
WIZER: Automated Model Improvement in Multi-Agent Social-Network Systems

Alex Yahja¹ and Dr. Kathleen M. Carley²

¹ Computation, Organizations, and Society Program, Institute for Software Research International, Center for the Analysis of Social and Organizational Systems, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 `ay@cmu.edu`

² Institute for Software Research International, Center for the Analysis of Social and Organizational Systems, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 `kathleen.carley@cmu.edu`

Summary. There has been a significant increase in the use of multi-agent social-network models due to their ability to flexibly model emergent behaviors in complex socio-technical systems while linking to real data. These models are growing in size and complexity which requires significant time and effort to calibrate, validate, improve the model, and gain insight into model behavior. In this paper, we present our knowledge-based simulation-aided approach for automating model-improvement and our tool implementing this approach (WIZER). WIZER is capable of calibrating and validating multi-agent social-network models, and facilitates model-improvement and understanding. By employing knowledge-based search, causal analysis, and simulation control and inference techniques, WIZER can reduce the number of simulation runs needed to calibrate, validate, and improve a model and improve the focus of these runs. WIZER automates reasoning and analysis of simulations, instead of being a multi-agent programming language or environment. We ran a preliminary version of WIZER on BioWar – a city-scale social agent network simulation of the effects of weaponized biological attacks on a demographically-realistic population within a background of naturally-occurring diseases. The results demonstrate the efficacy of WIZER.

1 Introduction

Currently, a paradigm shift is occurring in how we model and think about knowledge, individuals, teams, groups, networks, organizations, markets, institutions, and other societal systems due to developments in the field of computational modeling and analysis [2][6][17][33][40][20][41]. Computational modeling and analysis has emerged as a useful scientific tool for addressing socio-technical problems with complex dynamic inter-related parts, such as natural disaster response and biological attacks. These problems do not occur in vacuum but within a context constrained by social, organizational,

geographical, technological, regulatory, cultural, and financial factors. As opportunities and challenges emerge dynamically in, say tsunami relief, existing rescue and aid plans often need major adaptations. For members of a rescue and aid team to cohesively follow a joint course action, it helps if the development in the environment, the consequent change of plans, and the effects of the intervention carried out according to the plans can be thought over and analyzed both in advance and in real/ongoing time.

There has been a rapid increase in the use of multi-agent models [12][27][29] – as well as social network analysis [42] – to address complex socio-technical problems. Model assessment – determining how valid, how explainable, and how robust a model is – is becoming a major concern [11]. Indeed, identifying reliable validation methods for complex systems such as electronic medical surveillance systems is a critical research area [35]. Calibration and validation serve as a foundation for model improvement through simulation and inference.

Models contain both explicit and implicit assumptions about some portion of the real world. These assumptions form abstractions of reality and these abstractions may or may not be sound. Moreover, the real world changes continuously and in unexpected ways. A cohesive joint course action by a group(s) responding to ongoing socio-technical problems is crucial to the efficiency and success of the action. How to adapt existing plans in ongoing socio-technical environments and how to coordinate members of a group(s) depend on how valid the underlying models and assumptions are. It is also desirable to automate improvement of models and assumptions based on live empirical data. Validation and model-improvement serve as a foundation for the coordination of large number of agents and their distributed tasks, goals, and organizations to deal with live socio-technical problems. The required fidelity of the model varies as a function of the research, policy, and/or mission questions being asked. Calibration, validation, and model-improvement are hard due to the changes in the real world, altered goals, inherent assumptions and abstractions, and human cognitive limitations such as bounded rationality [39].

Information exploitation is a technique that has yet to be fully employed to deal with the problem of calibration, validation, and model improvement. (The term "validation" will be used from now on to denote calibration, validation, and model-improvement.) Few multi-agent simulations have exploited the depth and breadth of available knowledge and information for validation that resides in journals, books, websites, human experts, and other sources. Typically, simulation results are designed solely for human analysis and validation is provided by subject matter experts announcing that the model "feels right" face validity. While this may be sufficient for small-scale simulations, it is woefully inadequate for large-scale simulations designed to inform decision-makers. In particular, automated help for validation and analysis is crucial. However, little work to date probes the important aspect of automating validation and analysis (this is conventionally left to humans to perform: there is an invisible wall of separation between simulation and analysis/knowledge inference). To successfully automate validation and analysis, domain knowl-

edge must be exploited, for example by an expert systems inference engine. A simulation and inference engine that can do virtual experiments and knowledge inference in concert would facilitate focused search by using both the simulation engines search space and the inference engines knowledge space to arrive at better parameter and meta-model values for validation. This paper describes our approach for doing knowledge-based simulation-aided validation in multi-agent social-network systems, embodied in a tool called WIZER (What-If AnalyZER). WIZER applies knowledge control of the simulation, inference and intelligent search in multi-agent social-network simulations.

The results presented in this paper are based on WIZER runs using BioWar. BioWar is a city-scale multi-agent social-network simulator capable of modeling the effects of weaponized biological attacks on a demographically-realistic population within a background of naturally-occurring diseases [7][8]. BioWar currently runs a few thousand to several million agents. Unlike traditional models that look at hypothetical cities (such as the Brookings smallpox model [18] and the SARS model [23]), BioWar is configured to represent real cities by loading census data, school district boundaries, etc. It models both healthy and infected agents as they go about their lives, enabling observation of absenteeism, drug purchases, hospital visits, and other data streams of interest.

