SAND 2011-3422C

Integrating Uncertainty Analysis into Complex-System Modeling for Effective Public Policy I: Preliminary Findings

Patrick D. Finley, Robert J. Glass, Thomas W Moore, Arlo L Ames, Leland Evans, Daniel C. Cannon, Jacob Hobbs
Sandia National Laboratories
Albuquerque, New Mexico, USA
Complex Adaptive Systems of Systems (CASoS) Engineering Initiative
pdfinle@sandia.gov

Victoria J. Davey
Office of Public Health and Environmental Hazards
Veterans Health Administration
Department of Veterans Affairs
Washington, DC, USA

Public policy seeks to influence complex natural, social, and engineered systems to achieve desired outcomes. Effective public policies are those which combine good outcomes with high reliability such that their choice is robust to a wide range of possible uncertainties. Modeling these complex systems and their potential response to proposed policies can provide decision-makers with an objective basis for policy design. Critical to this design process is the development of rigorous methods to evaluate and rank modeled policy effectiveness in context of model uncertainty.

A peer-reviewed complex system model of pandemic influenza propagation is used in a test case to illustrate the power of uncertainty-based public policy ranking. The networked-agent model calculates the effects of both social network-based community mitigation practices such as school closure and social distancing, and individually-based treatment options such as antiviral treatment and vaccination. A proposed uncertainty-based methodology is described. The roles of experimental design, input factor representation, and sensitivity analysis methodology illustrate a succinct methodology to rank pandemic disease control options. A combination of space-filling experimental designs, modeling policy options as continuous rather than categorical variables, and treed Gaussian process and polynomial chaos expansion-based sensitivity analysis are projected to yield a straightforward ranking of policy

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.

options that is robust to the identified aleatory and epistemic model uncertainties. These methods should have the additional advantage of requiring relatively few model runs to achieve a consistent and defensible ranking. Phase I of the experimental work to demonstrate the proposed uncertainty-based methodology is outlined. This early work indicates that analysis of published modeling results can generate a robust composite ranking of public policy options for mitigation of pandemic influenza.

1 Introduction

Many domains of public policy are well represented as Complex Adaptive Systems of Systems (CASoS), including social organizations, economies, and governmental bodies. Modeling of public policy options and tradeoffs as Complex Adaptive Systems (CASs) has shown promise recently in understanding the structure and dynamics of these systems with the goal of improving the quality and reliability of public policy decisions. Often modeling of CASs results in some increased understanding of the problem space, but little in the way of objective guidance for decision makers on how best to formulate policies and regulations to guide the systems in socially beneficial directions.

Computational models of CASoS permit exploration of a wide range of possible policy options. High Performance Computing (HPC) enables models to be run many thousands of times with different input values. Analysis of the model outputs generated from these suites of runs can provide guidance on how systems behave under different policy regimes. Vast amounts of data generated by these supercomputer deployments can be challenging to interpret. However, automated data reduction eventually provides estimates of possible system responses to policy changes.

Simple evaluation of CASoS models across a range of inputs provides much needed information to decision makers, but ignores one of their major concerns. Policy decisions not only consider the projected end point of the policy implementation, but also must consider risk. Few policy makers would choose a prospective policy which promises very good quantitative outcomes, but is likely, if it should fail, to fail with catastrophic consequences. CASoS-based social, technical, and governmental systems and models used for their analysis are fraught with massive uncertainties. An ideal policy choice is one which performs well under a wide range of uncertainties. Such a policy would be expected to have positive outcomes, but perhaps more importantly, to also have a low risk of catastrophic failure.

Methods from the field of uncertainty quantification can provide the information needed to classify potential public policy options based on both their outcomes and their risk of failure. These powerful techniques enable researchers to identify, categorize and manage model and system uncertainty from many sources. Such uncertainty estimates permit ranking of policies not only by objective quantitative performance metrics, but also by their robustness to unforeseen events. While these methods do not yield verifiable predictions of policy outcomes, the combination of modeling-derived metrics, and uncertainty-derived reliability estimates permits methodical ranking of policy options.

2 Overview of Uncertainty Analysis

Uncertainty analysis of model results provides quantitative estimates of the quality of computed model outputs. Sensitivity analysis identifies which model outputs are most responsive to uncertain inputs. Factor sensitivities combine with knowledge of model structure and dynamics to generate realistic estimates of total uncertainty in model outputs.

