ISIS: A Networked-Epidemiology Based Pervasive Web App for Infectious Disease Pandemic Planning and Response

Richard Beckman[*]
rbeckman@vbi.vt.edu

Keith R. Bisset[*]
kbisset@vbi.vt.edu

Jiangzhuo Chen[*]
chenj@vbi.vt.edu

Bryan Lewis[*]
blewis@vbi.vt.edu

Madhav Marathe[*†]
mmarathe@vbi.vt.edu

Paula Stretz[*]
pstretz@vbi.vt.edu

ABSTRACT

We describe Isis, a high-performance-computing-based application to support computational epidemiology of infectious diseases. Isis has been developed over the last seven years in close coordination with public health and policy experts. It has been used in a number of important federal planning and response exercises. Isis grew out of years of experience in developing and using HPC-oriented models of complex socially coupled systems. This identified the guiding principle that *complex models will be used by domain experts only if they can do realistic analysis without becoming computing experts.*

Using Isis, one can carry out detailed computational experiments as they pertain to planning and response in the event of a pandemic. Isis is designed to support networked epidemiology – study of epidemic processes over social contact networks. The current system can handle airborne infectious diseases such as influenza, pertussis, and smallpox. Isis is comprised of the following basic components: (*i*) a web app that serves as the user-interface, (*ii*) a middleware that coordinates user interaction via the web app with backend models and databases, (*iii*) a backend computational modeling framework that is comprised of highly resolved epidemic simulations combined with highly realistic control strategies that include pharmaceutical as well as non-pharmaceutical interventions and (*iv*) a backend data management framework that manages complex unstructured and semi-structured data.

Isis has been used in over a dozen case studies defined by the DoD, DHHS, NIH, BARDA and NSC. We describe three recent studies illustrating the use of Isis in real-world settings: (*i*) uses of Isis during the H1N1 pandemic, (*ii*) supporting a US military planning exercise, and (*iii*) distribution of limited stockpile of pharmaceuticals using public and private outlets.

[*]Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech. Blacksburg, VA 24061, USA.

[†]Department of Computer Science, Virginia Tech. Blacksburg, VA 24061, USA

KDD'14, August 24–27, 2014, New York, NY, USA.

ACM 978-1-4503-2956-9/14/08 ...$15.00.
http://dx.doi.org/10.1145/2623330.2623375.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services —*Web-based services*; I.6.3 [**Simulation and Modeling**]: Applications; I.6.7 [**Simulation and Modeling**]: Simulation Support Systems

Keywords

web app; computational epidemiology; simulation; HPC; public health

1. INTRODUCTION

A pandemic is an infectious disease outbreak affecting a large population across the entire globe, for example, the re-emergence of H1N1 influenza in 2009. Certain modern trends exacerbate the speed and severity of pandemics. This includes: (*i*) increased global population, (*ii*) increased mobility, both on locally and internationally, and (*iii*) increased populations of older and immunocompromised individuals. In spite of these factors, over the last 50 years, public health agencies across the world have made remarkable strides in reducing the social, economic and health impacts of such pandemics. Computational epidemiology is the development and use of computer models for understanding the spatio-temporal diffusion of disease through populations. Controlled experiments used to understand scientific phenomenon are much harder and often impossible to do when studying epidemiology, due to ethical, and often practical reasons. As a result, computational models play an important role in elucidating the space-time dynamics of epidemics. They also serve an important role in evaluating various intervention strategies, including pharmaceutical and non-pharmaceutical interventions [20, 14, 19, 18]. The role of individual behavior and public policies is critical in understanding and controlling epidemics.

Traditional mathematical models for studying epidemics focused on rate-based differential equation models [15]. In this approach, one partitions the population into subgroups based on various criteria (e.g., demographic characteristics and disease states), and use differential equation models to describe the disease dynamics across these groups. Although useful, the approach fails to capture the complexity of human interactions and behaviors. The diversity of human behavior makes planning and response complicated and motivates a measure-project-intervene modeling cycle. This naturally motivates an *interaction based approach*. It involves accurate modeling of the social interaction network and the disease dynamics. It uses an endogenous representations of individuals together with explicit interactions between these agents to generate and capture the disease spread across the social interaction network; see [23, 24]. In this sense, the approach goes beyond the mathematical modeling techniques discussed earlier, which assume homogeneous interactions within each segment of the population. It raises new technical difficulties though, including (*i*) synthesis of city and national scale social contact networks, (*ii*) high performance computing models to study disease spread and interventions over these networks and (*iii*) bigdata challenges spanning the ability to carry out complex computational experiments, data management and analytics. Nevertheless, recent advances in computing technology, machine learning, data mining and network science have made this approach a reality [10, 12, 22, 21, 11].