2 Validation Experience

The complexity of ensuring valid results of agent-based simulations is shown during the validation of BioWar outputs. BioWar has many input and model parameters and these parameters can be stochastic. Brute-force search in the space of input and model parameters to fit the non-computational data is all but impossible. BioWar also has a complex response surface(s) and is knowledge intensive. Putting BioWar in specification can be viewed as a multi-dimensional numeric and symbolic optimization problem, with the knowledge component (e.g., school district announcements). The validation experience shows that there is a need for:

- Sophisticated analysis and response techniques to optimize the space over which parameters must be varied for correctness, and thus increase the number of parameters which can be studied.
- Tools to semi-automatically create and execute parametric studies to minimize the manual intervention currently required for these studies.
- New approaches to simulation scaling so as to reduce the size of the simulations which produce validated output streams.

WIZER addresses the first two points above.

3 Related Work

Multi-agent systems are usually "validated" by strictly applying requirements engineering. In software engineering terms [32], validation means the determination of the correctness of the final program or software produced with respect to the user needs and requirements not necessarily the empirical data or the real world. The usual emphasis in multi-agent system development is on language, programming, and design principles such as agent autonomy, team work, roles/types, and interaction protocols [12][29]. Formal methods [14] used in software engineering for control and understanding of complex multi-agent systems lack an effective means of determining if a program fulfills a given formal specification [16]. Complex societal problems contain "messy" interactions, dynamic processes, and emergent behaviors, so it is often problematic to apply requirements engineering and/or formal methods.

Another validation method is evolutionary verification and validation or EVV [37][38], which utilizes evolutionary algorithms, including genetic algorithms and scatter search, for verification and validation. While EVV allows testing and exploitation of unusual combinations of parameter values via evolutionary processes, it employs knowledge-poor genetic and evolutionary operators, not the scientific method, for doing experiments, forming and testing hypotheses, refining models, and inference, precluding non-evolutionary solutions.

Docking is another approach to validating multi-agent systems [1]. Docking is based on the notion of repeating a scientific experiment to confirm findings or to ensure accuracy. It considers whether two or more different simulation models align (produce similar results), which is used in turn as a basis to determine if one model can subsume another. The higher the degree of alignment among models, the more they can be assumed to be valid, especially if one (or both) of them has been previously validated. The challenges in applying docking are the limited number of previously validated models, the implicit and diverse assumptions incorporated into models and the differences in data and domains among models.

One application of docking is to align complex multi-agent simulations against mathematical or system dynamics models. BioWar's anthrax simulation has been successfully docked against the IPF (Incubation-Prodromal-Fulminant) mathematical model, a variant for anthrax of the well-known SIR (Susceptible-Infected-Recovered) epidemiological model [9] and BioWar's smallpox model has been docked against a SIR model of smallpox [10]. While aligning a multi-agent model with a widely used mathematical model can show the differences and similarities between these two models, the validity is limited by the type of data the mathematical model uses. For example, the IPF model mentioned above operates on population-level data, so the result of the alignment will be only valid at the granularity of population-level data. Mathematical models also have difficulties representing non-numerical (symbolic)

knowledge, including the knowledge base underlying complex context-sensitive agent interactions.

Validating complex multi-agent simulations by statistical methods alone [24] is problematic due to the coarse granularity required for statistical methods to operate properly and the insufficient representation of symbolic knowledge. Statistical methods are good at describing data and inferring distributional parameters from samples, but statistic methods alone are insufficient to handle the highly dynamic, symbolic, causal, heterogeneous, and emergent nature of societal systems.

Complex multi-agent simulations are not normally validated using expert systems (such as OrgCon [5]) as it is thought that it is sufficient to let human experts alone perform the analyses, experiment design, and quantitative and symbolic reasoning. This view is especially prevalent as most simulations are in the realm of purely numeric models.

Human experts can do validation by focusing on the most relevant part of the system and thinking about the problem intuitively and creatively. These subject matter experts (SMEs) have the knowledge needed to judge model performance in their specialized fields. Applying learned expertise and intuition, SMEs can exploit hunches and insights, form rules, judge patterns, and analyze policies. Managed and administered properly, SMEs can be effective. Pitfalls include bounded rationality, implicit biases, implicit reasoning steps, judgment errors, and others.

Another approach to validation is direct comparison with real world data and knowledge. Validation can be viewed as experimenting with data and knowledge, using models as the lab equipment for performing computational experiments [21][3]. Simulation models to be validated should reflect the real world and results from experiments in simulation should emulate changes in the real world. If results from virtual or computational experiments are compared to real world data and match sufficiently, the simulation is sufficiently valid. Simulation [25][34] has an advantage over statistics and formal systems as it can model the world closely, free of the artifacts of statistics and formal systems.