2.1 Uncertainty

Uncertainty is simply a lack of certainty about a process, a quantity, or the state of a system. Uncertainty can apply to past and present events, but is most often ascribed to unknowable future events. While uncertainty is often interpreted as an undesirable property about a system, uncertainty also admits possibilities for advantageous events and innovations. Risk is a term that is often applied to uncertain situations or conditions; here we consider risk to represent undesirable states of uncertainty with implicit or explicit potential for significant loss.

Computer models exhibit a range of uncertainty types [Helton et al. 2007]. Lack of knowledge about true values for input parameters used for modeling runs is termed parametric uncertainty. Since research or expert opinion can often narrow down the certainty with which input parameters are known, parametric uncertainty is often termed reducible uncertainty. Stochastic uncertainty results from an inability to forecast future events, such as wind direction next Tuesday, which may be implicit within a computer model. Structural uncertainty is a measure of how close a model captures the system of interest. Models constructed with incorrect mathematical formulations or insufficient detail may be poor analogs of real systems and thus have high structural uncertainty. An alternative categorization for uncertainty terms both structural uncertainty and parametric uncertainty as epistemic uncertainty, since they both arise from a lack of knowledge about values and processes. Stochastic uncertainty, on the other hand, captures fundamental inability to specify future events and is often termed aleatory uncertainty.

2.2 Design of Experiments

Using models to evaluate public policy options requires that they be run in a methodical manner that is focused on the issues at hand. Design of Experiments (DOE) entails planning model runs and parameter variations to answer a question adequately and efficiently [Santner et al. 2010]. DOE has long been a staple of bench research; a wide variety of methods and techniques have been developed to address cumulative measurement errors and other complications affecting hands-on experimentation. Computer experiments differ from bench experiments in that models can be configured to run deterministically and produce the exact same output for a given input without the variability seen in traditional experiments. Thus, many of the experimental designs developed for bench experiments are not appropriate for computer models. Rather than exposing measurement variation within experiments, designs for computational experiments focus on efficiently covering the multidimensional space of parameters adequately. Space filling designs ensure that parameter space is adequately sampled, and is the design of choice for uncertainty analysis using computer models.

2.3 Sensitivity Analysis

Sensitivity Analysis (SA) apportions variability in the output of a computer model to uncertainties in the constituent model inputs [Saltelli et al 2008]. As such, SA is a foundational method for understanding and measuring parametric uncertainty. Sensitivity analysis generates metrics termed *Sensitivity Indices* for each parametric input for a model which represents the impact that that parameter has on the value of the model output. These sensitivity indices are valuable for determining how best to decrease uncertainty in a model. Parameters with small sensitivity indices have very small effect on the output values; thus, effort to reduce uncertainty associated with those parameters may not translate to substantially decreased output uncertainty. Efforts to measure or generate more certain parameter estimates would be better spent on inputs having large sensitivity indices. Refining these parameter values would be expected to greatly improve overall model uncertainty.

Sensitivity Analyses fall into two broad categories, univariate and multivariate. Univariate methods are more intuitive to execute and interpret, whereas multivariate methods provide more information.

Univariate SA methods involve running the computer model many times with all inputs but one staying fixed. The effect of the single varying input parameter on the final output value is then plotted in a simple scatterplot. This method provides a wealth of information on general trends within models due to individual inputs. However, non-additive effects from combinations of inputs cannot be resolved with these simple tools.

Multivariate SA techniques methodically vary all inputs for each computer runs and rely upon sophisticated mathematical procedures to compute sensitivity indices. Techniques such as the Sobol method can generate accurate estimates the relative contributions of each input, but can require enormous numbers of model runs to generate the needed information. In contrast meta-models or surrogate models use fewer runs to generate a representative response surface for the model. The response surface is in turn used for detailed sensitivity analysis calculations [Storlie et al. 2009].

2.4 Uncertainty Quantification

Determining the sources of uncertainty and tracing effects of uncertainty throughout a model is termed uncertainty quantification (UQ) [Helton et al. 2007]. Each uncertain parameter is investigated to determine which distribution most accurately captures the possible values for the input. Similarly, each source of stochastic uncertainty is rigorously identified and its effects determined. The combined effects of these carefully estimated uncertainties are traced through to determine their possible effects on output uncertainty. Combined nested loop Monte Carlo model execution protocols permit concerted calculation of combined effects of epistemic and aleatory uncertainties. Results of these analyses are traditionally presented in horse-tail plots wherein individual time histories of many runs are plotted on a common set of axes.