Public health epidemiology provides an important societal application to apply computational thinking and exploit big data. The big data challenges come in numerous forms: (*i*) a large number of diverse data streams pertaining to social, health, economic, policy and infrastructure realms that need to be synthesized to develop realistic social contact networks; (*ii*) the size of data sets; (*iii*) modeling the epidemics spread, and the impact of interventions, requires substantial computing resources; furthermore this produces data that is several orders of magnitude larger than the input data; (*iv*) the measure-model-analyze-act-measure loop implies that the input data and data generated by models has to be analyzed in near-real time.

2. OUR CONTRIBUTIONS

Here we describe Isis (*Interface to Synthetic Information Systems*): a web-browser based modeling and decision support environment for public health epidemiologists. Isis can be used for planning, situation assessment, counter-factual analysis and response as it pertains to infectious disease epidemiology. The current system can handle airborne infectious diseases such as influenza, pertussis, smallpox, etc. Isis is a part of a larger integrated cyber-environment and is comprised of the following basic components: (*i*) a web-browser based user-interface (i.e., a web app), (*ii*) a middleware that coordinates user interaction with the UI and with the backend models and databases, (*iii*) a backend computational modeling framework that is comprised of highly resolved epidemic simulations combined with highly realistic control strategies that include both pharmaceutical as well as non-pharmaceutical interventions and (*iv*) a backend data management framework that manages complex unstructured as and semi-structured data. Isis can support interaction based computational epidemiology discussed earlier.

Isis is designed specifically so that public health analysts can focus on design and analysis of complex computer experiments to support planning and control of epidemics rather than becoming computer scientists. Nevertheless, the backend is comprised of powerful high performance computing based modeling and data management methods. The group has been developing models for large complex social systems for over a decade. The motivation to develop such pervasive computing web apps providing seamless access was derived from our earlier experience, wherein we realized that complex models although scientifically interesting are of little use if they cannot be made accessible to a domain expert who is usually not a computer scientist.

The supporting underlying mathematical models have been reported in our earlier work – our focus here is on computational and data infrastructure that enables the development and deployment of the pervasive and scalable web app so that it can be used by public health analysts. The focus of this paper is on problem description, design decisions, and deployment challenges.

Isis has been used to support a number of federal planning and decision support exercises. We describe three such studies: (*i*) use of the tool during the H1N1 pandemic outbreak, (*ii*) supporting a DOD planning exercise for military preparedness, and (*iii*) socio-economic analysis of anti-viral distribution. We have chosen these examples for three reasons. First, they demonstrate the range of computational experiments that can be carried out using the tool. Second, they demonstrate the use of Isis in important real-world settings as opposed to an academic exercise. Finally, each of the studies identified new challenges that served as requirements for further development of the tool.

3. THE WEB APP

The web app is an interactive, easy to use, graphical user interface and is designed to support the running of experiments consisting of numerous simulations that generate distributions of outcomes to gain an appreciation of the time-varying state (the dynamics) of an epidemiological event. The tool specifically supports exploration of the variability of outcomes in this highly stochastic process. Experimental outcomes are used in analysis reports, which are a kind of distribution of numerous replicates of an experiment and are generally viewable as plotted graphs. The Isis tool is intended primarily to facilitate both the planning and course of action of analysis activities for analysts.

The Isis web app allows a user to set up detailed factorial experiments. Using a simple interface to an underlying digital library, a user can specify the following for their experiment: (*i*) a social contact network; (*ii*) a within-host disease progression model; and (*iii*) a set of interventions. Each intervention requires additional details such as compliance level, sub-populations to which the interventions are applied and intervention triggers. An experiment consists of sweeping one or more parameters across a user-specified range of values. See Figure 1 for an illustration of Isis experiments. After setting up the experiment, the user is provided access to the results of the simulations. A set of basic analyses are performed automatically and the results are displayed. See Figure 2 for an illustration of Isis analyses. The standard plots and epidemic curves provide very detailed information about the epidemic. Additional information such as the spatio-temporal dynamics and disease dendrogram (how the disease moved over the social network) is also available. A key aspect of Isis is its simplicity – we can train public health analysts to make effective use of the system in about three hours. Somewhat counter-intuitively, by hiding the computational models from the end user, Isis makes them much more accessible. See `http://ndssl.vbi.vt.edu/apps/isis/` for further details.