There is related work in engineering design methods using Response Surface Methodology or RSM [28] and Monte Carlo simulations [36] to do direct validation, but only with numerical data and limited to a small number of dimensions. RSM is collection of mathematical and statistical techniques for the modeling and analysis of problems in which a response of interest is influenced by several variables. It can include virtual experiments using Monte Carlo simulation. It usually tests only a few variables and operates to find the best fit equation so that the correlation of equations predictions with real data is statistically significant.

4 Our Approach: Knowledge-Based Simulation-Aided Model-Improvement

WIZER (What-If AnalyZER) is a coupled inference and simulation engine that improves upon Response Surface Methodology to deal with the high dimensional, symbolic, stochastic, emergent, and dynamic nature of multi-agent social-network systems. Viewing simulation systems as knowledge systems, WIZER is designed for controlling and validating them directly with empirical data and knowledge using pattern analyses and knowledge inferences (mimicking those of SMEs) and virtual experiments (mimicking those of RSM).

WIZER integrates an inference engine and simulation virtual experiments to do calibration and validation for model-improvement and to provide explanations. It improves on RSM features by performing knowledge-intensive data-driven search steps via an inference engine constrained by simulation outputs, instead of just doing statistical and mathematical calculations. WIZER facilitates knowledge-based simulation control and simulation-assisted inference, enabling reasoning about simulations and simulation-assisted reasoning. It enables the management of model assumptions, contradictory or incomplete data, and increases the speed and accuracy of model validation and analysis. It is capable of explaining the reasoning behind inferences using both the simulation and inference engine. Search in WIZER is performed using both simulation and knowledge inference. The amount of search is reduced as the knowledge inferences, empirical data and knowledge, and virtual experiments constrain the search space.

WIZER seeks to emulate scientists doing experiments and analyses via the scientific method, instead of simply emulating an experimental setup. While other toolkits such as Swarm (<http://wiki.swarm.org>), TAEMS [13][26], and Repast (<http://repast.sourceforge.net>) are designed with the goal of assisting the design and implementation of agent-based systems, WIZER is designed to help with scientific experimentation, validation, analysis, and model improvement. WIZER is conceptually able to run on top of any simulation system, including those constructed using Swarm and Repast toolkits. WIZER is basically a logical reasoning, experimentation, and simulation control engine with statistical and pattern recognition capabilities. This is similar to techniques scientists employ when designing, executing, and analyzing experiments. WIZER differs from Evolutionary Programming [19] as it does not need a population of mutation candidates and the mutation operator. Instead, WIZER applies knowledge inference to simulations to design the next simulation run, based on scientific experimental method. If the result of inferences mandates a radical change, a revolution would occur. WIZER also differs from Evolutionary Strategies and Genetic Algorithms [15] as it does not use recombination/crossover operators. In short, WIZER employs a unique logical reasoning, simulation control and scientific method for doing virtual experiments. Explaining what a simulation system does and what happens in

simulation to SMEs is important from validation perspective. Utilizing its inference engine, WIZER can provide automated explanation of the happenings and emergent behaviors within a multi-agent simulation system.

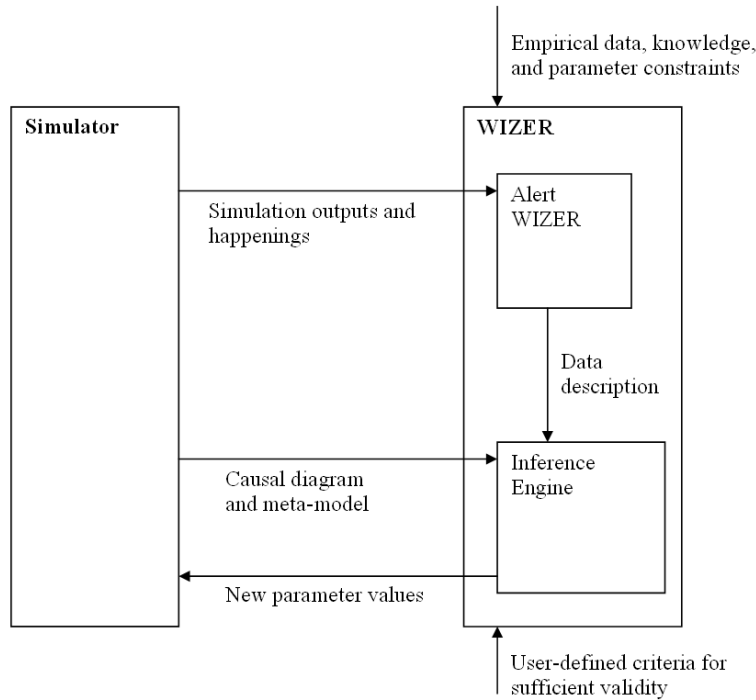


Fig. 1. WIZER Diagram

As shown in Figure 1, Alert WIZER takes in the simulation output data and determines which data streams of the simulation outputs do not fall within the empirical data value ranges and how. The WIZER Inference Engine takes the simulators causal diagram of what parameter influences which output data and the empirical constraints and confidence intervals on parameters to make a judgment on which parameters to change and how (including causal links and the model or agent submodel itself, if necessary). This results in new parameters for the next simulation. This simulation in turn yields new outputs which are fed back into WIZER.

