncertainty: Cannot make reductions. Must come up with strategies to make design flexible.
    - Move editorial to epistemic, once the environment is more well known.
      - Related to latent variables. If all are not understood, defined as latent. Once understood, they may become editorial variables.
      - Can be accounted for.
    - Robust design does not guarantee robustness.
    - System may have feedback loop to change in response to unknown variables.
    - Risk in the system vs. risk in the model
      - Should be a loop. The system is created, the model is made, the uncertainty is made, the model is improved, etc.
  - How do we understand and manage risk in the modeling approach
    - Not in the actual construction of the system
    - The system is out of scope for understanding the risks around the model.
    - Reframed to say that the idea is to communicate the wide variety of range of things that could happen.
  - Suppose the system is a dice roll
    - Know the model, 1/6 probability
    - How is risk managed in the system?
    - What is risk?
    - How is risk managed for the system?
  - Decide what we focus on.
    - Manage uncertainty in modeling and simulation.
    - Is that risk rather than uncertainty?
  - Risk is contributed by many sources of uncertainty
    - Focus more on epistemic uncertainty
    - Focus more on how we manage the part we can reduce.
    - Uncertainty times consequence.
      - Manage uncertainty or consequence?

- Consequence may be VVUQ
- How do we manage risk with respect to uncertainty
  - The model deals with uncertainty. Risk is in the use of the model
    - Puts the model into a use environment
    - What do we mean by risk in the model?
  - In the electric grid, they don't care about the uncertainty, just the risk in their choices?
  - Uncertainty must be communicated into risk.
    - How to we translate uncertainty to risk?
  - Assume a really defined risk
    - Map uncertainty into risk
    - How do we reduce or manage the uncertainty?
    - Do not talk about how the risk is assessed?
    - Manage and reduce uncertainty.
      - Which leads to communication issue
  - Prefer management of risk rather than reduction of risk.
    - Reduction of uncertainty may be reduction of information
  - To some extent, it is more important to understand the impact of uncertainty.
    - Probability may not have a known distribution.
    - Probability may be subjective.
    - Distribution can be found from three or two points
  - Epistemic uncertainty
    - How do we stabilize uncertainty even in the face of new information.
    - Can we bound the uncertainty?
    - Management is the stabilization of uncertainty.
    - Can be characterized, may not always be stabilizable.
    - Use of a portfolio of decisions to manage and stabilize the system
    - Not complicated, but one that can't be predicted.
    - How to make decisions
  - How do we manage the model of the system?
    - Reduce or understand?
  - We can use the models to find strategies that are more robust.
  - What does it mean to provide details on uncertainty?
    - Risk can be put into models all the time.
    - Difference between simulation and analysis
    - The result of the analysis may be an actual risk
    - Simulations don't have inherent risk
      - Analysis does
    - Only appropriate to use cases
  - Many think simulations implicitly include analysis. The two are logically independent, but used together
  - The purpose of the modeling is to help inform the decision.
  - How is modeling risk managed by the modeler?
  - How is risked managed by the decision maker using the model?
    - These are different questions
    - They tend to be mixed.

- o Word model is overused
  - ▪ Simulation != analysis models
- o Research opportunity:
  - ▪ Should risk be focused on the decision, or on the model? How to focus on these?
  - ▪ How is weiji managed?
- Nature of evidence
  - o What is evidence to say that a prediction is good?
    - ▪ Robustness
    - ▪ Qualitative and quantitative
      - Makes sense
      - Really good at generating hypotheses.
      - If the model can withstand the hypothesis
    - ▪ If uncertainty may become constant
      - There is confidence in what we do not know.
  - o Reputation, presentation, etc.
    - ▪ All contributes to how well the model is received.
    - ▪ Is not nature of evidence in communication, but not necessarily to the decision maker
  - o VV ASC took over the years about the developer reputation.
  - o Not only personal, but institutional.
  - o What is legitimate evidence in a complex systems model?
    - ▪ What constitutes legitimate evidence in a complex systems model.
    - ▪ Evidence that you have confidence in a complex systems model.
  - o Explore the possibility of using argumentation theory
    - ▪ A communications technique (rigorous)
- Extrapolation
  - o Personal knowledge and expertise
  - o Risk plays a huge role
    - ▪ If the model is used to predict things in the future, then there is a big risk
    - ▪ Fitness for purpose
      - Here's my model, here's my question. Is this useful given the assumptions in this context?
      - Prediction without any data.
  - o How can you quantify the erosion in confidence as you extrapolate?
  - o If the model is used outside of its bounds, outside of its domain.

Meeting officially ends at 10:34 AM

7.2 Session 1 Scribe Notes and Initial Report Out

III. Risk

- What is risk?

- How is it managed?

- Nature of evidence

- EXTRAPOLATION

What is risk?

- Probability times consequence
- Risk perception: perception/anticipation of negative consequences
- Risk is personal
- Characterization of confidence
- Each stakeholder may have a different perception of risk
- Unanticipated consequences
- 危机 (Wei Ji) Danger/opportunity: high risk high reward
- Risk attitude influences decision making
- Qualitative vs. quantitative risk
- Risk vs. uncertainty

Research Topics:

- Translation from uncertainty to risk.
- How do we extend what we have as accepted approaches for quantitative risk to include qualitative risks such as decision maker's perception?
- How do we extend the concept of risk from the output of a model to also include uncertainty inherent in model and inherent in how model is used?
- Communication of risk

Communication of Risk

- No good methods for communicating uncertainty
- Method of characterizing risk (physical world vs. user): equations may mean nothing to user
- We don't know how model will be used: user has to calculate risk based on their use

Research Challenge

- Methods for effectively communicating risk

How is risk (in uncertainty) managed?

- Do nothing at all (not acceptable!)
- Characterization of uncertainty
- Stabilize/bound uncertainty
- Build robust strategies in complex system models

Research Opportunities

- How is modeling risk managed by the modeler?
- How is risk managed by the decision maker using the model?
- Should risk be focused on the decision vs. on the model?
- How is 危机 wei ji managed?

Nature of evidence

- Robustness (answers robust, convince user) in face of overwhelming uncertainty
- Models as testbed for hypothesis
- How well you present/your background/etc.: do you convince user?
- Your history, your institution's history

Research Questions
- What is legitimate evidence for confidence in a complex systems model?
- Explore the possibility of using argumentation theory (rigorous communications technique).

EXTRAPOLATION

- Running model outside the bounds (domain) for which we have knowledge
- Always wrong
- Personal knowledge/expertise
- Risk plays a huge role: big risk in extrapolation
- Prediction without any data
- May work for a short time period (or in spatial domain) beyond bounds; very difficult further into future

Research Questions:
- How can you quantify the erosion in confidence as you extrapolate?

7.3   Session 2 Note-taker Notes

<div align="center">
Break out III: Risk Session 2
June 23, 2016
Lead: Wei Chen
Scribe: Matthew Lave
Discussion begins:
</div>

Overview
- Do not need to rewrite. Will revisit previous session
- Identify the potential funding opportunity in this area
- Identify some research proposal topic
- Volunteers

Review
- Three parameters, Bayesian analysis more explicitly
  - May not be necessary,
- Probability: probability of occurrence
  - Probability that something bad could happen. Probability of failure.
- Consequence: could be economic, lives, cost, etc.
  - Impact on people
  - Magnitude of consequence
  - Could be positive, e.g., market penetration of PV systems
- Consider both probability of, and the magnitude of
  - How would magnitude be translated into the cost of an outcome
  - May not be explicitly dollars
- More generally, this cost is to the specific individual.
  - Company A's penetration could negatively impact company B
- All quantities will be mad in the context of the decision maker

Research Topics:
- What does it mean to link these general questions and subtopics to complex systems and VVUQ
- Regarding how the model is used:
  - More specifically, how risk factors into making the decision
  - Why do we care about VVUQ?
- Translation from uncertainty to risk and the decision making criteria.
  - Why not decision making confidence?
- How to translate uncertainty in the model into some quantification or expression of risk to the desired user?
  - Totally different (Dennis)
    - We must include the magnitude and probability of a decision for both costs and benefits.
  - Does the uncertainty in the model inform the cost?
  - Models themselves don't have risk
  - Models have uncertainty

- o Many models made are risk assessment models
- o So the risk is still not in the model
- o Model has uncertainty, risk is how you use it?
- How do we extend the concept of risk from the output of a model to also include uncertainty inherent in a model and inherent in how the model is used?
  - o Trying to make sure that the model maker has properly characterized the risk for the decision maker
  - o Different from establishing confidence in the model
- For complex models, which are adaptive, the uncertainty will almost always be huge. There may not be much confidence in the overall result, but confidence in how the model captures the phenomena.
- How to think and talk about uncertainty and risk; what errors we are trying to avoid?

Communication of Risk
- [communication of risk is] not an untreated topic. The topic should not just be about the communication of risk
- Research challenge:
  - o Methods for effectively communicating risk associated with modeling the system

Management of Risk
- Is it obvious that it is not acceptable to do nothing?
  - o No, it is not always obvious
- Gather more valuable information to reduce risk
  - o Some risk will be irreducible
  - o Any information about the future will be extrapolative
  - o Occasionally more information creates more error
- We will not know if a model is correct
  - o To know that, we would need to gather more data
  - o Or, is it to gather more hypotheses?
  - o The more unknowns thrown at the model, the more a model's uncertainty has been characterized
  - o There are some things where data cannot be gathered
    - ▪ But there may be systems where data can be gathered
  - o Gathering more information often opens up to greater uncertainty
    - ▪ There may be a difference between complex systems and typical systems
  - o More data does not always make decisions any better
- Gathering data may return more information about what is known, but not what is unknown.
  - o Which is why it may be necessary to try and stabilize the uncertainty
- Know as much as we can about uncertainty, not always reduce it
- What kind of better information is necessary to reduce or characterize uncertainty
  - o May need to better understand the use context of the decision maker
- Could reduce risk by gathering more data rather than reducing uncertainty
- Discuss with the group responsible for the value of information
- Relativity of risk
  - o My risk is based on the definition:

- Gives the same risk as high probability low consequence and low probability high consequence.
- Risk definition not sufficient
- Cost should not necessarily be an objective cost, but a perceived cost as a decision maker
- Importance of consequence, informed potentially by prospect theory
  o Not everybody considers nuclear events, even though they are high consequence. Some high consequence events we do not consider

Nature of Evidence
- Anything to add based on the other groups' discussions?
  o What is the evidence for increasing confidence of the user. Not trying to convince, just trying to help them understand our confidence.
  o Increase user's confidence
    - How do you mix confidence and uncertainty?
    - Not convinced there needs to be another probability of the probability
    - Confidence should be implicitly captured
    - Confidence may not be quantified, in the example of extrapolation
  o How do you express confidence?
    - Confidence that the model maker has that they are giving good numbers
    - Confidence of the model maker.
  o Credibility vs. confidence in the model.
  o Terminology problem
  o George:
    - Does the model change in different cases
- Uncertainty, confidence (of model maker), confidence (of the user)
- Analysis product stability as confidence, and its stability
- When the model developer delivers the model, they should quantify the uncertainty, and specify for the use the confidence it will return the correct response
- The last one may be the most influential
- If the decision maker knows there is large uncertainty, and trusts the model maker, then the risk can be sorted out
- The modeler will need to be motivated to produce high confidence and low uncertainty
  o That is desirable, at least

List of potential sponsors
- Lot of VVUQ interested sponsors
  o DARPA
  o DOE
  o NSF
  o Intelligence Community
  o Users, e.g. EPRI, National Gas Association, WECC
  o DoD
  o DHS
  o FDA
  o Weather

- o NASA
- o NOAA
- o FEMA
- o Insurance Companies
- o NIH

Proposal Ideas
- What is legitimate evidence for confidence in a complex systems model?
  - o Evidence vs. Process
  - o Experimental pieces about what kind of evidence presented to decision maker, and how to present it?
- Not sure how to express evidence for an arbitrary user, but for a general model
  - o Unsure about confidence for a user and for a model
  - o How to build confidence in a model
  - o Confident management models
    - ▪ Have been implemented, will be tested, but not conducted just yet.
- George and Asmaret:
  - o How to build modeler and user confidence in complex systems models
- Dennis:
  - o Argumentation theory and communication aspect
- Must have a definition of terms. Must have a glossary in the workshop write up.
  - o We are talking past each other, and do not have a common definition.
- Workshop lexicon
- Value of more information to risk management?

7.4 Session 2 Scribe Notes and Final Report Out

Potential Sponsors

- DARPA
- DOE
- DOD
- DHS
- NSF
- FDA
- Intelligence Community
- Users: EPRI; National Gas Assoc.; WECC; insurance companies
- NOAA; NASA
- NIH
- FEMA

Proposal and Publication Ideas

- How to build modeler and user confidence in complex systems models? Translation from uncertainty to risk and the decision making criteria. What is legitimate evidence for confidence in a complex systems model? George and Asmeret
- Methods for effectively communicating risk associated with modeling and the use of models for complex systems. How to think and talk about uncertainty and risk; what errors we are trying to avoid (Type III and IV). Argumentation theory and communication aspect. Dennis
- Value of more information to risk management (joint with topic 1)?

- Standardized workshop lexicon in context of complex systems.