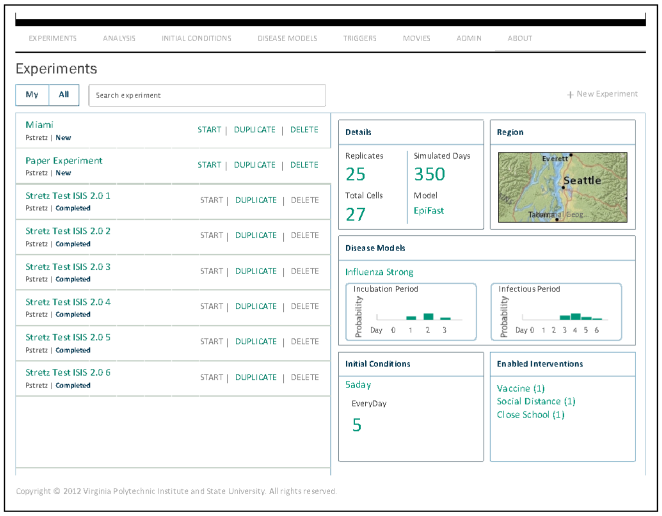

Figure 1: Experiment page of ISIS.

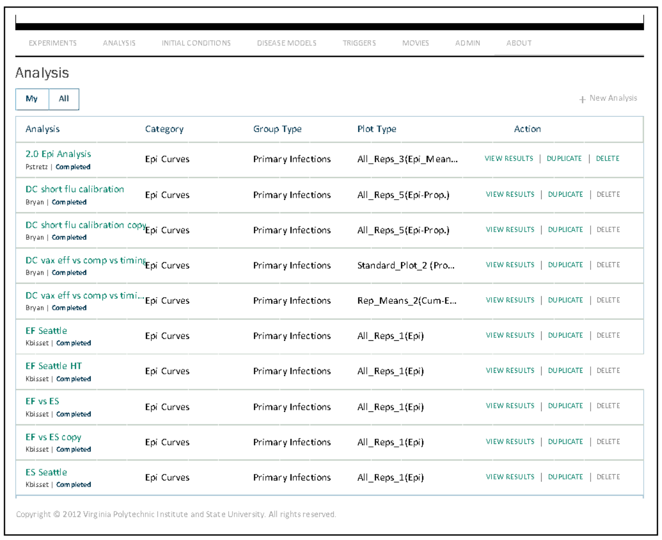

Figure 2: Analysis page of ISIS.

3.1 Designed Experiments in ISIS

ISIS is structured to easily make comparative studies of possible interventions using factorial or fractional factorial statistical designs. Rather than running and analyzing one simulation at a time, multiple simulations with differing combinations of interventions are simulated and analyzed as a way to compare the relative effectiveness of the interventions and to assess any non-linearity in these effects. The possible interventions in ISIS are: vaccinate a portion of the population, institute social distancing, close the work places or the schools, and use antivirals as a treatment or as a prophylaxis.

Figure 5 shows one screen of an experimental setup where interventions involving vaccination, social distancing, and closing schools are to be simulated and compared. The levels of vaccination required in the simulation experiment are shown in the figure: 0%, 10% or 20% of the population is vaccinated. In the experiment each of the other two interventions also has three different compliance levels. After 1% of the entire population is infected, 0%, 25% or 50% of

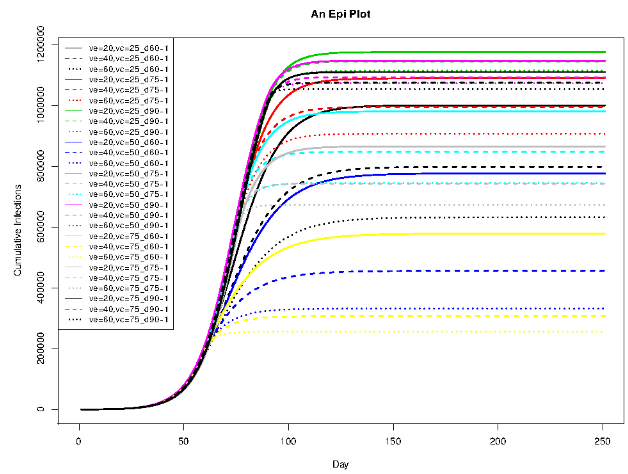

Figure 3: Analysis plots of a vaccination strategy experiment in the Washington, DC region comparing efficacy, compliance and timing of the vaccine.