This cycle repeats until a user-defined validity level is achieved. Thus, WIZER consists of:

- A system for determining which outcome variables match or fall within the acceptable range of the real data Alert WIZER. This system will create an "alert" when there is not a match. Inputs to Alert WIZER include real

and virtual data. Real data include various types of data such as subject matter experts (SMEs) estimation of behavior, 1st, 2nd, and 3rd order statistics for data streams at the yearly, seasonal, monthly, and day of week level, and actual streams of data. Alert WIZER includes statistical tools.

- An intelligent system for identifying which of the "changeable" parameters should be changed and how to improve the fit of the virtual to the real data the WIZER Inference Engine. This component uses a database relating parameters to the variables and modules they impact. This includes assumptions about the expected range for parameter values (according to SMEs) or best guesses, thus placing confidence measures on parameters.
- A local response surface analysis feature that can run simple virtual experiments for parametric studies.

The knowledge bases in the inference engine are populated with the knowledge about the simulator, simulation outcomes, domain facts and knowledge, assumptions, ontology, problem solving strategies, information about statistical tools it employs and other data. The knowledge bases contain both knowledge (hard or certain rules and facts) and assumptions (soft or uncertain rules and facts). Simulation outcomes provide measurements of the degree-of-support an assumption has. These different types of knowledge are included to enable the inference engine to reason about its reasoning. For example, knowledge about the simulation allows the inference engine to back up its symbolic reasoning with simulation outcomes and also to reason about the simulation. Part of the knowledge base is portable between simulations, but users need to provide the remainder.

The emergence of causal links based on low-level interactions can be probed by the inference engine, including probes to see what an individual agent does in its life and what events affected this agent and why, in addition to sample based probes. For sample based probes, WIZER conducts inferences based on the application of its included statistical tests.

The WIZER Inference Engine was inspired by the rule-based Probabilistic Argumentation Systems (PAS) [22] for handling assumptions. While a rule-based system is sufficient if knowledge engineers are able to check the causal relations inherent in rules, for large knowledge bases manual checks are cumbersome and prone to errors. Thus, there is a need for automated and formal causal checking. Fortunately, causal analysis has been treated mathematically [31]. WIZER uses a novel probabilistic argumentation causal system (PACS), which utilizes the probabilistic argumentation [22] in causal analysis [30]. Users of WIZER specify which rules are causal in nature and WIZER is capable of suggesting causal links and performing empirical computations to provide justification for these causal links. Results from social network analysis form one silo of domain knowledge fed into the WIZER inference engine. The inference engine in turn, along with the execution of virtual experiments in simulations, provides knowledge-based grounding for the emergence and

evolution of social networks from low-level agent behaviors and interactions. The causal mechanisms encoded in WIZER enable formal computation of interventions or actions, instead of mere observation. This allows WIZER to make changes in parameters, causal links, and meta-models, and to analyze the consequences. In other words, WIZER can emulate what scientists do by changing and analyzing experiments.

Causal analysis involves mechanisms (stable functional relationships), interventions (surgeries on mechanisms), and causation (encoding of behavior under interventions). Associations common in statistics can characterize static conditions, while causal analysis deals with the dynamics of events under changing conditions. Simply turning off some potential causal links and re-simulating is insufficient and while counterfactual testing – checking would happen if (true) facts were false – can uncover causal effects, the method can fail in the presence of other causes or when other causes are preempted and it ignores the sufficiency aspect. These weaknesses of this (global) counterfactual test can be addressed by sustenance, providing a method to compute actual causation [30]. Sustenance means minimally supporting an effect. Actual cause is computed by constructing causal beams and doing local counterfactual test within the beams. Causal beams are causal links that have been pruned to retain a subset of causal links that sustains the occurrence of an effect. Dynamic beams are simply causal beams with a time dimension [30].

To account for the probability of causation, the causal model [31][30] specifies the use of Bayesian priors to encode the probability of an event given another event. It does not distinguish between different kinds of uncertainty. It is unable to model ignorance, ignores contradictions and is incapable of expressing evidential knowledge without the use of the probability distribution format. Since the intended use of WIZER is to do validation in environments with incomplete, contradictory, and uncertain knowledge and because WIZER needs to clearly delineate between assumptions and facts, we need an improved causal model, built by borrowing concepts from the Probabilistic Argumentation Systems (PAS). Table 1 shows the encoding of facts, assumptions, and rules for rule-based systems using probabilistic argumentation, while Table 2 shows the encoding of facts, assumptions, and causations for causal analysis enhanced with PAS-like assumption management. In both tables, let P_i be proposition i , a_i be assumption i , *causes* be the causation operator, and \Rightarrow be the implication operator.