8. DAY TWO BREAKOUT GROUP D ON VVUQ-INFORMED MODEL DESIGN (FACILITATOR, WEI CHEN; SCRIBE, MATT LAVE)

8.1 Session 1 Note-taker Notes

Attendants: Laura, Alfred, Steve, Jackson, Sasha, Thomas, George, Dan, Craig, Jim, Ben

Alfred: Engineers avoid the VVUQ in complex systems – we hope to bring them together to empower the engineers to be ready with systems that are complex. You can build reliable systems for difficult systems (stock market) despite complexity. How do you verify model for chaotic systems for irreproducible data?

Steve: Assume labeled data, need methodology to do data compression on approximate model. Trying to bring the machine learning theory to the concept of modeling. Instead of trying to make correct, approximate uncertainty/probability of the machine learning growth of the model. VVUQ comes in in quantifying the correctness of the learning. Trying to describe a less strict VV that is quantitative (inside the circle vs "outside" the circle). Some states are very predictable, while others are unpredictable. We want to characterize the prediction over time.

Jackson: Working with simpler models – lot of detail is superfluous/misleading, unclear improvements. Simpler models are easier to analyze, consider using several simple models. Crosschecking missing details – how do we avoid black swan problem? How do we choose to put the system together subjectively – using design constraints to make VVUQ easier. Digital systems can have lots of hacks, etc and the composite can be unpredictable because of minimal constraints. Using resiliency to constrain the system to perform VVUQ better. Inherently multi-level analysis of the constraint we put on the design. Simple as we can abstractly. DOD Resilience: system ability to meet a changing set of requirements. Making the system of the model VVUQ-able to make the model VVUQ-able. Identifying observables that affect the function of the system for simpler models.

Sasha: Fundamental differences between physical and social systems. Physical systems have fundamental laws that make validation simple. Social systems do not have such fundamental laws making VVUQ difficult. Social systems also can change their behavior due to changing reasons differently from the speed of light not changing what is measured. Can we VV social systems? Model is ultimately to drive a decision – decision-driven development. Social models become adaptive models similar to game theory – what we get out of the model is important to whether it is correct. Stock market is more on getting profits then modeling perfectly – predicting one stock goes up is good enough. Physics has one model for "everything" – social is for predicting "anything". Also we cannot see what different parameters would have done empirically in social stock market – social modeling is more difficult. Moving towards incremental validation of social models – validating over time.

George: Propagation of disease is fundamentally integer problem opposed to continuous. We need to show either that our continuous model is applicable or we attempt to find integer solutions to integer problems. Possible solutions: recursion relations, modular arithmetic etc. Discrete models can be more accurate – using the grid size at which the model was derived. Difference equations vs Maps vs recursion relations all the same – language problem.

Dan: The case when you can't make experimental measurements: H1N1 progression of the disease. You can't experiment and you don't necessarily have the data coming back (county reporting). Don't design to solve the specific problem – take into account the data that you will be able to get in the future.

Craig: Qualitative vs quantitative model design. Qualitative: inherently vs not enough data (SME generated). No quantitative way to design the qualitative assumptions. Is there a way to build it in to the modeling process? Maybe such controversial qualitative quantities aren't useful. And if they are they need more input. Quantifying metrics such as resilience. Is there a way to quantify the error this method introduces?

Jim: The rules by which complex systems operate is simple, but parameter set is huge. Opposite of scientific space. VVUQ of a large parameter space is much more difficult. Provide a dual of the VVUQ of low parameter space for large parameter spaces. Try to invert the rules/parameters so that the parameter space is smaller.

Alfred: Do VVUQ by synchronizing real and virtual system then if they match you are absolutely correct. Attendants: Laura, Alfred, Steve, Jackson, Sasha, Thomas, George, Dan, Craig, Jim, Ben

Alfred: Engineers avoid the VVUQ in complex systems – we hope to bring them together to empower the engineers to be ready with systems that are complex. You can build reliable systems for difficult systems (stock market) despite complexity. How do you verify model for chaotic systems for irreproducible data?

Steve: Assume labeled data, need methodology to do data compression on approximate model. Trying to bring the machine learning theory to the concept of modeling. Instead of trying to make correct, approximate uncertainty/probability of the machine learning growth of the model. VVUQ comes in in quantifying the correctness of the learning. Trying to describe a less strict VV that is quantitative (inside the circle vs "outside" the circle). Some states are very predictable, while others are unpredictable. We want to characterize the prediction over time.

Jackson: Working with simpler models – lot of detail is superfluous/misleading, unclear improvements. Simpler models are easier to analyze, consider using several simple models. Crosschecking missing details – how do we avoid black swan problem? How do we choose to put the system together subjectively – using design constraints to make VVUQ easier. Digital systems can have lots of hacks, etc and the composite can be unpredictable because of minimal constraints. Using resiliency to constrain the system to perform VVUQ better. Inherently multi-level analysis of the constraint we put on the design. Simple as we can abstractly. DOD Resilience: system ability to meet a changing set of requirements. Making the system of the model VVUQ-able to make the model VVUQ-able. Identifying observables that affect the function of the system for simpler models.

Sasha: Fundamental differences between physical and social systems. Physical systems have fundamental laws that make validation simple. Social systems do not have such fundamental laws making VVUQ difficult. Social systems also can change their behavior due to changing reasons differently from the speed of light not changing what is measured. Can we VV social

systems? Model is ultimately to drive a decision – decision-driven development. Social models become adaptive models similar to game theory – what we get out of the model is important to whether it is correct. Stock market is more on getting profits then modeling perfectly – predicting one stock goes up is good enough. Physics has one model for "everything" – social is for predicting "anything". Also we cannot see what different parameters would have done empirically in social stock market – social modeling is more difficult. Moving towards incremental validation of social models – validating over time.

George: Propagation of disease is fundamentally integer problem opposed to continuous. We need to show either that our continuous model is applicable or we attempt to find integer solutions to integer problems. Possible solutions: recursion relations, modular arithmetic etc. Discrete models can be more accurate – using the grid size at which the model was derived. Difference equations vs Maps vs recursion relations all the same – language problem.

Dan: The case when you can't make experimental measurements: H1N1 progression of the disease. You can't experiment and you don't necessarily have the data coming back (county reporting). Don't design to solve the specific problem – take into account the data that you will be able to get in the future.

Craig: Qualitative vs quantitative model design. Qualitative: inherently vs not enough data (SME generated). No quantitative way to design the qualitative assumptions. Is there a way to build it in to the modeling process? Maybe such controversial qualitative quantities aren't useful. And if they are they need more input. Quantifying metrics such as resilience. Is there a way to quantify the error this method introduces?

Jim: The rules by which complex systems operate is simple, but parameter set is huge. Opposite of scientific space. VVUQ of a large parameter space is much more difficult. Provide a dual of the VVUQ of low parameter space for large parameter spaces. Try to invert the rules/parameters so that the parameter space is smaller.

Alfred: Do VVUQ by synchronizing real and virtual system then if they match you are absolutely correct.

8.2   Session 1 Scribe Notes and Initial Report Out
VVUQ Informed Model Design (**THEME: Design Models for V&V/UQ from the beginning**)

We need to consider VVUQ right from the start of the modeling process. Unique position of Sandia. Sandia is recognized as a leader in VVUQ. VVUQ for Complex Systems is difficult because of the nature of Complex Systems, strategically important. VVUQ for Complex Systems requires a culture change. Currently the VVUQ community is not well connected to the Complex Systems community, but there is a lot of ad-hoc VVUQ in the Complex Systems community, both in 'applied work', such as predicting stock values, copper prices, weather, climate, and in 'theoretical work', i.e. how to verify a model for chaotic systems, where experimental time series are generally not reproducible?

Lead: Alfred Hubler

Is this a research area?
What is the approach?

Steve Verzi: Labeled data, methodology to do data compression, apply theory of machine learning to modeling of complex systems. Let's give up on trying to prove model is correct, just estimate the quantifiable error (probability the approximation is good). Supervised learning: need "truth" data.

Jackson Mayo: Working with simpler models (easier to analyze, bound, understand its behavior, can analyze multiple, diverse, simple models). Resiliency, adaptivity, stability can take advantage of constraint we put on design. DoD definition of resiliency: ability to adapt to new requirements. Idea: design of a resilient system will make it more amenable to V&V. Two ways to make a system resilient: one is constraining it, one is to open it up (moving defenses).

Sasha Outkin: Modeling differences in physical systems vs. social systems. Validation differences. Turnaround of model is fast. Game theoretic probability. What you get out of the model is useful for making decisions, even if it is wrong (stock market trading). Good if you can predict anything, not everything. Purpose of modeling is more utilitarian (applying model for decision), but if partially correct model can give good decisions. Focus on achieving desirable outcomes from model vs. getting model right.

**Incremental validation **

George Barr. Re-examine model integer problems with continuous approximations. Rewrite the recursion relations (difference equations, maps), develop integer solutions to what are integer problems. Need to do model analysis up front to determine time-scales

Dan Pless. Design models not just to answer the question but also to be able to incorporate data that you have access to now and may be able to collect in the future.

Craig Lawton: Qualitative vs. quantitative aspects of model design. Two types of qualitative: inherently vs. lack of data (rely on SMEs). Have one set of SMEs to develop input data, another set of SMEs to examine is it reasonable. Second set of SMEs provides the "face validation." What to do if it is valid or invalid? Quantify or bound how the SME differences introduce error into the model. Metrics such as resiliency and diversity that would like to quantify.

 Jim Stewart: Rules by which complex systems operate are relatively simple but parameter space is huge (reverse for engineering). Invert that: provide a "dual" of the model. Rules become parameters and parameters become rules, then apply UQ/SA on the rule space.

Alfred: synchronize a real system and a virtual system, then you are able to control the real system while ?

8.3   Session 2 Note-taker Notes

Breakout 4 – After Lunch
Model Design (VVUQ informed)

Attendants: Laura, Alfred, Steve, Jackson, Thomas

Only clarifying discussions recorded – large quantities of talk on how to format/reword research ideas.

Discussion of Jim's idea:

What does it mean that the rules become the parameters?

In physics the boundary conditions are the parameters, and the differential equations are the rules – invert for complex systems. We may be able to get better results with the boundary conditions than the inside rules.

Usual is change the parameters and look at the effects – we therefore change the rules and look at the effects

The top down of complex systems gets married to the bottom up of physical systems – we know the whole system io in contrast to each component io

In Conway's Game of Life go to the emerging microstructure to find the parameters

Complex systems don't have to be true, just predict – we must reproduce the behaviors, not necessarily physically correct

We shake a beaker and look at the waves and apply an arbitrary mapper without analyzing underlying rules

We invert by inferring microscopic behavior from macroscopic behavior

On metrics:

We cannot presuppose that metrics of resiliency exist – but we must try to quantify them if possible

On George:

We use differential equations that were averaged from a distribution and we do not keep in mind the averaging scale when we attempt to discretize the equations

For more coarse and finer grain questions we find artifacts when we use an improperly scaled discretization

On ball:
Because the ball is at rest for a finite period due to elasticity of the ball – dependent on how much the ball is inflated. F=ma applies for a model that is infinitely stiff (inelastic collision). Further discussion stifled due to time.

8.4 Session 2 Scribe Notes and Final Report Out
VVUQ Informed Model Design (**THEME: Design Models for V&V/UQ from the beginning**)

We need to consider VVUQ right from the start of the modeling process. Unique position of Sandia. Sandia is recognized as a leader in VVUQ. VVUQ for Complex Systems is difficult because of the nature of Complex Systems, strategically important. VVUQ for Complex Systems requires a culture change. Currently the VVUQ community is not well connected to the Complex Systems community, but there is a lot of ad-hoc VVUQ in the Complex Systems community, both in 'applied work', such as predicting stock values, copper prices, weather, climate, and in 'theoretical work', i.e. how to verify a model for chaotic systems, where experimental time series are generally not reproducible?

Possible research call for ideas:
1. **Call: How can we cope with and leverage the evolving, adaptive structure of the system we are modeling for VVUQ? Explore the conditions for the emergence of resiliency and diversity.**
2. **Call: Develop methods for confidence in the credibility of the complex system model, accounting for the purpose of the model.**
3. **Call: Managing complex systems with controlled, synchronized, real-time virtual systems**
4. **Call: Develop methods for inferring the underlying microscopic rules (e.g. individual agent) from macroscopic behavior (system level).**
5. **Call: Develop accurate mathematical methods for macroscopic discrete time and space models of continuous and discrete systems**

Call: How can we cope with and leverage the evolving, adaptive structure of the system we are modeling for VVUQ? Explore the conditions for the emergence of resiliency and diversity.
Possible approaches:
1. Jackson Mayo. Work with simpler models (easier to analyze, bound, understand its behavior, can analyze multiple, diverse, simple models). Resiliency, adaptivity, stability can take advantage of constraint we put on design. DoD definition of resiliency: ability to adapt to new requirements. Idea: design of a resilient system will make it more amenable to V&V. Two ways to make a system resilient: one is constraining it, one is to open it up (moving target defenses).
2. Adaptive competition: competing subject matter experts (SMEs). Craig Lawton: Qualitative vs. quantitative aspects of model design. Two types of qualitative: inherently vs. lack of data (rely on SMEs). Have one set of SMEs to develop input data, another set of SMEs to examine is it reasonable. Second set of SMEs provides the "face validation." What to do if it is valid or invalid? Quantify or bound how the SME differences introduce error into the model. Develop metrics for resiliency and diversity.
3. (Sasha and Jackson) Adaptive updating: Turnaround of model is fast, and the model is updated based on previous iteration and current data.

Call: Develop methods for confidence in the credibility of the complex system model, accounting for the purpose of the model.