2.4 Decision Analysis

Once the potential effects of uncertainty are measured and apportioned by various SA and UQ methods, policy options represented as parameter settings are analyzed to find the most advantageous combinations. Often simple visual inspection of model performance verses variability allows the most robust and effective policy to be immediately determined. For more involved models, optimization and search methods can be used o find the options which best combine objective performance and low susceptibility to unforeseen circumstances.

3 Applying the Methods

We re-examine the published findings of a peer-reviewed complex system model to demonstrate the potential applicability of uncertainty analysis on public policy decision making. Davey et al. [Davey 2008] used an agent based model of disease propagation through a stylized community to evaluate prevention and mitigation of pandemic influenza. This investigation investigated the effects of a range mitigation strategies or policies on the severity and duration of epidemic outbreaks. Many different parametric combinations were fed into the disease propagation model to determine the interventions which were most likely to reduce illness, death and economic cost.

This study will examine the data which was generated by the Davey investigation to determine whether additional uncertainty-based methods and analyses could refine the published findings. The present study is organized into four distinct sequential phases:

1. Applying additional analyses to the data generated by Davey et al.
2. Running the Davey model using different experimental designs
3. Modifying the Davey model to make it more amenable to advanced analyses
4. Applying quantitative decision analytical methods to model data to rigorously rank policy options.

This paper presents some initial results from Phase 1 of the extended project.

3.1 Model Description

Loki-Infect is a networked agent-based computational model developed by the National Infrastructure Simulation and Analysis Center (NISAC) at Sandia National Laboratories. In this model, agents represent individuals of various age classes who are linked to each other within and among social groups (such as households, neighborhoods, school classes, clubs, businesses, etc.) to form an explicit contact network reflective of a multiply-overlapping, structured community. Behavioral rules for individuals, their interactions, and the performance of network links are specified to model the spread of influenza. Community mitigation strategies are implemented through modifications of these behavioral rules when a given strategy is imposed during a simulation. Intervention strategies are listed in Table 1

Table 1: Intervention Strategies

Category	Symbol	Intervention
Network-Based	S	Schools closed.
Network-Based	C	Social distancing of children and teenagers.
Network-Based	A	Social distancing of adults and seniors.
Case-Based	Q	Household quarantine.
Case-Based	T	Antiviral treatment.
Case-Based	P	Household member antiviral prophylaxis.
Case-Based	E	Extended contact prophylaxis

Davey et al. ran the model for a wide range of compliance values, disease infectivity values, mitigation initiation and cessation times for combinations of intervention strategies. The model was run 100 times for each set of distinct input parameters to explore effectiveness of disease containment across different randomly generated community social networks. The model was run over 2,000,000 times to fully cover the parameter space.

Model results were presented in a series of crosstab tables which showed the mean number of infected people for each combination of intervention strategies. A small portion of a crosstab from the original study is reproduced as Table 2.

Table 2: Example model run results. Cell values represent mean number of infected people from 100 runs. Infectivity=0.75, Compliance=90%

	None	A	C	C,A	S	S, A	S,C	S,C,A
None	2780	1872	1111	624	221	207	124	119
T	1560	765	373	241	164	150	122	117
Q	984	562	267	237	178	151	125	141
P	711	379	217	184	161	138	114	123
Q,T	600	324	218	159	140	132	119	120
Q,P	329	298	166	160	148	129	121	130
E	251	208	149	150	146	134	106	111
Q,E	267	187	138	145	122	117	104	108

2.2 Data Analysis

Standard statistical techniques can extract useful information from the data set shown in Table 1. Davey et al. looked at a range of variables affecting intervention implementation that are not represented in the simplified data subset shown in Table 2. However, this subset serves well to illustrate potentially useful methods.

Mean counts of infected people shown in the cells of Table 2 clearly are greatest in the upper-left corner and decrease regularly to a minimum in the lower left corner. This corresponds to large numbers of infected individuals in cells representing few or no interventions grading to fewer infected persons in cells representing layered strategies of many interventions applied in concert. However, it is not clear from Table 2 exactly which interventions are demonstrably superior to others. Recall that the cell values in Table 2 are mean values from 100 individual model runs.