Table 1. Rule-Based Encoding

Type of Knowledge	Logical Representation	Meaning
A fact	$P1$	$P1$ is true
A rule	$P1 \Rightarrow P2$	$P1$ implies $P2$
An uncertain fact	$a1 \Rightarrow P1$	If $a1$ is true then $P1$ is true
An uncertain rule	$a2 \Rightarrow (P1 \Rightarrow P2)$	If $a2$ is true then $P1$ implies $P2$

Table 2. Causation Encoding

Type of Knowledge	Logical Representation	Meaning
A fact	P1	P1 is true
A rule	P1 causes P2	P1 causes P2
An uncertain fact	a1 \Rightarrow P1	If a1 is true then P1 is true
An uncertain rule	a2 \Rightarrow (P1 causes P2)	If a2 is true then P1 causes P2

We call Table 2's formalism the probabilistic argumentation causal systems (PACS). WIZER includes both rule-based and causal formalisms. PACS algorithmic details are derived from both PAS [22] and causal analysis [31]. Simulation virtual experiments can be seen as a proxy for real world experiments when doing real world interventions would be unrealistic or unethical. Causal analysis uses computations based on real-world experimental and non-experimental data. WIZER adds another dimension to causal analysis: allowing quasi-experimental that is, simulated data.

The internal workings of the WIZER Inference Engine are complex, but its basic operations are simple. Let $P = p_1, \dots, p_n$ be propositions, $A = a_1, \dots, a_n$ be assumptions, h be the hypothesis and $K = c_1 \cap \dots \cap c_n$ be the knowledge base of clauses, where c_i is an element of the set of all possible A and P clauses. Let a be the (conjunctive) arguments supporting h . We have

$$a \cap K \Rightarrow h$$

or equivalently

$$a \Rightarrow \neg K \cup h$$

or equivalently

$$\neg(\neg K \cup h) \Rightarrow \neg a$$

$$K \cap \neg h \Rightarrow \neg a$$

In other words, if we know K and h , we can compute the supports, that is, the arguments supporting h . The hypothesis h is a clause produced by Alert WIZER after comparing simulation data streams with empirical data. After finding the arguments supporting h , the degree of support can be found, defined as

$$dsp(h, K) = prob(a \text{ support } a \text{ of } h \text{ is valid} \mid \text{no contradiction}, K)$$

Similarly, the degree of plausibility can be found, defined as

$$dpl(h, K) = prob(\text{no support of } \neg h \text{ is valid} \mid \text{no contradiction}, K)$$

These two measures are used to determine which arguments are the most relevant to the hypothesis at hand, pinpointing which parameter values, causal links, and/or submodels should be changed. In other words, hypothesis h is the input to WIZER Inference Engine and the arguments supporting h are the output, leading to changes in parameter and meta-model values.

The operations described above are performed for both rule-based and causal clauses. Then, for clauses denoted as causal, additional operations are performed to see whether and to what degree the causal relations are empirically correct, partially based on the degree of support and the degree of

plausibility. Sustenance, causal beams and actual cause are also computed. WIZER also performs virtual experiments as needed.

The intertwining causal computation and virtual experimentation capability of WIZER enhances PACS and is useful in simulations to:

- Provide a formal computational means to convert simulation results or happenings to user-friendly causal sentences and also a mechanism to arrive at probability distributions or profiles for assumption variables.
- Allow probing of existing and potential causal assumptions and links and examination of the robustness of causal links using empirical data and quasi-experimental data obtained by simulations based on other known mechanisms and empirical data. For example, a simulation may have modeled Washington DC and policy analysts would like to know the effects of quarantining certain city blocks or closure of some major roads to mitigate the spread of smallpox. The mechanisms, data values, and stochastic processes in the city model themselves do not contain direct answers to the above causal question. Utilizing causal computation would allow this question to be answered.
- Allow the formal modeling of interventions in simulations.
- Allow symbolic values/events to be considered in determining causal relations. For example, the recent shortage of flu vaccine caused the CDC to recommend restrictions on who received the vaccine, resulting in a stockpile of unused flu vaccine, partly because some eligible people believed that none were available due to the news. WIZER would be able to probe similar kinds of cause and effect relationships.
- Allow experimentation and simulation control. As WIZER modifies, runs, re-modifies, and re-runs simulations, it uses causal mechanisms to keep track of and help inform what causes a certain series of modifications to work or fail and to suggest possible next steps.
- Allow better inference by letting the inference engine run simulations in the midst of causal inferences as needed. This allows the examination of the empirical claims of causal inferences.
- Provide a way to automatically tweak agent meta-models and individual agents so that they are both realistic and able to coordinate in a realistic environment.

5 Run Setup and Empirical Data

WIZER was used to validate BioWar. As mentioned earlier, BioWar [7] is a city-scale spatial multi-agent social-network model capable of bioattack simulations. BioWar has a large number of variables and interactions. Application of the Spiral Development model [4] to BioWar code development means that any previous validation of model predictions may no longer apply to a new version.

We have implemented Alert WIZER, which takes the empirical data on school absences, workplace absenteeism, doctor visits, emergency room visits, with additional emergency room visitation data from SDI (Surveillance Data Inc.), and over-the-counter drug purchase data. It also uses the outputs of the BioWar simulator and conducts minimum bound checking, maximum bound checking and mean comparison.

The following empirical data was used to compute the empirical bounds and means for the Alert WIZER:

- NCES Indicator 17, 2002 (Year 2000 data), for calculating school absenteeism <http://nces.ed.gov/programs/coe/2002/section3/indicator17.asp>
- CDC Advance Data, from Vital and Health Statistics, no. 326, 2002, for calculating ER visits <http://www.cdc.gov/nchs/data/ad/ad326.pdf>
- CDC Advance Data, from Vital and Health Statistics, no. 328, 2002, for calculating doctor visits <http://www.cdc.gov/nchs/data/ad/ad328.pdf>
- 1997 US Employee Absences by Industry Ranked for determining work absenteeism <http://publicpurpose.com/lm-97absr.htm>
- Over-the-counter (OTC) Drug Sales extracted from Pittsburgh Supercomputing Centers "FRED" data containing pharmacy sales data.