1. Confidence in the credibility of a model. Steve Verzi: Use methodology to do data compression, apply theory of PAC (probably approximately correct) machine learning to modeling of complex systems. Let's give up on trying to prove model is correct, just estimate the quantifiable error (probability the approximation is good). Supervised learning: need "truth" data.
2. Sasha Outkin. Decision-driven development of models. What you get out of the model is useful for making decisions, even if it is wrong (stock market trading). Good if you can predict anything, not everything. Purpose of modeling is more utilitarian (applying model for decision), but if partially correct model can give good decisions. Focus on achieving desirable outcomes from model vs. getting model right.
3. Dan Pless. Use data that you can acquire for validation vs. performing validation on customer question for which you may not have data. Design models not just to answer the question but also to be able to incorporate data that you have access to now and may be able to collect in the future.

Call: Managing complex systems with controlled, synchronized, real-time virtual systems
1. Approach: Have a virtual system model of the system which is synchronized with the real system and is used to provide control.

Call: Develop methods for inferring the underlying microscopic rules (e.g. individual agent) from macroscopic behavior (system level).
Possible approach:
(1) Rules by which complex systems operate are relatively simple but parameter space is huge. Invert that order: provide a "dual" of the model, where rules become parameters and parameters become rules. Then, apply UQ/SA on the rule space. Loose analogy: in physical systems, we know all the little pieces (e.g. we understand the microstructure), but not necessarily how they fit together vs. complex systems where we know the input-output relationship of all the components (we know the macrostructure) but we don't know all the "inner parts" of the components. These inner parts are what we will try to infer/quantify the uncertainty.

Call: Develop accurate mathematical methods for macroscopic discrete time and space models of continuous and discrete systems
1. George Barr and Alfred Hubler. Re-examine model integer problems with continuous approximations. Rewrite the recursion relations (difference equations, maps), develop integer solutions to what are integer problems. Need to do model analysis up front to determine appropriate time-scales.

9. WORKSHOP MISCELLANEOUS INFORMATION

9.1 Photos

9.2   List of Attendees

Last	First	Affiliation
Adkins	Carol	SNL
Adams	Brian	SNL
Appelo	Daniel	UNM
Backus	George	SNL
Barr	George	Consultant
Brodsky	Nancy	SNL
Brown	Nathaniel	SNL
Brown	Tammy	SNL
Chen	Wei	Northwestern
Cook	Ben	SNL
DeLaurentis	Daniel	Purdue
Denman	Matt	SNL
Engi	Dennis	Consultant
Finley	Pat	SNL
Flanagan	Tatiana	SNL
Frazier	Chris	SNL
Gabert	Kasimir	SNL
Griffith	Richard	SNL
Hoover	Marcey	SNL
Hubler	Alfred	UIUC
Johnson	Curtis	SNL
Jones	Kat	SNL
Jones	Dean	SNL
Keller	Elizabeth	SNL
Kleban	Steve	SNL
Lave	Matt	SNL
Lawton	Craig	SNL
Lilje	Anne	SNL
Mayo	Jackson	SNL
Mockus	Linas	Purdue
Nateghi	Roshanak	Purdue
Naugle	Asmeret	SNL
Oberkampf	Bill	WLO Consulting
Parrott	Lori	SNL
Pless	Dan	SNL
Romero	Vicente	SNL
Siirola	John	SNL
Smith	Mark	SNL

Stewart	Jim	SNL
Stracuzzi	David	SNL
Sundaram	Shreyas	Purdue
Swiler	Laura	SNL
Thompson	Bruce	SNL
Trucano	Tim	SNL
Tsao	Jeff	SNL
Turnley	Jessica	Galisteo Consulting
Verzi	Steve	SNL
Vugrin	Eric	SNL
White	David	SNL
Wildey	Tim	SNL
Witkowski	Walt	SNL

Academic Alliance Portion of Workshop and Presentations:

The afternoon of June 23rd, several participants from the Academic Alliance Schools (Dan DeLaurentis, Alfred Hubler, Shreyas Sundaram, Linas Mockus and Daniel Appelo) had an opportunity to present their current research ideas and interests. These presentations can be distributed upon request through Sondra Spence.

9.3   Instructional E-mails

Instructions and Overview – 5/20/2016

Dear CSVVUQ Workshop Participant:

Thanks very much ahead of time for your participation in our June 22-23, 2016 workshop on "Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation and Uncertainty Quantification." We look forward to a stimulating exchange of ideas and to the development and articulation of a long-term research agenda for this exciting new scientific domain.

Enclosed please find a draft agenda (we will separately provide details on local hotels, directions to the workshop site, and other event information).

- On Day One morning and early afternoon, we are planning plenary talks to help us frame the problem as well as concretize it with specific examples from the electric grid and cyber domain. The rest of Day One afternoon, we want to hear your ideas in this problem space. To that end, **please send Sondra Spence (sspence@sandia.gov) by June 1 a one-paragraph brief Research Idea (as well as one-paragraph brief bio) that we can pre-organize into topical track sessions**. Your Research Idea can be informal and speculative, and you should feel free to change it in real-time in response to the plenary talks you will have just heard; we are planning for a five-minute time segment for you to present your Research Idea, followed by a five-minute group discussion.

- At the end of Day One, having heard the problem framing and your colleagues' "seed" Research Ideas, we will be asking each of you to propose a few "second-pass" Research

Ideas. These "second-pass" Research Ideas could be exactly one of the "seed" Research Ideas, could be a synthesis of a few of them, or could be brand new. These "second-pass" Research Ideas we will bin, and Day Two will be devoted to discussing/adding/organizing them into Potential Research Areas and Sub-Areas.

Our hope is ultimately to articulate these Research Ideas, organized into Research Areas, in a journal article to be shared with the larger research community. To that end, we will be asking for volunteers to help us post-workshop with further discussion, development, and co-authoring of that article. We will also be asking some of you (stay tuned!) to help with leading/facilitating various portions of the workshop itself.

Finally, please note that, after the main workshop adjourns at 2:30 p.m. on June 23rd, there are planned follow-on activities associated with Academic Alliance participants. We ask each of our Academic Alliance participants to bring a ten-minute presentation outlining your research interests and potential areas of collaboration, without feeling restricted to the topics discussed in the main workshop. Sandia participants in the main workshop are encouraged to attend the Academic Alliance activities as a means to catalyze potential future collaborations.

Thank you, again, for your time, interest, and participation in this important activity. If anything is unclear, please don't hesitate to contact one of us.

Best regards,

Jeff Tsao (jytsao@sandia.gov), Steve Kleban (sdkleba@sandia.gov), Tim Trucano (tgtruca@sandia.gov)

Sandia National Laboratories

Seed Research Idea Reminder 1 – 6/2/1026

Dear CSVVUQ Workshop Participant:

Thanks much for the "seed" Research Ideas many of you have submitted these last few days. We just wanted to remind those of you who haven't sent yours in yet to please do. We will be "binning" these next week into our Day One track sessions.

We wanted to emphasize that this is indeed a "working" workshop, and we are looking forward to all of us sharing ideas. We encourage all of you (including managers, students, discussion leads, facilitators – just not our plenary speakers, who will be sharing their ideas more formally) to submit a "seed" Research Idea to share with the workshop. Again, these can be speculative and informal, no need for extensive preparation.

Thanks again, and we'll keep you posted as the workshop approaches.

Jeff Tsao (jytsao@sandia.gov), Steve Kleban (sdkleba@sandia.gov), Tim Trucano (tgtruca@sandia.gov)

Sandia National Laboratories

Dear CSVVUQ Workshop Participant:

By now, you've received from Sondra our pretty-much-final agenda. It has listed our Day One breakout sessions (the result of our binning your "seed" Research Ideas): New VVUQ Theory, Concepts in Complexity, and Decision Making & Applications. **If you didn't submit a seed Research Idea, please do, it is not too late -- the workshop will benefit from your ideas.** If you did submit a seed Research Idea, please send Sondra your slides, if you would like to show some, a day or two before the workshop. We will have an abstract booklet available at the workshop, so you will be able to read the abstracts of all the seed Research Ideas. The abstracts will be organized into the three above breakout sessions, so you will know which breakout session to participate in. We will send a preliminary abstract booklet out in a few days (it won't be final as some new seed Research Ideas might be submitted and we might also do some last-minute re-shuffling).

Regarding the three-to-four "second pass" Research Ideas that you will be generating at the end of Day One, there are no constraints. They could include your own or someone else's seed Research Idea. They could be brand new ideas that emerge as you listen to the Day One talks and engage in discussion. As one of you pointed out, no one will have heard all of the seed Research Ideas, unfortunately, as they will be presented in the breakout sessions. But, again, we will have an abstract booklet available at the workshop.

For our Academic Alliance Participants, post–workshop we will be gathering to explore collaboration opportunities. It would be very helpful if you were to give a 15m presentation on your research interests (and also send your slides to Sondra a day or two before the workshop). Those research interests could intersect the topic of the workshop, but they do not need to. Sandians who were part of the workshop are invited to attend, but we will also be inviting other Sandians in other domain areas. So presenting the full range of your research interests would be appropriate.

Thanks, and we're looking forward to a productive workshop!

Jeff

10. WORKSHOP PRE-WORK

10.1 2016 March 28 Mini-Workshop Agenda and Overview

Exceptional service in the national interest — Sandia National Laboratories

Towards Verification, Validation, and Uncertainty Quantification of Complex Systems Models

1. Overarching Goal + Plan (15m)
2. CS VVUQ Team + Recruiting (30m)
3. Mini-Workshop Agenda (15m)
4. Classes of KOIs and Uncertainties (15m)

Work at Sandia National Laboratories was supported by Sandia's Solid-State-Lighting Science Energy Frontier Research Center, funded by the U.S. Department of Energy, Office of Basic Energy Sciences. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

CS Staff Brainstorming · Tentative Plan · 2016 Mar 07 · 1/7

Overarching Goal:
Map out a 10-year research plan for quantifying knowledge gained from simple-to-complex systems *models*

Systems:
Simple to Complex

Feb-Jun

Organizing Principles:
A Uncertainties (Steve V Tatiana F)
B KOIs (Mark S Asmeret N)

Hierarchically Abstracted Exemplars:
C Cybersecurity (Steve V Asmeret N 1 Domain staff)
D Smart Grid (Mark S Tatiana F 1 Domain staff)

Mar-Jun

VVUQ:
Old to New

E Extend VVUQ already developed for complicated systems (2 VVUQ staff 2 CS staff)

F New VVUQ that doesn't yet exist for complex systems (2 VVUQ staff 2 CS staff)

Apr-Jun

Calendar:

Mar 28	Internal ¼-day mini-workshop (A-D)
Jun 21-23	National 1.5-day workshop (A-F)
Aug 30	Workshop report / paper?

CS VVUQ Team and Recruiting

- **4 CS Staff**
  - Asmeret Naugle (1463), Mark Smith (6133), Tatiana Flanagan (6921), Steve Verzi (6132)
- **2 Domain Staff**
  - Cybersecurity (Jeff will recruit)
    - Stephen (Todd) Jones (5638, was PI for the moving target defense work that was used as an exemplar last year)
    - JD Doak (8962, is leading/teaching a cybersecurity course given at Sandia, and knows many other people from 8900)
    - David Stracuzzi (1462)
  - Smart Grid (Mark and Tatiana will recruit)
    - Matt Lave (6112) (PV and distributed systems integration)
    - Kevin Stamber
    - Jack Flicker (1768)?
- **6 VVUQ Staff: Tim T and James S will make suggestions, Jeff will recruit**

196

March 28 Mini-Workshop Agenda?

- **8:30-9a** — **Overarching (practical, not philosophical) Goals**
- **Today's Goals/Plan**
- **9-10:30a** — **4 CS Presentations (15m+5m)**
  - **Organizing Principles: Uncertainties, KOIs**
  - **Exemplars: Smart Grid, Cybersecurity**
- **10:45a-12p** — **VVUQ Roundtable Comments 6x(5m+5m)**
- **12-1p** — **Lunch**
- **1-2p** — **2 Breakout Sessions**
  - **Opportunities for Extending Existing VVUQ**
  - **Opportunities for Developing New VVUQ**
- **2-2:30p** — **Report Back**
- **2:30-3p** — **Next Steps**
- **3p** — **Adjourn**

Knowledge-of-Interest (KOI) as Organizing Principle for CS VVUQ?

10.2 2016 March 28 Mini-Workshop Participants

Attendees: Jeff Tsao, Brian Adams, Asmeret Naugle, George Backus, Bill Rider, Matt Lave, Tim Wildey, Mark Smith, Kasimir Gabert, Dean Dobranich, Stephen Verzi, Laura McNamara, Tatiana Flanagan, Tim Trucano, Sondra Spence, Miles Hall, Katherine Jones, Patrick Finley, Steve Kleban, Christina Ting

10.3 2016 March 28 Mini-Workshop Note-taker Notes

1. **Jeff**-The overarching goal for this meeting is to scope out potential intersections between the complex systems and VVUQ groups. I also wanted to try and describe the working definition of "complex systems" and solicit feedback on that.
   a. **George**-The model and domain of a complex system is inherently hard to manage and to define. I agree that the model captures the adaptive part of the system, but the key issue revolves around how the output compares to the outcome; complex and adaptive systems can be very different.
   b. **Tim Wildey**-Regarding the term "knowledge of interest", do you have any examples that we should keep in mind?
      i. **Asmeret**-Insight or policy outcomes, both from a quantitative and qualitative standpoint.
2. **Tatiana's** presentation on challenges of difficult-to-complex systems models
   a. **Brian**-For the first V in VVUQ, can you give us some examples and solution techniques and how they're implemented? We are not looking at finite element codes here.

      i. **Tatiana**-In slide three for instance, there is dimensional reduction; in the case of systems with a lot of data and low system to noise, there is pattern recognition. I do not think that uncertainty quantification is well understood yet.