Analysis of Variance (ANOVA) of data matching corresponding to the values shown in Table 2 indicates which interventions can be clearly differentiated from others based on their mean efficacy at preventing influenza infection. Table 3 summarizes ANOVA results from a synthetic data set matching the mean values shown in Table 2 and corresponding standard deviation values quoted in the by Davey et al. Table 3 shows that the data represent six groups of significantly different mean values. Five distinct groups exist in the upper left of the table indicating treatments which result in mean infected counts of greater than 500, while all interventions resulting in less than 500 mean infected represent a single group which cannon be resolved into smaller sub-groups at 95% confidence. These findings may be overly optimistic; ANOVA only strictly applies when the distributions of the individual means is approximately Gaussian, which is unlikely in this case.

Table 3: ANOVA Results. Cell colors indicates clusters of mean values which are significantly different from others at 95% confidence

	None	A	C	C,A	S	S, A	S,C	S,C,A
None	2780	1872	1111	624	221	207	124	119
T	1560	765	373	241	164	150	122	117
Q	984	562	267	237	178	151	125	141
P	711	379	217	184	161	138	114	123
Q,T	600	324	218	159	140	132	119	120
Q,P	329	298	166	160	148	129	121	130
E	251	208	149	150	146	134	106	111
Q,E	267	187	138	145	122	117	104	108

ANOVA is unable to differentiate between similar-valued cells in the strategy matrix because the cells vary not only in mean value, but also in variability among the 100 individual values that the mean cell values represent. Figure 1 displays a scatter plot of the synthetic data set constructed from the data specifications in the Davey et al. paper. The figure shows both the variation in mean values for different protocols (combinations of interventions) and substantial difference in the reproducibility of the model runs.

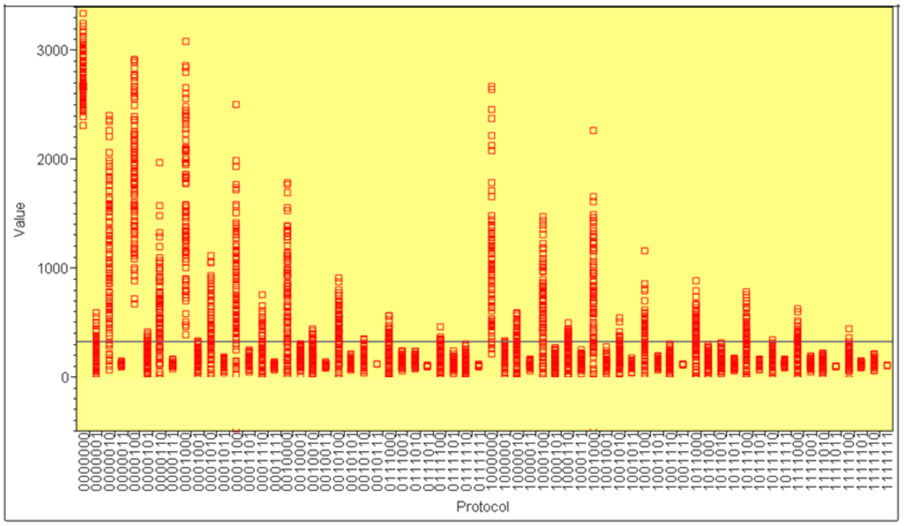

Figure 1: Scatter plot of data shown in Table 1. Vertical axis is number of infected individuals per run. Horizontal axis shows different intervention combinations. Red dots represent results from individual model runs

Model results can be better understood by considering both values and variability simultaneously. Figure 2 shows how the data represented in Figure 1 maps to such a display. Figure 3 shows how a simple scatter plot of variability as a function of mean infected count can yield a categorization of interventions into efficacy and reproducibility. Mean infected person count and standard deviation values from Figure 2 are plotted in the left pane of Figure 3. Note that model results define a broad arc of values in the Low Infectivity/Low Variability category through the Low Infectivity/High Variability category to the High Infectivity/High Variability category. Those options falling in the lower left region are those which would be most attractive to policy makers. They combine good outcomes and show little sensitivity to variability between model runs.