BioWar simulation outputs include the data streams matching the above empirical data such as daily absences for each school.

6 Preliminary Results

WIZER was run on "Challenge 3" and "Challenge 4" data from BioWar [8] using an implementation of Alert WIZER. Challenge 3 data consists of 4 data streams with 10 simulation runs for each attack case (no attack, anthrax attack, and smallpox attack) for each of 4 cities. The city population and locations (buildings and facilities) were scaled at 20%. The parameters were adjusted following an execution of preliminary inference engine steps based on a partial causal diagram of BioWar. We present the means from the four Challenge 3 simulation output data streams in Tables 3-6.

Table 3 shows that the simulated means of school absenteeism rates for normal simulation cases (no bioattack) fall between lower and upper empirical bounds for the simulations of Norfolk, Pittsburgh, San Diego, and "Veridian Norfolk" (a part of Norfolk specified by Veridian, Inc.). For anthrax attack cases, the simulated means are higher than normal means but still lower than the empirical higher bounds. This is plausible as the empirical higher bound contains (contagious) influenza outbreaks and other disease cases. For smallpox attacks, however, the simulation mean for one city – San Diego – is higher than the empirical higher bound. Smallpox is highly contagious so this is also plausible. For other cities, the simulated means of school absenteeism remain within expected bounds.

Table 3. School Absenteeism

City, scale	Lower bound	Higher bound	No attack	Anthrax	Smallpox
Norfolk, 20%	3.04%	5.18%	3.45%	3.75%	3.55%
Pittsburgh, 20%	3.04%	5.18%	3.52%	4.67%	4.46%
San Diego, 20%	3.04%	5.18%	3.78%	3.81%	5.57%
Veridian Norfolk, 20%	3.04%	5.18%	3.73%	4.05%	4.31%

For the workplace absenteeism (Table 4), the simulated means are within the empirical bounds for normal (no attack) cases for all the cities. In case of anthrax attack, the workplace absenteeism means are higher than those for normal cases; and in three of four cities, higher than the empirical higher bound. For smallpox attack, the simulated means are higher than those for normal cases, and higher than the empirical higher bound for one of the four cities.

Table 4. Workplace Absenteeism

City, scale	Lower bound	Higher bound	No attack	Anthrax	Smallpox
Norfolk, 20%	2.30%	4.79%	2.72%	4.65%	2.82%
Pittsburgh, 20%	2.30%	4.79%	2.77%	5.79%	3.99%
San Diego, 20%	2.30%	4.79%	3.26%	4.99%	5.78%
Veridian Norfolk, 20%	2.30%	4.79%	3.16%	5.50%	3.81%

Table 5 shows that for doctor visits the simulated means for the four cities fall within the empirical bounds for normal (no attack) cases. For anthrax attack cases, the simulated means are higher than those for normal cases for two cities, and slightly lower for two other cities. For smallpox attacks, the means are higher than those for normal cases for three cities and the same for one city. The results for attack cases are imperfect but indicate correct trends. All means for anthrax and smallpox attacks are within the empirical bounds.

Table 5. Doctor Visit per Person per Year

City, scale	Lower bound	Higher bound	No attack	Anthrax	Smallpox
Norfolk, 20%	0.415	1.611	0.499	0.476	0.499
Pittsburgh, 20%	0.415	1.611	0.493	0.485	0.573
San Diego, 20%	0.415	1.611	0.726	0.753	0.796
Veridian Norfolk, 20%	0.415	1.611	0.707	0.821	0.738

For emergency room visits (Table 6), the simulated means for four cities fall within the empirical bounds for normal (no attack) cases. For anthrax attacks, the simulated means are higher than those of normal cases for two cities and slightly lower for two others. For smallpox attacks, the simulated means are higher than those for normal cases for three cities and the same for one city. The results for attack cases are imperfect but indicate correct trends.

Table 6. Emergency Room Visit per Person per Year

City, scale	Lower bound	Higher bound	No attack	Anthrax	Smallpox
Norfolk, 20%	0.056	0.232	0.112	0.108	0.112
Pittsburgh, 20%	0.056	0.232	0.109	0.106	0.129
San Diego, 20%	0.056	0.232	0.149	0.159	0.188
Veridian Norfolk, 20%	0.056	0.232	0.161	0.187	0.168

Challenge 4 data has 12 data streams: school absenteeism, work absenteeism, doctor visits, emergency room visits, emergency room visits using the Surveillance Data Inc. data, and seven drug type purchase data streams. Table 7 shows the percentage of validated data streams for six cities for the no attack case.