          1. **Brian**-What simulation technologies are in use?

              a. **Bill**-With an atomic level of simulation, you always have to verify it.

              b. **Tatiana**-Uncertainty quantification is contingent on what models you're using.

              c. **Steve Verzi**-How do we validate the feedback? Regarding the first v, how do we know the agents are representative of what they're supposed to be?

  b. **Jeff**-One thing that has come up is that machine learning or deep learning is based on correlation, so there is no need to have causal models. This is becoming a bigger deal in terms of how you understand systems. There is causality built in deeply here at Sandia. However, there are some aspect-to-correlational applications if you don't have models; can you build something at that point that builds confidence in the correlation?

      i. **Tatiana**-Reconciling soft data when things are fuzzy is always a challenge, especially when you are trying to do uncertainty quantification.

3. **Matt's** presentation on complex systems and the electric grid

  a. **Jeff**-So for this sensor data integration system, are these modeled separately or integrated?

      i. **Matt**-Currently there is not enough sensing to integrate the modeling but it would be interesting to see how to integrate the data.

  b. **George**-What type of model is this?

      i. **Matt**-It is an electric model with multiple assumptions. The difficulty is in computing the model with many inputs.

  c. **Laura**-I assume industry is doing a lot of these models?

      i. **Matt**-In a static sense yes, but generally not in real time on the electric grid; utilities are very secretive about this.

      ii. **Bill**-How variable are these quantities of interest? Do we have a concrete understanding of what the variability looks like?

          1. **Matt**-A little, the quantities of interest are quite variable, so it depends on the various inputs. Lots of utilities don't know exactly how many inputs customer's photovoltaics are generating.

          2. **George**-What is the regularity size?

  d. **Laura**-One of the things you need for V&V is data. Sometimes data is not shared which is an issue, especially if the data you need is proprietary in nature.

      i. **Matt**-It's not so much a propriety issue; it's more like they are not using the data so it's not as useful as it could be.

  e. **Jeff**-In some cases still there is not enough data, but in ten years lack of data will likely not be an issue. If we can focus our efforts toward methodologies we will be ahead of the curve.

      i. **Laura**-We also need to isolate what data is valuable and if it is collected correctly.

  f. **Dean**-Regarding the big utilities, is there optimization with this?
    i. **Matt**-Yes, there is some control over storage and operation for regularity policies for optimizing the grid.
  g. **Bill**-When simulating, is there more of a focus on multiple or single applications?
    i. **Matt**-We've done both; for multiple, we would focus on a question such as 'how much photovoltaics can the distribution grid handle?'
    ii. **Bill**-What about the single simulation?
      1. **Matt**-One scenario could be focused on analyzing if you had three photovoltaics at a single point. We do not model the human input though.
    iii. **Brian**-Regarding the human element, when they put smart meters in, people seem to have differential pricing; is there evidence that people's behavior changes?
      1. **Matt**-Going forward this human aspect is going to be increasingly important. For example, getting paid for the time of use is going to be a major driver for the consumer market. Currently, people do not seem to have been greatly impacted yet.
  h. **Tim Trucano**-how long do you run these simulations for?
    i. **Matt**-Typically for around a year. While it is interesting to look at week-long simulations sometimes, that is mostly a worst case scenario exercise.
4. **Kasimir's** presentation on complex VVUQ
  a. **Jeff**-Does your virtual machine have the time delay in the network accounted for?
    i. **Kasimir**-Yes, that is an ongoing effort.
  b. **Steve Verzi**-Are the distinguishers a threshold, or is there an underlying proxy?
    i. **Kasimir**-There is an absent value for two term machines, with one real and one emulated.
  c. **George**-We want to reduce the UQ, so the argument is to allow a certain level of failures to keep us honest as opposed to instilling complacency that nothing can go wrong.
  d. **Laura**-For the past 20 years there's been a data revolution, so there has been retrospective bias. We are also still learning about these systems and how unpredictable they can be.
    i. **Kasimir**-For solution verification, some of these quantities can be adapted to analytics solutions. You should be able to measure the virtual or emulated meter. Another takeaway from this LDRD is that the quality of interests comes from the question; as soon as you change a single question, you might have to reevaluate the model.
  e. **Brian**-Is there a notion of approximation error?
    i. **Kasimir**-Yes, can adjust when verifying. With the infection graph over time, we went from malware to agent-based simulation, and then did calibration to line them up.
  f. **Jeff**-When deciding what quantities of interest to pull, how much was driven by the level of ease it would be to VVUQ? Or was it mostly driven by the question?
    i. **Kasimir**-Mostly based off of the questions being asked.
    ii. **Tim Trucano**-There is a meta-discipline behind this project; all validation means to us at this point is that there is a comparison to a model that is

supposed to be real world. Judging whether the model was actually predicting what is was supposed to be predicting a major challenge. We need to sharpen the definition of validation, and base it off of subject matter and the problem at hand. We also need a definition for what validation data means. Even if you need the data for validation, it doesn't mean you can actually measure the data. The quantity of interest is very meaningful, especially as it relates to validation.

1. **Jeff**-Sometimes there are quantities of interest driving toward a VVUQ, even if the question has not been asked yet.

   iii. **Tim Trucano**-Part of the question is producer vs. user error. Looking at the question is going to be very different depending on respective point of view.

   iv. **Brian**-What do your customers do with the models and validation at this point?

      1. **Kasimir**-Sometimes they are engineering based, other times they are looking at the past and seeing what else could have been done.
      2. **Brian**-It is useful to look at how these models would be used to influence decision making abilities.

5. **Asmeret's** presentation on knowledge of interest
   a. **Jeff**-For basic research, is everything about decision making?
      i. **Steve Verzi**-What about an autopilot directly linked to the control? What happens if the info is bad?
         1. **Asmeret**-It might mean that the systems were not that complex.
   b. **Laura**-One of the things with V&V is that we conceive of a benchmark and then it to measure how close we are and how good something is based off of that benchmark. Where mod and sim is best is if there is a community built around it to shape data collection and experiments.
   c. **George**-When we talk about VVUQ for complex systems, both sides of this discussion have questions about what that means, and how solid our collective knowledge is of the system. Those questions significantly affect how we build the model and do V&V with the model.
      i. **Bill**-One of the issues is thinking about what 'good enough' means. Models inform decisions but are not the basis of the decisions. If scenarios are key then we should use models to try and understand the dynamics of the system at a very deep level.
      ii. **George**-Who exactly is the 'we' in this?
         1. **Asmeret**-Me, Mark and Jeff primarily, along with some conversations with Tatiana and Kasimir.
   d. **Bill**-What is important is zeroing in on this knowledge of interest notion of success stories; things we've understood and confirmed through validation that have truth to them or explanatory power. In terms of mod and sim, these stories are very important.
      i. **Asmeret**-With many of our initiatives validation is very hard to confirm. However, we have been able to provide insights that have proven to be accurate.

      ii. **George**-This goes back to the science of falsification. If there is no way to make something equal, then we can be close enough to determining that it's not going to work. Using the model to explain why a decision will not work helps a lot. Within the idea of understanding causality, modeling is useful.

      iii. **Laura**-A lot of this also depends on the domain. For instance, there are communities that accept mod and sim in decision making but others that define it more as a research tool. We have to be careful to not fetishize the model. It is no more than a grouping of math and software; we must take a step back and focus on V&V and make note of what we are trying to model.

          1. **Brian**-We also have to acknowledge the increasing complexity in the present day, which is more frequently dictating our use of mod and sim techniques.

      iv. **Bill**-There is a reoccurring view of mod and sim, that it is the final solution to reconcile the chaotic world. There is a large degree of illogical exuberance with this; V&V can be perceived as a way of extinguishing the mod and sim optimism.

6. **Tim Trucano** -Round Table discussion for V&V
   a. **Dean**-Some ideas that have resonated with me are that 'all models are wrong but some are useful' and 'an experiment is viewed as the truth partially exposed, while a simulation is a half-truth fully exposed'. We must find synergy between these two concepts.

      i. **Tim Tucano**-The community understands mod and sim is just one part of building a body of evidence.

      ii. **Dean**-Optimization under uncertainty is important, especially considering the uncertainties going into it.

          1. **Tim Trucano**-What is more dangerous, the experiment or the simulation?

              a. **Dean**-The synergy is the most important part. There is less danger than if you just consider one at a time.

      iii. **George**-In the sense of progressive hedging, this is interesting feedback for the question of whether a process is deemed to be 'good enough'. This area is currently in development, with contacts such as John Hart and JP Watson working on it. This work should be more robustly distributed throughout the lab.

      iv. **Patrick**-I have found that if you couch the presentation in the framework that 'this may not be the very best answer but it is a good and safe answer', people are more willing to present their findings and take a stance.

   b. **Tim Trucano**-In terms of qualifying a weapon for example, there are different elements of optimization feeding into a body of evidence, which then feed into a holistic sense of judgement.

      i. **Mark**-So we are not just looking at the tail of the event?

          1. **Dean**-No, we must look at the whole continuum, based off of best estimates.

2. **Tim Trucano** -We can do a lot of work providing evidence within a margin of safety but the key thing is that even if the work is quantitative, it doesn't mean that the final decision remains quantitative. Ultimately, many of our decisions are qualitative in nature. Within agent-based models, qualitative decisions are becoming more frequent.
3. **Jeff**-The more orthogonal the perspectives, the easier it is to zero in on whether a decision is correct or not.
4. **Tim Trucano**-The evidence is filtered through Dean's expertise. V&V evidence is more like a legal argument than a scientific one, with qualitative pieces of evidence arising frequently.
5. **Jeff**-Is there a measure of orthogonality?
   a. **Steve Verzi**-Statistically, there are measuring tools that can be applied.

7. **Tim Wildey**-V&V and UQ is often labeled as something different, so people might be participating in it without acknowledgement. How do you all define V&V and UQ and the specific applications associated with them? What about characterizing reducible uncertainty?
   a. **Tim Trucano**-Reducible uncertainty equates to lack of knowledge.
   b. **Kasimir**-In the cyber community there is brainstorming everything that might be important in a model, and then ranking in terms of importance, and then focusing on the most vital aspects. Dealing with uncertainty, we have run sensitivity studies, and have built lab models outside of the virtual models and then tried to test.
   c. **Matt**-A lot of it goes back to whether something seems right, which is not very quantitative. Generally at the beginning, validation is a loose quantity, so there is a lot of room for improvement.
      i. **Tim Trucano**-Does validation have to be quantitative?
      ii. **Laura**-No but it should be systematic.
      iii. **Matt**-For us, there is a divide between qualitative questions like 'are you allowed to put photovoltaics on your rooftop?' vs. more quantitative questions like, 'if you exceed a certain level, what penalties do you then get?'
   d. **Brian**-What is the state of the human modeling validation?
      i. **George**-There are a lot of views on human behavior; from an economic point of view, using stochastic or agent based models are one of the most difficult places for this.
         1. **Bill**-This goes back to whether the models we use are being used correctly and applied in the right areas.
   e. **Brian**-If you use an agent based model, how do you evaluate a single agent?
      i. **Steve Verzi**-There are microsystem models for single people.
      ii. **Laura**-With an agent based model, what can happen is that you can have a model that looks correct, but could still be wrong based off of incorrect assumptions.
      iii. **George**-The sampling processes for the models could also prove to be invalid.

f. **Jeff**-So it seems like there are two main drivers with people: they are uncertain, and they adapt. For the problems that we care about, which class is most relevant with respect to agent-based modeling?

    i. **Laura**-This is a sociology vs anthropology vs psychology question, with fundamental attribution error tied in. We must look at the degree of influence of the human network on the individual, as it is very difficult to model individual agents because there are so many influential variables.

    ii. **Tatiana**-Usually we try and trim the assumptions for the model as much as possible.

g. **Jeff**-We often care more about figuring out how to avoid worst-case scenarios. For grid and cyber, is that true?

    i. **Matt**-More or less, utilities' biggest concerns are with respect to targeted attacks on the grid. Our work is about integrating renewables at a certain rate and mitigating expensive upgrades.

8. **Bill Rider**-One of the biggest things I have noticed thus far in the conversation is the notion of 'what kind of questions are you asking' and 'are your models appropriate to answer these questions?' All of us are interested in the average behavior of the given system, but then there's also how we understand or engineer our systems to keep the really bad things from happening. The classical community is invested in what they're already doing with average behavior. Increasingly though, society is asking about worst case scenarios. Is the model we are solving suitable to addressing the question we are asking it? Frequently we are asking questions that models can't robustly address. What does V&V mean for both the classical side and the more recent side? There is a core of determinisms regarding how we look at validating and science. There is also a core level of uncertainty that will not go away, simply because we cannot account for everything in an increasingly complex environment.

a. **Brian**-There is a broader context for V&V and UQ. I don't have a firm idea about how my tools are used, so getting the classical physics community to show how the tools are being applied will be important.

b. **Laura**-What is the goal for this workshop?

    i. **Asmeret**-Generating goals for addressing issues for the larger workshop.

    ii. **Steve Verzi**-Our external advisory board suggested hosting a workshop, so we want to work with the V&V community to generate communication both ways.

c. **George**-At the moment, how can we use uncertainty to our benefit? Knowing your model is within a certain set of boundaries can help make decisions. When we do agent based modeling, if the classical modeling does not align, then our level of understanding can dictate our decisions.