Consideration of the run results from the perspective of a public policy decision maker suggests that economic and reliability considerations might also come into play. Policy makers would probably prefer solutions which required fewer separate interventions to achieve a desirable outcome compared to those requiring many layered interventions. Solutions based on fewer interventions would be expected to cost less and be easier to coordinate and deploy in the field. Figure 4 illustrates that both average infected values and variability show strong correlation with number of interventions included in the strategy. Model runs with a larger number of layered interventions performed much better generating fewer infected individuals and less variation. These data suggest that selection of an effective and robust policy involves tradeoffs among conflicting characteristics of cost/complexity, effectiveness, and reliability.

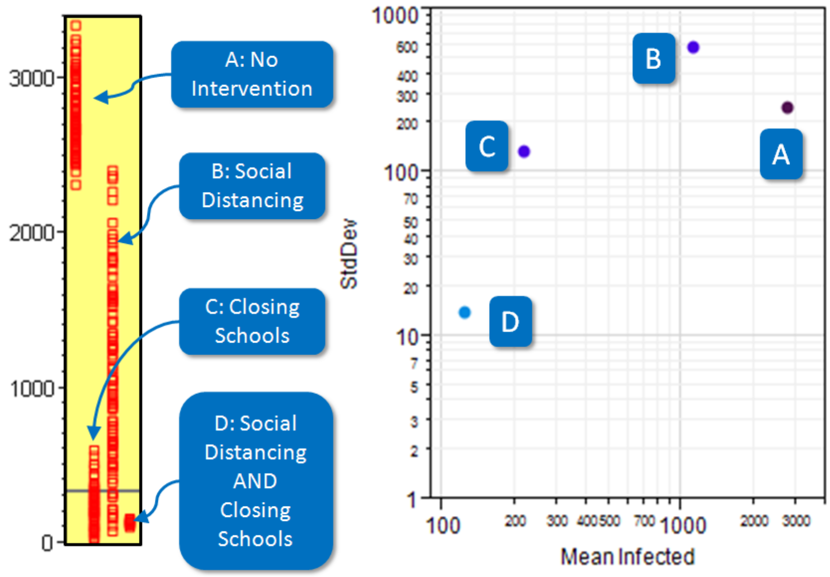

Figure 2: Mapping of model results to scatter plot

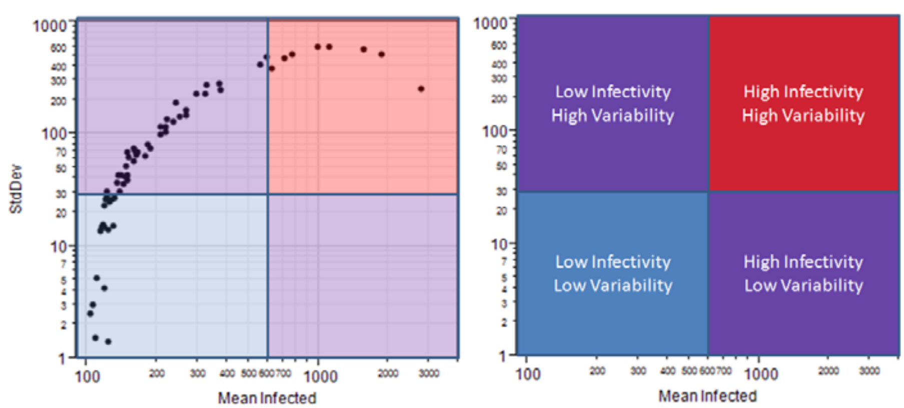

Figure 2: Conceptual categorization of scatter plots. Model results falling in lower left quadrant represent effective, robust choices.

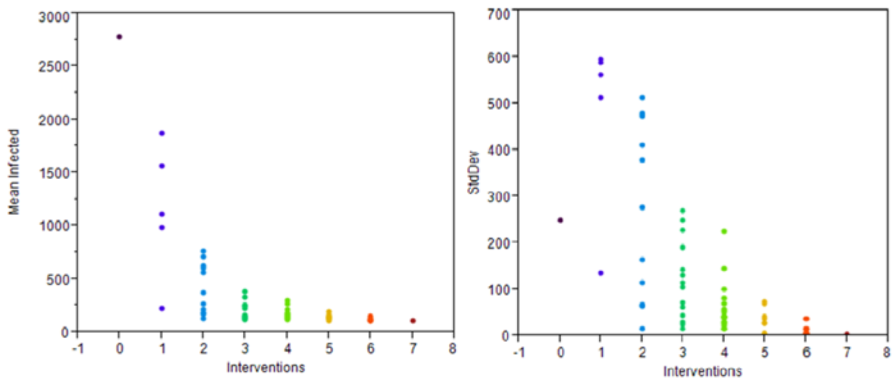

Figure 4. Strategy effectiveness and variability as a function of number of interventions

An alternate view of the relationships diagrammed in Figure 4 can be seen in the scatterplot in Figure 5. This plot demonstrates that best performing low variability intervention strategies found in the lower left region of the graph are characterized by larger numbers of layered interventions. Options which would be simpler to field and presumably cost less are arrayed in the upper right portion of the graph, indicating that these options allow more infections and more potentially risky variability.