Table 7. Percentage of "Challenge 4" Simulation Output Data Streams Validated

City	Data Streams Validated
San Francisco	5 out of 12, or 41.67%
San Diego	7 out of 12, or 58.33%
Pittsburgh	7 out of 12, or 58.33%
Norfolk	6 out of 12, or 50.00%
Hampton	4 out of 12, or 33.33%
Washington DC	4 out of 12, or 33.33%

7 Discussion

Automation of simulation experiment control and analysis is rarely viewed as a critical feature of simulation systems; instead, experimental control, analysis, intervention, validation, and model-improvement are left for humans to perform. Most simulation platforms aim to provide tools to ease the coding of simulation systems, rather than automating the analysis, control, validation, intervention, and model-improvement. WIZER indicates that such automation can be very useful, especially when dealing with socio-technical and

public health problems which have a high degree of uncertainty and interactions. Based on empirical data and knowledge, simulations can bound the inferences and allow the empirical claims of the inferences to be investigated. At the same time, knowledge-based inference and control of simulation can reduce the number of simulation searches and virtual experiments that need to be conducted. Simulations and inferences on them here act like a dynamic version space on both search and knowledge spaces.

The results presented in this paper are preliminary. More WIZER and simulation runs are needed to get better statistics – such as the median and variance –, and to evaluate error margins, the effects of sample choices, search space traversal, and the performance of combined simulation and knowledge search, including the metrics for measuring the amount of search reduction in both search space and knowledge space. The performance of WIZER will be compared with that of human subject matter experts.

8 Acknowledgements

This research was supported, in part, by DARPA for work on Scalable Bio-surveillance Systems, the NSF IGERT9972762 in CASOS, the MacArthur Foundation, and by the Carnegie Mellon Center on Computational Analysis of Social and Organizational Systems. The computations were performed on the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputing Center. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA, the National Science Foundation, the Pittsburgh Supercomputing Center, the MacArthur Foundation, or the US Government.

References

1. Axtell R, Axelrod R, Epstein J, and Cohen M (1996) Aligning Simulation Models: A Case Study and Results. *Computational and Mathematical Organization Theory*, 1 no. 2: 123-141.
2. Axelrod R (1997) Advancing the Art of Simulation. In *Proceedings of the International Conference on Computer Simulation and the Social Sciences*, Cortona, Italy.
3. Bankes SC (2004) Models as Lab Equipment: Science from Computational Experiments. In *Proceedings of North American Association for Computational Social and Organizational Science (NAACSOS) Conference (2004)* ISRI, CASOS, Carnegie Mellon University, Pittsburgh, PA.
4. Boehm B (2000) *Spiral Development: Experience, Principles, and Refinements*, Spiral Development Workshop. Special Report CMU/SEI-2000-SR-008. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA.
5. Burton R and Obel B (1998) *Strategic Organizational Diagnosis and Design: Developing Theory for Application*, 2nd ed. Kluwer Academic Publishers, Dordrecht, the Netherlands.

6. Carley KM and Prietula M (eds) (1999) *Computational Organizational Theory*. Lawrence Erlbaum Associates, Mahwah, NJ.
7. Carley KM, Fridsma D, Casman E, Altman N, Chang J, Kaminsky B, Nave D, and Yahja A (2003) *BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events*. In *Proceedings of North American Association for Computational Social and Organizational Science (NAACSOS) Conference 2004*, Pittsburgh, PA, http://www.casos.ece.cmu.edu/casos_working_paper/carley_2003_biowar.pdf.
8. Carley KM, Altman N, Kaminsky B, Nave D, and Yahja A (2004) *BioWar: A City-Scale Multi-Agent Model of Weaponized Biological Attacks*. CMU-ISRI-04-101 Technical Report, Institute for Software Research International, Carnegie Mellon University, <http://reports-archive.adm.cs.cmu.edu/anon/isri2004/CMU-ISRI-04-101.pdf>
9. Chen L-C, Carley KM, Fridsma D, Kaminsky B, and Yahja A (2003) *Model Alignment of Anthrax Attack Simulations*. *Decision Support Systems in the special issue on Intelligence and Security Informatics*, in press.
10. Chen L-C, Kaminsky B, Tummino T, Carley KM, Casman E, Fridsma D, and Yahja A (2004) *Aligning Simulation Models of Smallpox Outbreaks*. In *Proceedings of the Second Symposium on Intelligence and Security Informatics*, Tucson, Arizona, June 10-11, (2004) Also in *Springer-Verlag Lecture Notes in Computer Science Vol. 3073*.
11. Cioffi-Revilla C (2002) *Invariance and Universality in Social Agent-Based Simulations*. In *Proceedings of the National Academy of Sciences of the U.S.A.*, 99 (Supp. 3):7314-7316.
12. Dastani M, Dix J, and Seghrouchni AEF (eds) (2004) *Programming Multi-Agent Systems*. *Lecture Notes in Artificial Intelligence 3067*, Springer Verlag, Berlin, Germany
13. Decker K (1996) *TAEMS: A Framework for Environment Centered Analysis and Design of Coordination Mechanisms*. In *Foundations of Distributed Artificial Intelligence*, Chapter 16, G. O'Hare and N. Jennings (eds), Wiley Inter-Science, pp. 429-448.
14. Dershowitz N (eds) (2004) *Verification: Theory and Practice*. Springer Verlag, New York, NY.
15. Dianati M, Song I, and Treiber M (2003) *An Introduction to Genetic Algorithms and Evolution Strategies*. Technical Report, University of Waterloo, <http://www.swen.uwaterloo.ca/~mdianati/articles/gaes.pdf>
16. Edmonds B and Bryson JJ (2004) *The Insufficiency of Formal Design Methods: the Necessity of an Experimental Approach for the Understanding and Control of Complex MAS*. In *Proceedings of AAMAS 2004*, ACM.
17. Epstein JM and Axtell RL (1996) *Growing Artificial Societies*. MIT Press, Cambridge, MA.
18. Epstein JM, Cummings DAT, Chakravarty S, Singa RM, and Burke DS (2004) *Toward a Containment Strategy for Smallpox Bioterror: An Individual-Based Computational Approach*. Brookings Institution Press, Washington, DC. <http://www.brook.edu/press/books/towardcontainmentstrategyforsmallpoxbioterror.htm>
19. Fogel LJ (1999) *Intelligence Through Simulated Evolution: Forty Years of Evolutionary Programming*. Wiley Series on Intelligent Systems, New York, NY.
20. Gilbert N and Troitzsch KG (1999) *Simulation for the Social Scientist*. Open University Press, Berkshire, United Kingdom.