9. **Brian**-Jeff asked about research destinations. I am convinced that computational science is often not science. Using models and computational models is often not scientific in nature. Can we put V&V discipline into those models? With UQ, do we need to redefine our parameters? Could reduced essence be applicable?

a. **Tim Trucano**-Parsimony is assumed, while parametric uncertainty is based off of correlation vs causation. Epistemic uncertainty does not provide enough of a foothold in some of the models.

b. **Dean**-It's very difficult to log what is occurring within some of these complex models. With simple models, you need to understand it much more thoroughly. You can find synergy between these two concepts by building a continuum of models. What questions are you really trying to answer? Often times you think you know what question you're trying to answer but mod and sim forces you to understand the other causes and effects of other questions. It really is a learning process.

c. **Patrick**-Regarding simple vs. complex models, NIH has funded very complex models focusing on people and attracting diseases. On the other hand, CDC has very few, simple models. With the Ebola Crisis, there was a scenario for both models and yes they were more or less equivalent. The problem arose when you had to adapt to a new location; simple models were easier to use, while the more complex models were absolutely useless. This underlines the importance of understanding the assumptions you are working with when building the model and how those assumptions present barriers and limitations. There is a key difference between an operational and a planning model regarding time and actionable details. I want to see V&V used on the fly.

      i. **Laura**-Having a model as part of an iterative process is invaluable.

d. **Tim Trucano**-V&V needs to start on day one.

10. **Laura**-It is easy to get caught up in the various equations that make V&V more difficult. We must keep track of systematic V&V anchors. V&V supports organizational collective thinking about tools. The models do not make predictions, humans make predictions.

a. **Brian**-Is there some kind of V&V police system within this community?

      i. **Bill**-This goes back to the body of evidence; what will be accepted as pertinent evidence, what will be excluded? Verification evidence at Los Alamos is almost always discounted, at SNL it is looked at more thoroughly.

11. **George**-With resilience in complex systems, how do we quantify it? What do we do when we have V&V? Maybe we should focus on confidence with the intervention.

a. **Steve Verzi**-It isn't a single intervention, it's a suite.

b. **Tim Trucano**-The key is the inability to do validation. If you can't, then it's a risk issue, and you must then manage and account for the risk in the model.

c. **Laura**-It's not just about the physics piece, it's about how you perceive of these models.

d. **Bill**-Lots of times there is a lack of completion within these models, so we have to question whether the body of evidence is sufficient. The policing aspect becomes relevant when a community is stuck in certain validation habits without sufficient accountability and biases that might have generated success in the past.

12. **Patrick**-Sandia is well known for classic V&V but ad hoc methods are also quite advanced in appropriateness and frequency of use. There are still gaps in the models though. If we can get people to become comfortable with models that we take as second nature it would be beneficial.

a. **George**-The economic models are reproducible because of the community based around the models. We can learn from failure and bad choices. This community of practice is almost as important as the mechanistic domain.

13. **Mark**-There should be an ecosystem of learning rather than just a one-stop, one point of view process. There must be a continuum. I think we collectively believe in the process of building a model.
14. **Steve Kleban**-For research challenges, the pushback was based on theory. This year we've gone with a validation emphasis. Instead of just asking how we bring recognized VVUQ to the community, there is more of a shared emphasis on complex systems. What we are doing with validation is very difficult to do compared with what is being commonly published by other entities, so there is room for us to further explore this territory.
15. **Tatiana**-There is a monthly meeting we host for complex systems as part of the community effort.

10.4 Plenary Presentation Dry Run Notes

1. **Jeff Tsao**
   a. Introduction is solid
   b. Slide 2-Models that are adaptive and not difficult?-Botnets maybe?-Curtis
      i. What do we mean by adaptive?
      ii. Do we want to keep the continuum arrows? Tatiana
         1. Yes on the x axis-Curtis
   c. I have a rough slide summarizing Laura's yellow box -Steve V.
   d. If we just call it complex systems it might lead to arguments about what defines a complex system-Asmeret
   e. Note where the journal articles should be submitted, maybe produce a list-Laura
   f. Breakout session leaders? Wayne, Jenn, Laurie Parrot as a backup, Alfred, waiting for Dan DeLaurentis
   g. Are we expecting everyone to come together after the breakout session?-Tatianna
      i. Journal article could draw together 0-5 of our talks, don't know yet-Curtis
   h. Are there activities Friday?-Laura
      i. Academic alliance and dinner activities after, should mention it
2. **Laura Swiler**
   a. Steve V. will send over comparative diagram
   b. Big V and little v, do not need to define-Curtis
   c. Big V is more quantitative and little v is more qualitative, you don't need to change anything -Steve V.
   d. People will call it what they want-Curtis
   e. The purpose of this talk is to get an idea of how other people do it, don't change anything-Asmeret
   f. We assume that the model and experiment are both covering the same truth-Steve V.
   g. Are you running with a traditional V&V definition?-Curtis
      i. A large reference book came out in 2010 which gives a sense on how recent this all is. While we have bodies of documents about procedure and process and definitions, it is still an evolving piece of work. There is not a lot of consistency for metrics used, experiment procedure, etc. Sometimes it's very hard to do, with point series and different values-Laura
      ii. Is the scientific process based in formal validation? I am going to call it that and focus on careful science-Steve V.
      iii. I would suggest prefacing everything by saying that V&V is a relatively new field, that we can't do half of the things we want to do, and that we're trying to stretch this not yet fully developed science to fields it's not ready for-Curtis
   h. The bottom two bullets are excellent on slide 29. Kasimir can get as much data as he wants, he just needs to put a sensor out. It's not good or curated data, but we have a lot of it-Curtis
   i. There is lots of text on every slide, maybe also limit the number of slides-Eric
      i. Copies of slides, white papers, and research ideas will all be handed out-Jeff

j. The details for V&V for cyber can get very complex, might not be best to include here-Kasimir
k. Reword theory and data rich bullet-Steve K.

3. **Stephen Verzi**
   a. Mention classification on slide 5-Curtis
   b. The 6th bullet on slide 5 is specific to complex systems, might want to emphasize tie-ins-Eric
   c. Maybe have a bigger image of the Balci model in the background
   d. Validity is inserted as an absolute concept vs some acceptable tolerance? Eric
      i. Make sure to mention how you are defining formal validation
   e. Slide 11-I would argue that people run experiments on tornadoes and hurricanes-Eric
   f. 'Dymatica' is the title of the image at the bottom left of slide 16-Asmeret
      i. What is the difference between complicated and complex modeling equations?
         1. Regarding the diagram that Jeff was showing, the stuff that I'm looking at is more complicated rather than complex-Curtis's area is more complex-adding parameters to the lower left hand yellow box that Jeff showed is what I'm going for with tons and tons of parameters added-Steve V.
         2. Use the same wording in the quad chart as in your presentation-Asmeret
         3. Complicated is a climate model for example, if you add anything biological to it it's complex-Curtis
   g. The individual run vs model concept you used last time was useful, could be a good thing to bring up on slide 20-Curtis
   h. The graphics are very busy and hard to see, unless you go through the steps of them it's not easy to get anything from it. I would also simplify the graphics; the graphics on slide 16 should also be enlarged and elaborated on and there's a part of defining an ABM and saying it's not a complex model that should be taken out on slide 18-Tatianna
      i. Complex vs complicated can confuse people if you are not explicit-Curtis
      ii. I would recommend dropping slide 18, folks like Bill will know about emulators already, the message was not clear, are you using the emulator for insight or as a replacement in a validator exercises? I will not go into depth on the emulation, but if you think it's important for complex systems you should try and include it-Laura
      iii. When we say emulator we mean model we don't understand-Kasimir
      iv. There is differing terminology for emulators so it's not a great term, you might want to avoid it if possible-Laura
      v. I would just leave the emulator aspect out, it's confusing-Steve K.
   i. Spend a couple slides going into depth on the Balci diagram-Asmeret
   j. There is a slide talking about risk; you could say that the risk of using this model is low-Mark
   k. Maybe have a slide about what we mean by credibility, it could help frame things -Asmeret

l. On slide 4 you can do a modeling process that becomes adaptive-talk about how it could be an adaptive process as foreshadowing for Curtis-Curtis
    i. I don't like the term experimental model on the Diagram from Balci-Laura
    ii. We can make our won via a graphic artist-Jeff
m. On Slide 9 the inherent system unpredictability is missing-Curtis
    i. Very little value of predicting just one thing-Steve V.
n. On slides 10 and 16 it is very risky to bring up humans; when you focus on humans people start thinking about how humans are irrational-Curtis
o. Slide 14 is excellent, shakes us free of the notion that models are only SD or AV-Curtis
p. If you do talk about creditability at the beginning, is there another nuance you can bring up about the expenses associated with formal validation?-Jeff
    i. There is a moment in slide 14 where you can bring up that if the system is going to validate every 24 hours, there must be a cheap way of doing it; real time use is one thing we don't have-Curtis

4. **Matthew Lave**
a. Slide 5-What is traditional or nontraditional? They've got mathematical models to maximize profit for the electrical grid already-Eric
    i. Stochastic modeling is not routinely done-there are two kinds that are nontraditional, system and model, that are often blurred together; it is very confusing arguing about the definition of things-Mark
    ii. We're being inconsistent with terminology between Jeff, Steve, and Matt-Asmeret
    iii. Maximizing profit should be re-labeled, maybe replace with production efficiency or something that seems less human-Curtis
    iv. You might want to say what the violations are of line limits early on and emphasize why it matters-Curtis
    v. On slide 18, you can add in collection of this data; ask questions about whether we're getting the right data, how we can minimize it, what the risks are of optimizing that data vs collecting as much as possible, etc. I like the conciseness of slide 18-Also, it was not clear to me that there were three separate human operators in slide 14-Steve V.
    vi. I would emphasize the control systems aspect and also, with the stochastic optimization model validation, the control aspect with large state spaces and large numbers of control options- you really want to emphasize that the power utility can only control so much-Laura
        1. What does validation really mean for that kind of model? It is another animal if you're dealing with an unknown future event-Eric
    vii. Is the underlying assumption that the grid operator is who the model is made for? As opposed to Curtis's presentation?
        1. Yes, the biggest thing is that the grid operator is making the decisions, incentives will control how decisions are made from a customer point of view-Matt
            a. Is that where it differs from cyber then? You have to model on the internet to make choices vs. on the grid?-Jeff

       i. In the long-term everything will be a cyber system-Kasimir

      ii. In operation of the grid, it's the model of the operation of the grid that's being modeled here; when we add customers who add PV to the system, it could add another level of modeling that Curtis covers-it could also tie in with cyber in a sense with these actors who are not intending to be adversaries but that are intending to maximize their profits-Steve V.

    b. Hacking against the system is another thing we've touched on-Matt

  b. I'm voting for difficult vs non-difficult for terminology-Jeff

       i. I'm for traditional vs nontraditional-Asmeret

      ii. People might say traditional V&V is statistics based vs Basian-Eric

5. Kasimir Gabert

  a. There is a different metric on the y axis than what you describe on the previous page for the results area, maybe be more clear as what you are including-Asmeret

       i. Take the time to explain what the slide is on the results-Eric

      ii. Maybe include two bullets for elaboration-Curtis

     iii. Show before and after?-Mark

     iv. Either get rid of the legend or describe what it does-Steve V.

      v. Don't say it's not important-Mark

  b. Make sure to mention both the software and hardware components in your Emulytics environments -Laura

  c. Regarding the distinguisher, is it the inverse of the Turing Test?-Steve K.

       i. Say what you're going to program the distinguisher to do-Laura

      ii. If you had lots of criteria, what is your parameter for your distinguisher?-Steve K.

         1. It is a way of taking something that is undefined and producing an answer-Kasimir

     iii. How does the distinguisher tie in with the V&V?-Eric

         1. This would be the smallest building block-Kasimir

         2. Mention that is has to do with validation more explicitly-Mark

  d. Try and apply presentation terminology to people outside of Sandia, techweb for example-Mark

  e. We want to say that we can expand the V&V techniques so we can verify this model-Eric

       i. There is some truth to having to dumb down the model enough to be able to validate it-Curtis

  f. Did the distinguishers you built give you confidence about distinguishing one way or another?-Steve V.

       i. We did find that our models are insufficient for predicting small packet performance-Kasimir

  g. Can you add some text about the question you're trying to answer using the distinguisher?-Asmeret

6. **Curtis Johnson**
    a. I liked the model spectrum a lot-Kasimir
    b. Maybe use the swimming example for model adaptation-Eric
    c. Would you say that adaptivity requires intentionality?-Steve K.
        i. Depends what you mean by intentionality-Curtis
        ii. Random mutation can result in a dominant species depending on environment-Eric
    d. In your green and blue population examples, maybe make more explicit what the rules are in the model-Mark
    e. On slide 10, "the fecundity doesn't matter" is an untrue statement, also mention that the reproductive rate is in the assumptions-Asmeret
    f. Take out some of the examples that are less necessary, drop the 'born on date also matters' example, -Laura
        i. Maybe briefly mention that slide but don't linger-Kasimir
    g. Mention cyber or electro grid example instead of iPhone on slide 12-Jeff
    h. What's mutual adaptiveness? Are the pictures representative of that?-Eric
    i. What do you mean by adaptation? Is it intentional or random? Maybe briefly mention-Eric
        i. Adaptedness is the result of lots of generation, being well-suited to the environment, adaptation produces a variety of things because it does not know what kind of environment it will face-Curtis
    j. Include Sandia National Laboratories in your presentation, take blue text out of blue backgrounds-Sondra
7. Next big thing to do is the white papers-Jeff

10.5 Summary Transcripts of Plenary Talk Dry Runs

Matt Lave:

I'm going to talk about complex systems and the electric grid and give some examples for us to think about. On slide two there is an overview of the electric grid and the conventional way that it's been operated. On the far left in the black there's generating stations which could be coal, natural gas, or nuclear power plants. Those are generally big generating stations that produce the power that is then injected into the electric grid. From the generating station there's a transformer which takes it to the transmission lines which are spread across the U.S. These lines take the power generated and transport it to where it will eventually be consumed. Then there's the distribution grid which is where the electricity is delivered to the consumer. There are different levels of transmission grade customers including factories, commercial buildings, houses, etc. Now on slide three, it's interesting to note that the U.S. transmission grid is broken up into three separate areas: the western and eastern interconnects and Texas. Texas has its own transmission grid because if they secede they will still have power. Each of the different nodes could have some control; power can flow across the grid through many different paths. Different paths are used depending on which generators are producing power and where the power is demanded. Because of that, it is an adapted system. The power flow magnitude and direction will change both on short time scales and long. A significant effort is spent on outage planning which are low probability events. Often this type of planning is referred to as accounting for N-1(if you lose a component of the grid, you lose one of these electric lines or a generator on the transmission

grid) contingencies. There are always operating plans in place to account for what will happen if you lose one of these nodes.