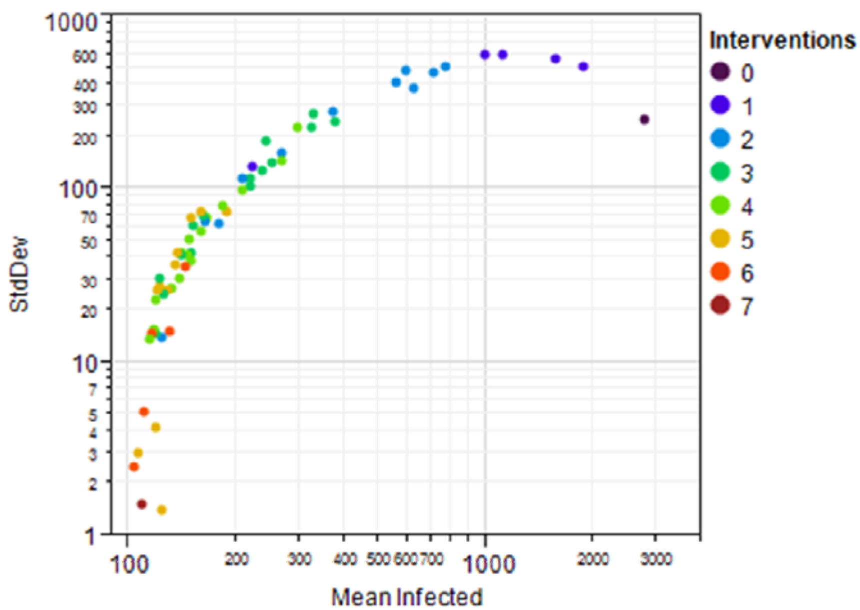

Figure 5: Dependence of intervention effectiveness on number of layered interventions

One approach to resolving the conflicting aims of public policy decision makers in evaluating intervention effectiveness and efficiency is to look for the best performing mix of interventions for each cost category. Figure 5 shows that within each category of intervention counts, intervention combinations possess differing effectiveness and robustness to variation. These plots show that little improvement occurs in opting for policies relying on four interventions relative to those requiring only three. Similarly, the best performing intervention strategy for the two intervention case is about as effective as the best performing four intervention strategy. This figure suggests that a ranking of effectiveness by intervention count class may be useful to allow decision makers to effectively gauge "bang for the buck" of different intervention approaches (Table 4).