21. Guetzkow HS, Kotler P, and Schultz RL (eds) (1972) *Simulation in Social and Administrative Science: Overviews and Case-Examples*. Prentice Hall, Englewood Cliffs, NJ.
22. Haenni R, Kohlas J, and Lehmann N (1999) *Probabilistic Argumentation Systems*. Technical Report 99-9, Institute of Informatics, University of Fribourg, Fribourg, Switzerland.
23. Huang C-Y, Sun C-T, Hsieh J-L, and Liu H (2004) *Simulating SARS: Small-World Epidemiological Modeling and Public Health Policy Assessments*. *Journal of Artificial Societies and Social Simulation*, 7, no. 4. <http://jasss.soc.surrey.ac.uk/7/4/2.html>
24. Jewell NP (2003) *Statistics for Epidemiology*. Chapman and Hall/CRC, Boca Raton, FL.
25. Law AM and Kelton WD (2000) *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York, NY.
26. Lesser V, Decker K, Wagner T, Carver N, Garvey A, Horling B, Neiman D, Podorozhny R, NagendraPrasad M, Raja A, Vincent R, Xuan P, and Zhang XQ (2004) *Evolution of the GPGP/TAEMS Domain-Independent Coordination Framework*. *Autonomous Agents and Multi-Agent Systems*, Volume 9, Number 1, Kluwer Academic Publishers, pp. 87-143.
27. Lucena C, Carcia A, Romanovsky A, Castro J, and Alencar P (eds) (2004) *Software Engineering for Multi-Agent Systems II*. *Lecture Notes in Computer Science* 2940, Springer Verlag, New York, NY.
28. Myers RH and Montgomery DC (2002) *Response Surface Methodology: Process and Product Optimization using Designed Experiments*, 2nd ed. John Wiley and Sons, New York, NY.
29. Nickles M, Rovatsos M, and Weiss G (eds) (2004) *Agents and Computational Autonomy: Potential, Risks, and Solutions*. *Lecture Notes on Artificial Intelligence* 2969, Springer Verlag, Berlin, Germany.
30. Pearl J (2000) *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, United Kingdom.
31. Pearl J (2003) *Statistics and Causal Inference: A Review*. *Test Journal*, 12 no. 2 (December): 281-345.
32. Pressman RS (2001) *Software Engineering*. McGraw-Hill, New York, NY.
33. Prietula MJ, Carley KM, and Gasser L (1998) *Simulating Organizations*. AAAI Press/The MIT Press, Menlo Park, CA.
34. Rasmussen S and Barrett CL (1995) *Elements of a Theory of Simulation*. *ECAL* 95, *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Germany.
35. Reifman J, Gilbert G, Parker M, and Lam D (2004) *Challenges of Electronic Medical Surveillance Systems*. *RTO HFM Symposium on NATO Medical Surveillance and Response: Research and Technology Opportunities and Options*, Budapest, Hungary. <http://www.rta.nato.int/Pubs/RDP.asp?RDP=RTO-MP-HFM-108>
36. Robert CP and Casella G (1999) *Monte Carlo Statistical Methods*. Springer Verlag, New York, NY.
37. Shervais S, Wakeland W, and Raffo D (2004) *Evolutionary Verification and Validation of Software Process Simulation Models*. In the 5th International Workshop on Software Process Simulation and Modeling, extended abstract.
38. Shervais S and Wakeland W (2003) *Evolutionary Strategies as a Verification and Validation Tool*. Portland State University draft paper, <http://www.sysc.pdx.edu/faculty/Wakeland/papers/EvoVandVRevD.pdf>

39. Simon H (1957) A Behavioural Model of Rational Choice. *Quarterly Journal of Economics* (69): 99-118.
40. Taber CS and Timpone RJ (1996) *Computational Modeling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-113. Sage, Newbury Park, CA.
41. Ward MD (1985) *Theories, Models, and Simulations in International Relations*. Westview Press, Boulder, CO.
42. Wasserman S and Faust K (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, United Kingdom.