Kasimir Gabert:

Roughly, you can look at cyber security as the behavior of technology and the behavior of people. With people, you have a very adaptive and complex system that's extremely difficult to model. There have only been a couple of cases at Sandia where we've really tried to model human behavior in cyber space. Most of our work has been concentrated in modeling the behavior or technology, including people when necessary to get a system to work. Cyber systems are built on top of computers which you can view from numerous different levels. At the lowest level you have transistors and electrons whose performance has been demonstrated to be chaotic at times. You can imagine a single bad bit coming over the wire and causing a massive state change in the entire system. A good deal of computer research is spent on trying to find these single bit phase chains. It's unlikely that any of the developed models will find them. However, we would like to validate the findings of cyber security events. Then, you can work your way through numerous levels from transistors to components to computer systems to massive systems with hundreds of thousands of computers. One example came in 2011 when Amazon had a massive outage. Basically, there was a minor configuration error while some machines were being updated. After the error had been corrected, all of these machines were out of sync and promptly shut themselves down while transitioning to a different data center as part of the design for resiliency. As a result, there was a massive re-mirroring storm that continuously propagated and resulted in a huge number of outages. Another example is thinking of systems like bitcoin, where you have all sorts of human interaction. The whole technology works because of miners and a financial incentive to burn electricity. This runs all the way between different nodes in the system, lightweight bitcoin nodes, all sorts of auxiliary servers that help translate people's quarries and propagate things. You end up getting massively complex systems. I will focus on a specific cyber model at Sandia that we've developed that hopefully we can validate.

Curtis Johnson:

I'm Curtis Johnson and I'm going to go back to a high level overview. Kasimir gave a really hard example of how to validate a specific model, now we're going to go way up to what little theory we have in complexity and that's mostly from biology and adaptive complexity. I have two goals for this talk: to lay out the fundamentals of adaptive systems, and then to use that to illuminate v&v challenges in this space. There is no consensus on many of these principles.

So everything in adaptive complexity is based on this process, the natural selection algorithm. We have some candidates and organisms that might recombine and mutate and create new organisms. We ask out of those new organisms, which can reproduce? Those that are better adapted to the new conditions and can accomplish those new functions are the ones who have more offspring. I'm going to show some really simple models to illustrate the adaptive process.

My starting point is a totally stable homogeneous environment, with only asexual reproduction, unlimited resources, no predators, no food chain, and no illness. Each concentric circle is a generation; even when I have two species, they have the same generational cycle. Let's say the green population is immortal and the blue population only lives to see their grandchildren. We go

from 1 to 2 to 6, and remember, because no one dies, all of these are alive. The green keeps growing, the blue dies and can only see its grandchildren. So green continues to grow faster than blue and they continue like that. Notice that green is bigger than blue but these two populations are in competition in no way, resources are unlimited, etc. These populations are a simple function of a linear thing; fecundity and longevity are all that matter. Now let's add copying errors. Let's say in the third generation it produces some weird thing. So that evolves very simply, the new thing starts reproducing unbounded in its own way, but our original blue population is largely unaffected; it gets to grow unbounded while the new thing grows unbounded next to it. I want to note here that we get these small mutations that create new populations so, unlike in physics models where we talk about a population of ants or in Kasimir's world where we talk about a population of computers, no two are exactly alike. This is different from Carbon Dioxide molecules where one is exactly like the other. When we're in these adaptive complexity worlds we have this cloud of pretty similar things but we don't have identical things. A modeling challenge is to decide whether that matters or not.

Here's another simple example: let's say green starts a generation earlier than blue. They both produce unbounded with green bigger than blue. Now, what if suddenly we bound the resources represented by those white squares. Let's also assume for both populations that the G-3 population died just before the current generation is produced. This scarce resource would immediately go to food but it could be a variety of different things. Now, if you're one of these critters on the inside, if you get your hands on one of these boxes you can produce two offspring. If not, you don't get to produce that generation. We will also give blue a slight advantage for getting those resources in that it gets a higher percentage of the boxes as a function of the total population. At this point we have way fewer blue than green because blue started later. But with that slight advantage resource wise, you get blue going from 25% of the population with scarce resources to 38% in just three generations. We went from models where fecundity and longevity were all that mattered to models where they almost don't matter, with the slight advantage given to blue overwhelming them almost completely. If we ran this model out 4 generations, ceteris paribus, green would die off. Suddenly, we've eclipsed those original models. Now what matters are competition and a relative performance for the scarce resources; absolute performance doesn't matter very much at all. I want to point out that with adaptive complexity, not only can the means change but the goal can also change. This will happen naturally in the real world with resources running low. Even the goal can change in a closed adaptive system.

Now I'm going to try and generalize this process by talking about any adaptive system, which includes engineered systems. Back to the original process, we have a bunch of entrants, we had organisms before, and we have some test process which has to be a similar process for all of the entrants. It does not have to be exactly the same. And then in some of them they survive biologically but all you need to say is that they met some threshold. So it doesn't have to be death that stops them. Then you have a generating process to create the new set of entrants; all we limit is that it has to first do something with prior design information and doesn't generate new information out of nowhere. Second, it has to account for who survived and who didn't. Taking that information, it takes past designs and modifies to produce a new generation. So this is how I'm taking the biology and saying 'this is the new process'. It uses the new designs and uses the success and failures of the old designs.

One example is that if we have different smartphones, and the width of these ovals applies to their current market share, maybe my test is sales and profits. How much of the market do I get, how much money do I make off of it, etc. Say iPhone loses a little bit of market share, everyone goes back and tries to come up with new designs; we don't know what the new designs are going to be, that's going to play out in the next generation. So that's an example outside of biology with generating a test process.

If we wanted to name this process, we've got this generator. It could be a manufacturing plant where you have entrants that have some kind of structure and way that they work. This could be a physical structure or in Kasimir's world or in biology it could be an informational structure. They endure a test where they try to do their function in a given environment and they have some relative success at that and that outcome data gets fed back up to the generator. A simple example would go back to the first model where you have a parent that produces offspring that survive or don't to reproduce which produces this generations gene pool which feeds back up. In this case, the offspring are the generator which is quite unique about biology, with the new parents carrying the results in their DNA within the whole gene pool. So the generator is really an entire population and not just a single parent.

Let's go back to a commerce example where you have a factory that produces a product that has sales and profits. There's feedback, as if the product produces too little profit the whole system dies. There's also feedback from the customer. The factory looks just like proteins; they can't replicate themselves just like iPhones can't. Sure enough though, DNA and RNA are used to fabricate proteins. Those proteins contribute to the survival and reproduction of their host and the organism, which changes the generation of this gene pool. I want to point out that with these test processes, functionality is born. With functionality, there is a goal, with survival leading to reproduction. Those that survive have this property of making themselves last and making their offspring last. Generating tests produces flexible means to achieve the end with the first fecundity-based scenario being a great example of this.

These cycles create hierarchies of ends and means and we'll see examples of that; we've already seen how the end can change. There are huge benefits to these generated tests cycles because they innovate and are creative. After several generations, the entities that you have are acutely adapted to whatever environment that they've survived. It's the difference for a company in buying an empty factory and buying another company. When you have a generated test cycle, you have this productive fertile thing that gets stuff done. When you have an empty factory you have a lot more stuff to do. You have to watch things that are alive though. The problem with generated test cycles is that they produce things like cancer. Our cells are all over the place which can result in problems. Because these test cycles have this optimization over time function, it does make one want to lean toward that kind of modeling. They have goals which make it tempting to skip past all of the individual generated test cycles and point to one general objective, and eliminate all of the little details and focus on the goal. However, based on my earlier point, the goal can always change too.

Another example of a generated test cycle is an investor, with specific bets and specific investments with prices that change. Profits and losses are then felt and knowledge is gained. Investors can also invest themselves out of the market if they're bad enough. Another example more related to V&V is if you have a machine learning classification of incoming network traffic

to distinguish between malware and goodware. I have the code which is the generator, it comes up with some candidate features from the data to try and distinguish between the bad and good. It tests against the data, which, in this case, the data is the environment. It either succeeds or fails; I then use some kind of method, maybe a rock curve to validate.

Another example, if you think of the concentric circles, you would think that there would be only one species because they slowly eliminate each other. One thing that maintains variation in the environment is that it is constantly changing and that it's not homogeneous. So if I'm being competed out of one niche, I can quickly go look for another, so you get lots of diversity. Of course the other critters are part of the variability in the environment. Over time, this creates very robust systems because you have to survive all of the cycles. If one is alive today, it has been at the end of a chain of endless success. Someone before you had to survive all of the environments that they were placed in to get here, which includes catastrophic events.

Over time, you get neutral adaptedness of the thing, its function, and its environment. You can guess a lot of things about a creature simply based off of a picture. You can use this mutual adaptedness of the structural environment to make a lot of useful inferences. I want to highlight some of the advantages of adaptive complexity for v&v, such as making inferences that can then be validated later. One thing that's hard about adaptive systems is that they retain all of this latent information that is difficult to tease out of the system. For example, I have x and y chromosomes. My x is active but doesn't produce female features in me because those features aren't exposed to selection in me. But I have a daughter and my x chromosome contributed just as much to her as my wife's did. So those conditions appear latent inside of me but when I have a daughter and conditions are right, suddenly they pop out. There is an iceberg phenomenon where you have a ton of functionality and info that is hidden unless the environmental conditions are right.

Organisms are basically local negative entropy machines that concentrate energy locally and they have to follow the second law. This makes them attractive targets for other organisms who want their energy and protein. They can eat each other but that's a one off shot. One thing that's nice is what organisms produce e.g. eggs, honeycomb, etc. Organisms are naturally drawn to each other. The next step was for one organism to train another organism to bring them food or energy like aphid herding ants for example or dogs.

I want to also point out that there are huge hierarchies of generated test cycles. On the biology side, you can have a population that's producing better and better individuals with learning cycles in them such as rewiring neurons or mitochondria. The same thing happens on an industrial cycle, with an industry cycling through companies who experiment with employees and try to get the most pleasure and benefit for their time. They have a design and sales cycle and suppliers and companies. These cycles are not the same, so it is really hard to model. You can have one generated test cycle where one organism needs another organism in order to succeed. Humans need each other to reproduce and to succeed. Mitochondria have to survive to reproduce and they have to reproduce into a human egg in order to change hosts. Ultimately, they only succeed when we succeed. This is another case where you can simplify the model. Even though mitochondria look independent, they really aren't. In business, the android platform gets updated, it gets installed on phones, it's either profitable and popular or it's not. When I run an android app I'm kind of a parasite on the platform. First I have to produce my app, and then I

have to get it downloaded which is tied in with the popularity of the platform. In these cases adaptive complexity might be simpler because you can say things like 'me and my mitochondria are a team'.

Now let's talk about models. Think of a continuum: on one end, there are things like a linear classifier or a single clustering algorithm or F=MA, very spare things. On the other end, there are meshed physics simulations, agent based models, emulations, very complex stuff. There are two extremes with these representational models. Photographs are appropriate examples of this; they're two dimensional, the color is wrong, leaving a lot of stuff out. On the other end of the model there are minimalist representations such as Georgia O'Keeffe paintings. I do not want to say one model is better than the other. It's tempting to say that the photograph is more accurate and has a higher fidelity. You have to keep in mind though that the photograph has many more instances of modeling error. It's correct to say that each model has its own uses and applicability. I'm really opening these models up in order to remind you all of model use pairs and of the idea that there isn't a perfect model for anything.

Adaptive modeling can be very hard because all of the types are heterogeneous, i.e. we have clouds of boxes, and we don't have the same kind of box. We have vast stores of latent and low information. Every generated test cycle has its own collection of local information which it may or may not ever share with other elements although it's using it and taking advantage of it. Think about Matt's model; everyone who owns a piece of the grid might be using their own or a shared model or be trying to model someone else's model. In adaptive complexity you see these important distant connections between things. With these generated test cycles many things try many different strategies. Threshold effects are also very common in adaptive models, which is how the latent information gets triggered. For example, when the temperature becomes great enough, a lot of genes become active in my body like a switch being turned on. This kind of thing happens all the time in biology. How can we validate that nothing important within the use-case was left out of the model? How do we capture the critical functionality and determine what is critical? Do we model the homeostatic mechanisms? This whole process can be very difficult and complex.