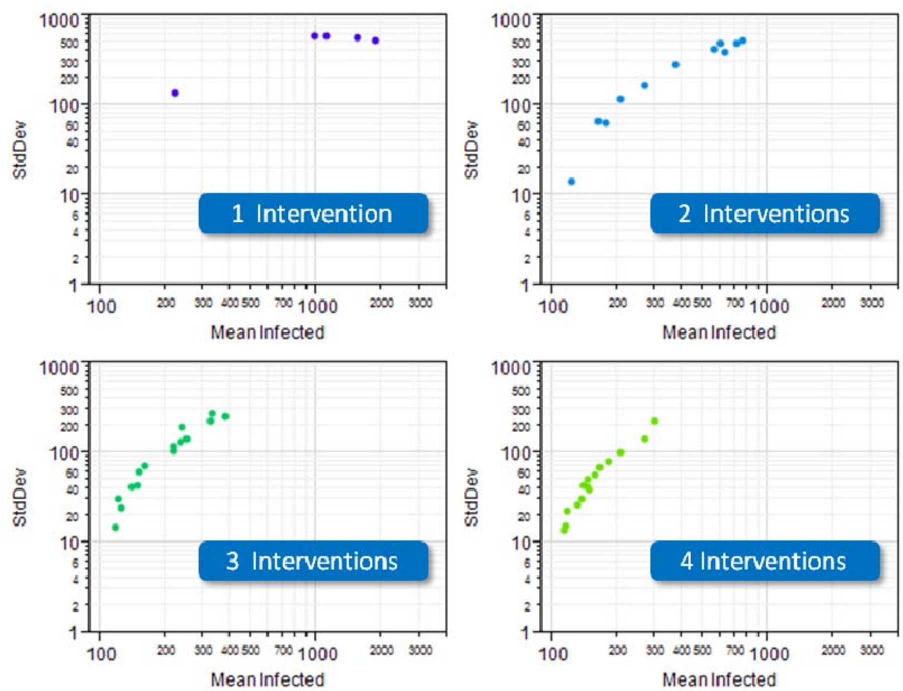

Figure 6: Performance of interventions classified by number of layered interventions

Table 4 reveals some not only the best performing suites of interventions for influenza mitigation, but also which specific interventions are most effective. All of the listed best-performing composite intervention strategies contain the school closure option (S). Child and teen social distancing (C) is the next most common component of the best-performing mitigation strategies. Of the case-based interventions, quarantine (Q) and antiviral treatment (T) appear to be effective in strategies reliant on few interventions, whereas prophylactic interventions (P and E) appear to work well only when applied in conjunction with many other interventions.

Table 4: Intervention Policy Rankings

Number of Interventions	Rank of Effectiveness	Intervention	Mean Infected	Standard Deviation
1	1	______S	220.5273	133.8141
	2	___Q___	984.4853	595.5609
	3	_____C_	1111.12	586.9511
2	1	_____CS	124.1364	13.89174
	2	T_____S	163.9688	65.28918
	3	___Q__S	178.0909	62.91812
3	1	____ACS	118.6429	14.75235
	2	___Q_CS	125.1429	24.43471
	3	T____CS	121.875	30.07342
4	1	TP___CS	114	13.6504
	2	T___ACS	116.75	15.32272
	3	T__Q_CS	118.5714	22.619
5	1	TP__ACS	123	1.414214
	2	TPE__CS	106	3
	3	T__QACS	119.6667	4.163332
6	1	TPEQ_CS	103.6	2.50998
	2	TPE_ACS	110.75	5.188127
	3	TPEQA_S	116.5	14.47165
7	1	TPEQACS	108.3333	1.527525

4 Conclusions

This report presents the initial findings of an extended effort to apply more sophisticated data analyses methods to complex adaptive systems. These advanced methods and techniques exploit measurements of uncertainty to extract more information from suites of model runs than simple summary statistics can provide.

This initial phase applied standard statistical approaches to published data from a peer-reviewed influenza propagation model. The exercise demonstrated that mean values from multiple model runs often do not consistently differentiate among input values representing the question under investigation. Simple Analysis of Variance tests showed that subtle differences between mean values for different modeled intervention configurations although evident in tabulated data are not statistically significant.

Incorporation of an additional factor into the analysis enabled a more detailed analysis. The number of interventions which must be applied to achieve a desired level of response is a factor which is likely to be of interest to public policy decision makers, since it directly relates to implementation cost and complexity. Separating the 64 modeled policy combinations into groups based on intervention count provided a simple ranking of policy options showing the most effective and low-risk policy options for different cost and complexity categories.

Phase 2 work underway now addresses the integration of rigorous uncertainty quantification to complex adaptive system models to enable robust decision making for public health policy questions.

Bibliography

[1] Davey, V.J., Glass, R.J., Min, H.J., Beyeler, W.E., Glass, L.M. (2008) Effective, Robust Design of Community Mitigation for Pandemic Influenza: A Systematic Examination of Proposed US Guidance. PLoS ONE 3(7): e2606. oi:10.1371/journal.pone.0002606

[2] Helton, J.C., Johnson, J.D., Oberkampf, W.L., Storlie, C.B.,(2007) A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. Computer Methods in Applied Mechanics and Engineering Volume 196, Issues 37-40, 1 August 2007, Pages 3980-3998

[3] Saltelli, A., Ratto, M., Andres, T., Campolongo, J., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. (2008) Global Sensitivity Analysis, Wiley.

[4] Santner, T.J., Williams, B.J., Notz, W.I. (2010) The Design and Analysis of Computer Experiments. Springer. .

[5] Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C.J. (2007) Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, Reliability Engineering & System Safety, Volume 94, Issue 11, November 2009, Pages 1735-1763, ISSN 0951-8320,