One interesting opportunity and problem is that economic and social models tend to focus on goals and functions. Lots of optimization models do this. Many engineers think, if there's an engineered system, we're going to model the physical system. It's a question of whether we should bother with that now. Most stock models prior to the micro trading era never bothered to model the trading infrastructure but modeled what investors are trying to optimize. Cyber and grid questions might be better served by these kinds of models rather than models that work so hard to emulate the physical system.

If you do the physical model alone, you miss things like if the powerline owners are trying to make a profit. Maybe they might find a reason to be happy with blackouts. In Kasimir's world, the malware person might not be trying to actually take down our network; they might just be after making some money. Embarrassment or fears of punishment are other things that physical models miss. People frequently do irrational things that don't make any engineering sense. How do we validate that a model pursues a goal or robustly or inconsistently performs a function like it's real-world counterpart? If you write an optimization function, how can you tell that it emulates the real thing? Mean field approaches (random distribution when we don't understand

what is exactly going on) can be very risky with this. A lot of times the erratic behavior of functional things is actually adaptedness. The most unlikely thing you'll see in biological behavior is randomness because there's been so much time to adapt. Nonetheless, the Chicago School of Economics proved that mean field models can be very effective. Loads of microeconomic and macroeconomic models assume wildly inaccurate things about people but still provide very useful insights for large scale populations.

The last big challenge is that you have moving parts within moving parts, with means, goals, and generated test processes changing. Adaptive complexity is game theoretic, based on outcompeting one another. In this case, it matters what population I'm in and what strategy I'll use. There isn't one best strategy or answer. There is also path dependency, with adaptedness based on recent paths in the environments that the system faced. So, some basic approaches: you can have relatively fixed, detailed kinds of models and operate them on small enough timescales where adaptation doesn't confuse everything. You can also have radically generalizable heuristics such as fight or flight, supply and demand, etc. These approaches can give you decent data that you can then validate against historical data. Alternatively, you can literally program in the generated test and grow neural nets where you can grow agents and produce realistic things. But you can never expose them to exactly the conditions that they did face and you can't guess what conditions they're going to face but you can create realistic and adaptive entities.

The correlation between function, environment, and structure gives us hope for inferring things. Adaptedness also gives us good reason to believe that the past will predict the future and that extrapolation is not a bad strategy. Even though we have all of these heterogeneous types, all of those types have to perform roughly the same function in roughly the same environment. We can assume homeostasis. We can know or infer the goals of systems. And we can also try these goal function driven models instead of detailed physical models when it makes sense which can capture simplifying constraints. There are things we can also do to mitigate some of our core problems. We can leverage our adaptedness, use optimization models to capture functionality and goals, leverage these means for an ends relationships, and use more effort to determine where our models are valid. We can give up on hard questions and focus on questions that our models can actually answer. We also can deal with some of these scaling issues by employing multiple models that feed each other.

10.6 Asmeret Naugle, Mark Smith, Tatiana Flanagan, Laura Swiler, Stephen Verzi, Curtis Johnson, Matt Lave, Kasimir Gabert, Jeff Tsao: Technical White Paper on "Complex Systems Models and Their Uses: Towards a New Science of Verification, Validation and Uncertainty Quantification"

Complex Systems Models and Their Uses: Towards a New Science of Verification, Validation, and Uncertainty Quantification

Asmeret Naugle, Mark Smith, Tatiana Flanagan, Laura Swiler, Stephen Verzi, Curtis Johnson, Matthew Lave, Kasimir Gabert, Jeff Tsao
Sandia National Laboratories, June 22-24, 2016

Purpose and Goals of the Workshop and this White Paper

This short white paper is a condensation of some of the thoughts that will be presented in the plenary talks at our workshop on "Complex Systems Models and Their Uses: Towards a New Science of Verification, Validation and Uncertainty Quantification." The intent of the workshop is to bring together two communities – the complex systems modeling community, and the VVUQ community – to propose and discuss research opportunities at their intersection. The intent of this white paper is to present one possible framing of the issues that we hope those research opportunities will address, as a means to stimulate thinking. We begin by reviewing current VVUQ approaches to physics-based models; then we describe some of the characteristics and uses of complex systems models that make current VVUQ approaches more difficult to apply; and finally we suggest the possibility of nonetheless building credibility in those models, an essential component of building user confidence in the knowledge gained from those models.
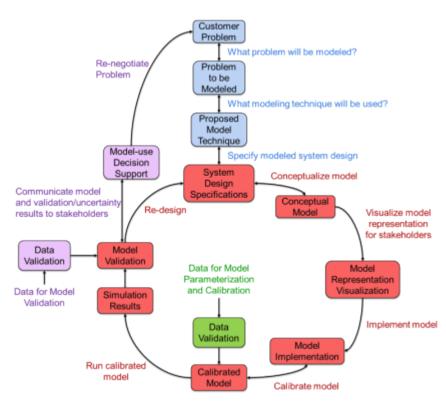
VVUQ for Physics-Based Models

Techniques for verification, validation, and uncertainty quantification (VVUQ) have been developed for a variety of fields, but much of the technical capability in this area has focused on improving the credibility of physics-based models in computational science and engineering. The goal of VVUQ in these fields may be to explicitly identify assumptions and limitations of a model, in order to understand credibility and ensure that the models are not oversold. While engineered systems may be relatively well-known compared to complex systems, prediction of outcomes is still very difficult. Data availability, high computational demand of simulations, imperfect models, under-resolved approximations, unknown parameters and boundary conditions, humans in the loop, and a desire to extrapolate beyond validation boundaries all contribute to difficulty in simulation, validation, and prediction. Nevertheless, VVUQ for physics-based models enables more rigorous evaluation of models, improving understanding of the limitations of these models and attempting to ensure credible outcomes. New development of VVUQ techniques for complex systems should leverage current VVUQ techniques for physics-based models wherever possible.

The computational science and engineering community has relatively well-established techniques and definitions that may benefit the development of VVUQ for complex systems. An overview of the modeling and VVUQ process, adapted from Balci[1], is shown below. **Verification** is often defined as addressing the question "Are the equations solved correctly?"[2]. This encompasses the formulation of the governing mathematical equations, the translation to a set of discretized equations that must be solved, the implementation of the equations in a computer code, and the use of algorithms and solvers to solve the equations. Verification involves software quality and associated topics such as design requirements, software change and revision control, the use of unit and regression tests covering intended features with core and edge cases, and memory analysis.

---

1. O. Balci. Verification, Validation, and Testing. In Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. John Wiley & Sons; 2007.

2. W. Oberkampf and C. Roy. Verification and Validation in Scientific Computing. Cambridge University Press; 1st ed., 2010.

Verification also involves numerical analysis to examine the properties the algorithm possesses, such as symmetry, stability, conservation, convergence, etc.

Validation addresses the question: "Is the model adequate to use for the intended application?" [2,3]. Validation involves a quantitative comparison between experimental data and computational simulation results. In the simplest sense, validation involves taking the difference between the experimental and simulation results, and assessing whether that difference is acceptable for the intended use. However, most validation approaches go beyond that. There are a number of validation metrics which have been developed to account for uncertainties and errors in both experimental data and simulations. Widely used validation comparison error metrics include the difference between simulation and physical data, defined by the ASME 20-2009 V&V Standard[3], as well as the uncertainty in this difference which includes parameter uncertainty, numerical discretization errors, and uncertainty in the physical measurement data. Other metrics such as Bayes factor and the area metric[4] are also used for validation.

VVUQ may also be supported by other capabilities. **Calibration** is the identification of optimal parameter values that yield simulation results which "best match" the experimental data in some sense. Calibration methods include linear and nonlinear regression approaches (e.g. find the parameters that minimize the sum of squared errors) and Bayesian calibration (find a posterior distribution on parameters that is informed both by the prior distribution and the experimental data). **Uncertainty Quantification** is the propagation of uncertain model inputs (usually characterized with probability distributions) to model outputs, resulting in a distribution on response quantities. Sampling methods are most commonly used to perform UQ. Based on the response distribution, one can assess the likelihood of typical or extreme outputs, determine the mean or median performance, understand the variability in the responses, and find probability of failure. **Sensitivity Analysis** is the identification of the most important variables affecting the response. It involves understanding how model outputs vary as inputs vary. Sensitivity analysis metrics are often correlation coefficients or variance-based indices.

While early work in the VVUQ area often emphasized the distinction between aleatory and epistemic uncertainty, recent work has focused more on uncertainties in model form and model selection. Bayesian methods have also become more popular. There is an increasing acknowledgement that hierarchical validation is not always appropriate, and that rigorous VVUQ can be very expensive, requiring substantial data and computational power. Data is, in fact, key in the VVUQ process, requiring experimental capabilities, benchmarks, and enough data for both calibration and validation. The emphasis on data has had substantial impact in the rigor of the VVUQ process. Some researchers in computational science and engineering fields have encouraged the idea that the hard-science concepts of validation should not be applied to systems for which uncertainty is so great that predictive models are inappropriate[5]. Others have suggested that evaluation of models should be considered far broader than current VVUQ bounds. McNamara[6] suggests that models are only inputs to forecasts made by

3. ASME V&V 20-2009 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: https://www.asme.org/products/codes-standards/v-v-20-2009-standard-verification-validation
4. Y. Liu, W. Chen, P. Arendt, and H. Huang. *Toward a better understanding of model validation metrics.* Journal of Mechanical Design, July 2011, Vol. 133. DOI: 10.1115/1.4004223.
5. H. Meir. Personal communication with Vicente Romero at Sandia, May 2016.
6. L. McNamara. *Why models don't forecast*. Sandia Technical Report SAND2010-5203C. Aug. 2010.

people, and Saltelli and Funtowicz[7] give a process for a 'sensitivity audit' that assesses credibility of the entire modeling process.

Ideally, VVUQ practitioners would ensure that models are sufficiently evaluated and that their capabilities are not oversold. Expert judgment and customer enthusiasm would not be considered sufficient, and instead data-based VVUQ techniques would be applied to evaluate models. In many systems of interest data availability is low, but data should still be a focal point of VVUQ. Specifically, VVUQ practitioners should understand and communicate how data was used to calibrate, evaluate, and validate the model, as well as what benchmark data are available. Sensitivity analysis, uncertainty quantification, and optimization should be used to understand model behavior, identify worst case scenarios, and assess the spread and likelihood of potential outcomes. Finally, risk management on the use of the model should be considered. Specifically, researchers should consider what risks are involved with using a model, and how the VVUQ process and outcomes might affect this risk.

Complex Systems Models

While most existing verification, validation, and uncertainty quantification (VVUQ) techniques were developed for studying engineered and physics-based systems, there has been recent interest in developing VVUQ methods for systems with characteristics that make them difficult to model and assess using current methods. These might include systems traditionally defined as complex, as well as other systems with characteristics dissimilar to those of physics-based models. For the purposes of this document and workshop, we will focus on systems and model-use pairs that have specific characteristics that make their modeling, assessment, and VVUQ substantially different from that of engineered and physics-based systems, and describe these using the term 'complex systems'.

Amenability to Current VVUQ Approaches

Systems that can be treated as a set of physics-based differential equations:	Systems that are more difficult because of complicating characteristics:	Systems that are even more difficult because of adaptive complexity:
• Have clearly defined quantities of interest • May have many variables and parameters, but only some are uncertain and data can be obtained	• Lack of established theory • Lack of data, high uncertainty • Non-standard 'knowledge of interest' • Feedback and nonlinearity • Etc.	• Goal-directed adaptations create non-random, hierarchically modular substructures

System and Model "Complexity"

Complex systems present particular difficulties for modeling and VVUQ. Applying existing rigorous quantitative assessment methods to these systems and models is sometimes possible, but in many cases a more qualitative approach may be necessary. These systems may present specific

---

7.   Saltelli and S. Funtowicz. *When all models are wrong*. Issues in Science and Technology, Winter 2014. Pp. 79--85.

characteristics that make them particularly difficult to assess with current VVUQ techniques and strategies. For example, there is often a lack of both theory and data available for defining and characterizing these systems, which may be compounded by inability to conduct experiments. In many cases, in fact, the behavior or scenario of interest has never occurred. Sometimes a perceived lack of data is actually caused by a low signal-to-noise ratio, with useful data nearly impossible to detect. Synergy and interaction between components make a reductionist approach to understanding these systems inappropriate. Non-linearity and feedback loops can add difficulty, and in some cases a model might even interact with and change the system of interest over time.

Models of these systems might also be used for different purposes than physics-based models. Prediction is often less-emphasized, with models instead used to answer questions about intervention effectiveness, system insight, risk, resilience, dynamic tendencies, or other behavior and knowledge of interest. Another challenge in modeling complex systems is determining what to include in the model and what to leave out. The importance of threshold effects in these systems might mean that small differences can have outsized effects. A wide variety of granularities can be used to model these systems, with benefits and drawbacks to any choice. These characteristics can make establishing model credibility difficult, since it is nearly impossible to establish that all important characteristics for the use case were included in the model. Similarly, any given model might focus on certain aspects of the system of interest, but no model can capture all of these aspects. Establishing model credibility would ideally involve ensuring

> **A few system/model characteristics that might make VVUQ difficult**
> - Lack of established theory
> - Lack of data or high signal-to-noise ratio
> - Adaptive behavior
> - High throughput
> - Heterogeneity of subcomponents
> - Multi-scale interaction
> - Bifurcations and phase change
> - Cascading behavior
> - Feedback loops
> - Non-linearity
> - Goal-driven and/or gaming behavior
> - Humans in the loop
> - Reliance on soft quantities
>
> **A few model uses that might make VVUQ difficult**
> - Qualitative questions
> - Need for real-time or quick turnaround
> - Feedback between model and system
> - Scenarios that have never occurred

that the model focuses on the same characteristics and functions as the real-world counterpart system, but ensuring and proving that the model focuses on the correct aspects is a major challenge for these models. Heterogeneity, latent information, and threshold effects might make mean-field approaches commonly used for physics-based systems inappropriate for application to complex systems.

In some cases, these systems will involve adaptive complexity, which we defined here to mean a system that evolves via a generate and test mechanism, a concept inspired by natural selection, and that we posit may be the universal engine for organized complexity. This model begins with a pool of *entrants* that go through a *test* process. Entrants experience varying success at

achieving their functional goals during the test. Test results are used to *generate* new entrants for the next test cycle such that the next generation is, at least on average, better suited to succeed under common test conditions. Along with ecological applications, generate and test processes can be found in any adaptive system, and can produce both behavioral (learning) and structural (evolutionary) changes. For example, the generate and test model applies to manufacturing (with products as the entrants and sales creating the test cycle), finance (with investments as the entrants and profits creating the test cycle), and machine learning (with candidate features as the entrants and tests against data creating the test cycle) applications. Generate and test processes create functionality in their associated populations, and ensure that structure, function, and environment are strongly linked. Adaptation within a variable environment can add to resilience and robustness. Generate and test cycles often appear in hierarchies, and can have many types of relationships with each other including: predatory, symbiotic, parasitic, and even enslavement relationships, in which a lower-level cycle relies on a higher-level cycle for survival, and the higher-level cycle is benefited by the lower-level.

Exemplar complex systems topics can help to illuminate potential uses of modeling and VVUQ techniques and issues. For example, the electric grid, while traditionally modeled as a physics-based system, can be considered adaptively complex for some modeling problems. At a basic level, electric grids consist of generators, transmission lines, distribution lines, and consumer loads, but these components are networked together into vast, complex machines that span sizable portions of entire continents. Widely distributed loads and generation much match across the system at all times but neither is known perfectly in advance, and control is shared among a large number of actors, each of whom is trying to maximize their individual interests, including efficiency, safety, reliability, and profit, through individual, behavioral 'generate and test' mechanisms.[8] An example of a use of interest to electric grid experts to which a complex systems model might be applied is: how should grid planning, design, and operation be managed to maximize grid resilience? Another exemplar topic of wide interest is cyber security. Cyber systems include computers and networks, as well as the software that dictates their purpose and function. People, including ordinary cyber users and system administrators as well as hackers, criminals, and terrorists, also play crucial roles in cyber systems and are ultimately the source of all cyber security concerns. An example of a use of interest to cyber security experts to which a complex systems model might be applied is: what are the best ways to block malicious internet activity, given that malicious agents are likely to adapt quickly to new strategies through 'generate and test' mechanisms leading to learning?

Approaches and Suggestions for VVUQ of Complex Systems Models

---

[8] National Academies of Sciences, Engineering, and Medicine. Mathematical Sciences Research Challenges for the Next-Generation Electric Grid: Summary of a Workshop. Washington, DC: The National Academies Press, 2015. doi:10.17226/21808, http://www.nap.edu/catalog/21808/mathematical-sciences-research-challenges-for-the-next-generation-electric-grid.
National Academies of Sciences, Engineering, and Medicine. Analytic Research Foundations for the Next-Generation Electric Grid. Washington, DC: The National Academies Press, 2016. doi:10.17226/21919, http://www.nap.edu/catalog/21919/analytic-research-foundations-for-the-next-generation-electric-grid.
Joseph H. Eto and Robert J. Thomas, eds. Computational Needs for Next Generation Electric Grid Proceedings. LBNL-5105E, 2011. https://certs.lbl.gov/sites/all/files/lbnl-5105e.pdf.

The strategies and techniques used for VVUQ of physics-based models will likely need to be adjusted for useful application to complex systems, especially since these systems often lack the theoretical, mathematical, and data rigor that many current VVUQ approaches were originally designed for. To address this issue, we suggest organizing VVUQ for complex systems under the umbrella of credibility. The overarching purpose of the VVUQ process is to assess and improve the credibility of models, allowing increased confidence in their use and in how well the models represent the real world. All of the strategies and techniques used for VVUQ of complex systems should be designed to improve credibility. These techniques are likely to include original or modified versions of current verification, validation, calibration, uncertainty quantification, and sensitivity analysis techniques, including those developed for physics-based models and others developed for complex systems[9,10]. Existing VVUQ techniques developed and/or used for complex systems applications span a wide range of strategies, including things like face validation, Turing tests, extreme condition tests, and behavior reproduction. New strategies, including both quantitative and qualitative techniques, should also be developed. A wide variety of techniques and strategies are likely to be useful, since the variety of questions, modeling paradigms, and systems of interest applicable to complex systems is likely to be very wide.

One of the first issues that must be faced when applying modeling and VVUQ to complex systems is explicit identification of the model-use pair. This involves two major factors: what is the intended use of the model, and what models are appropriate for this use? For example, is the model designed for prediction, or is it more appropriate for policy exploration? Is it a good model for conducting a risk analysis, or is it useful for demonstrating how a theoretical system might play out over time? It is important not only that the questions we ask of models are appropriate and well-defined, but also that model credibility is assessed specifically in relation to the questions and model uses of interest. This may require a deep understanding of the system to be modeled, as well as negotiation with the customer to ensure that expectations of project and model outcomes are reasonable. The model-use pair should help in identifying what level of credibility is possible, and might indicate specific techniques that should be used to assess and improve that credibility.

Assessing and improving the credibility of models of complex systems requires a balance of qualitative and quantitative strategies. Current VVUQ techniques should be used where appropriate, but it is important that these techniques not be applied in situations where they are inappropriate. For example, it may not be possible to assess a system that has no data available using current, physics-based validation logic and metrics. A model of such a system might still be assessed using a concept similar to validation, by comparing model results to analogous, synthetic, or expert data. The rigor of this technique would be different from the physics-based concept of validation, but could still improve the credibility of the model. In similar ways, many of the concepts used in physics-based VVUQ might be altered to more appropriately apply to models of complex systems.

Combinations of approaches might be particularly useful in studying complex systems, allowing researchers to consider multiple scales of behavior and/or use divergent methods. Understanding whether different approaches lead to similar answers, and potentially feeding models with information

8.   O. Balci. 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. Annals of Operations Research, 53 (1994): 121-173.

9.   J. D. Sterman. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin McGraw-Hill, Boston, Massachusetts. Pp. 858-889.

learned from differently-scaled models, could lead to better understanding of these systems. VVUQ of complex systems models should also consider the limits within which these models can be considered valid or credible.

Interestingly, the very qualities that make complex systems difficult to model might also present advantages to the modeler. For example, understanding that structure, function, and environment are highly correlated can allow information about two of these characteristics to lead to inference about the third. Heterogeneity is high in these systems, but is also limited by the functionality required by generate and test cycles. Dynamic stability can often be assumed. Finally, goals of the system can be inferred through knowledge about generate and test cycles, allowing the modeler to narrow down the detail required for a model.

In creating strategies and techniques for VVUQ of complex systems, we are attempting to add rigor to a process which is still being developed. As modeling techniques for complex systems evolve, associated VVUQ strategies and techniques should be developed to improve understanding of the credibility of these models and how they should be used. A balance must be found between rigorous, quantitative techniques and more qualitative, descriptive strategies. Finally, since complex systems encompass a huge variety of topics and types of models, VVUQ strategies must be flexible to ensure relevance to a variety of possible applications.

[Blank page following section.]

DISTRIBUTION

4   Lawrence Livermore National Laboratory
     Attn: N. Dunipace (1)
     P.O. Box 808, MS L-795
     Livermore, CA 94551-0808

1	David	Alderson	NPS
1	Daniel	Appelo	UNM
1	George	Barr	Consultant
1	Wei	Chen	Northwestern
1	Daniel	Delaurentis	Purdue
1	Dennis	Engi	Consultant
1	Alfred	Hubler	UIUC
1	Linas	Mockus	Purdue
1	Roshanak	Nateghi	Purdue
1	Bill	Oberkampf	WLO Consulting
1	Norman	Packard	UIUC
1	Shreyas	Sundaram	Purdue
1	Jessica Glicken	Turnley	Galisteo Consulting

1	MS0110	David	Womble	1220
1	MS0148	Paul	Yourick	0710
1	MS0344	Jim	Stewart	1441
1	MS0348	Duane	Lindner	8100
1	MS0351	Ben	Cook	1910
1	MS0351	Bill	Hart	1913
1	MS0351	Robert	Leland	1000
1	MS0351	Andy	McIlroy	1900
1	MS0359	Karla	Weaver	1911
1	MS0421	George	Backus	0159
1	MS0425	Elizabeth	Keller	0159
1	MS0440	Scott	Holswade	2200
1	MS0519	Laura	McNamara	5346
1	MS0620	Kasimir	Gabert	5638
1	MS0621	Todd	Jones	*5638
1	MS0724	James	Chavez	6000
1	MS0736	Richard	Griffith	6230

1	MS0747	Ace	Sorensen	6131
1	MS0789	Sondra	Spence	1931
1	MS0793	Johnathan	Cox	*4122
1	MS0801	David	White	9300
1	MS0815	Justine	Johannes	1500
1	MS0828	Vicente	Romero	1544
1	MS0828	Walt	Witkowski	1544
1	MS0830	Susan	Stevens-Adams	0431
1	MS0831	John	Zepper	5500
1	MS0836	Dean	Dobranich	1514
1	MS0933	Chris	Lamb	9526
1	MS0980	Steven	Gianoulakis	5550
1	MS0980	Amy	Shrouf	5554
1	MS1027	Curtis	Johnson	5635
1	MS1027	Christina	Ting	5635
1	MS1104	Carol	Adkins	6100
1	MS1104	Charlie	Hanley	6110
1	MS1137	Tatiana	Flanagan	6921
1	MS1137	Sasha	Outkin	6921
1	MS1137	Kevin	Stamber	6132
1	MS1138	Walt	Beyeler	6924
1	MS1138	Theresa	Brown	6924
1	MS1138	Thomas	Corbet	6924
1	MS1138	Steve	Kleban	6132
1	MS1138	Greg	Lambert	6132
1	MS1138	Lori	Parrott	6924
1	MS1138	Dan	Pless	6924
1	MS1138	Steve	Verzi	6132
1	MS1138	Eric	Vugrin	6921
1	MS1139	Nancy	Brodsky	6921
1	MS1169	Charles	Barbour	1300
1	MS1188	Nat	Brown	6131
1	MS1188	Tammy	Brown	6135
1	MS1188	Pat	Finley	6131
1	MS1188	Jared	Gearhart	6131
1	MS1188	Marcy	Hoover	6130
1	MS1188	Dean	Jones	6131
1	MS1188	Kat	Jones	6131
1	MS1188	Craig	Lawton	6135
1	MS1188	Nadine	Miner	6114
1	MS1188	Mark	Smith	6133
1	MS1188	Bruce	Thompson	6133
1	MS1243	John	Vonderheide	5520

1	MS1244	Stephen	Lott	5530
1	MS1318	Brian	Adams	1441
1	MS1318	Laura	Swiler	1441
1	MS1318	Tim	Trucano	1400
1	MS1318	Tim	Wildey	1441
1	MS1323	Bill	Rider	1446
1	MS1324	John	Feddema	1460
1	MS1324	Bruce	Hendrickson	1400
1	MS1326	John	Siirola	1464
1	MS1327	Phil	Bennett	1463
1	MS1327	Mike	Bernard	1463
1	MS1327	Asmeret	Naugle	1463
1	MS1327	David	Stracuzzi	1462
1	MS1327	Jennifer	Troup	1464
1	MS1327	Craig	Vineyard	1462
1	MS1327	John	Wagner	1462
1	MS1327	Christy	Warrender	1462
1	MS1348	John	Larson	4200
1	MS1415	Bill	Seng	1912
1	MS1421	Jeff	Tsao	1120
1	MS1427	Grant	Heffelfinger	1100
1	MS1484	Christopher	Frazer	10244
1	MS9001	Marianne	Walck	8000
1	MS9052	Matt	Lave	6112
1	MS9151	Heidi	Ammerlahn	8900
1	MS9151	Jim	Costa	8950
1	MS9151	Susanna	Gordon	8960
1	MS9152	Robert	Clay	8950
1	MS9158	Robert	Armstrong	*8956
1	MS9158	Jackson	Mayo	8953
1	MS9159	JD	Doak	8962
1	MS9159	Phil	Kegelmeyer	8900
1	MS9159	Karim	Mahrous	8962
1	MS9159	Jerry	McNeish	8954
1	MS9159	Cosmin	Safta	8954
1	MS9957	Amanda	Dodd	1914

[The housekeeping entries are required for all SAND reports.]

1	MS0899	Technical Library	9536 (electronic copy)

For LDRD reports, add:

1	MS0359	D. Chavez, LDRD Office	1911

For CRADA reports add:

1   MS0115   OFA/NFE Agreements   10012

For Patent Caution reports, add:

1   MS0161   Legal Technology Transfer Center   11500

Sandia National Laboratories