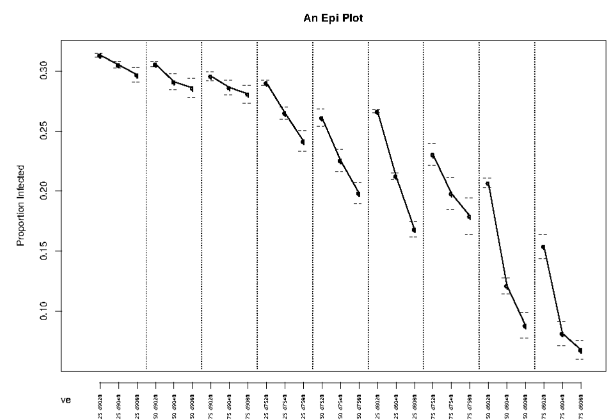

Figure 4: Analysis plots of a vaccination strategy experiment in the Washington, DC region comparing efficacy, compliance and timing of the vaccine.

the school children no longer attend school. Additionally, the entire population excludes 0%, 10% or 20% of nonessential activities. The experiment consists of simulations for all combinations of the levels of the interventions each replicated 25 times. Hence, the entire ISIS experiment is the factorial aggregate of $3 \times 3 \times 3 \times 25 = 675$ simulations.

The circles in Figure 6 are the mean number of persons infected (y-axis) for each of the treatment combinations on the x-axis. The x-axis has three rows one corresponding to each of the three interventions – vaccination (vc), close schools (sds) and general social distancing (sdg). The lines in the plot have no meaning and only serve to guide the eye. The x-axis is arranged so that the most important intervention is along the bottom row, the next most important in the next to last row, and so on. In this experiment, vaccination effects the number of infected persons the most as the infection or attack rate decreases most the as the vaccination rate

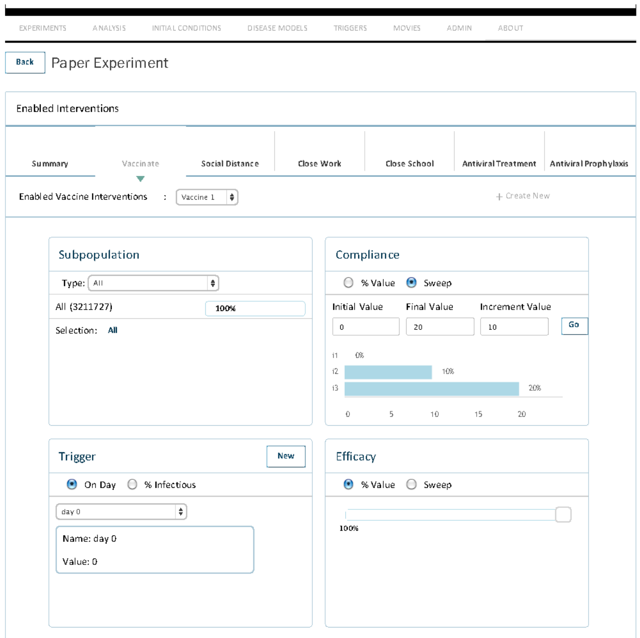

Figure 5: One of the three Isis screens for setting up a 3x3x3 experiment to assess the relative importance of vaccination, school closures, and social distancing. This figure shows that simulations are requested where 0%, 10% and 20% of the population is vaccinated. Similar screens exist for the other two interventions: social distancing and closing schools. In total 25 replicates of each of the 3x3x3=27 experimental combinations are simulated.

changes from 0% to 20%. Within each of the vaccination rates, closing schools has the largest effect at reducing the attack rate. It is interesting to note that the combination of the three interventions reveals a non-linear response in attack rates. That is, closing schools has a much larger effect on the attack rates if there is no vaccination than it does if 20% of the population is vaccinated.

3.2 Related work

Recently, there has been a flurry of activity aimed at developing easy to use interfaces to support computational infectious diseases epidemiology. Here we briefly discuss some of the prominent efforts.

1. Alex Vespignani and his colleagues at Northeastern University have developed EpiC and Gleamviz tools for global mobility and epidemic simulations. They developed a hybrid approach comprising of aggregate models for each county of interest and detailed networked connections between these counties capturing the flows of travelers based on daily airline passenger data.

2. Lauren Meyers and her group at the University of Texas at Austin have developed a set of tools DiCon for optimization and control problems related to epidemic dynamics, focusing on the logistics and supply chain [25].

3. The University of Pittsburgh group, a part of the NIH Modeling of Infectious Disease Agent Study (MIDAS) program, has developed a suite of tools for computational modeling, analysis and visualization of epidemiological processes. The tools use synthetic populations produced by RTI International [29] that are in turn based on an extension of the methodology developed by our laboratory. FRED [27] and GAIA [28] are the most related to our systems. FRED and GAIA are open sources systems and also support research in networked epidemiology.

4. The group at the University of Notre Dame is developing a cyber infrastructure for vector-borne diseases as a part of the Vector-Borne Disease Network consortium (VecNet); see [30]. VecNet is a part of the Malaria Education Research Agenda (MalERA) program established with the goal of reducing and hopefully eliminating malaria and funded by the Bill & Melinda Gates foundation.

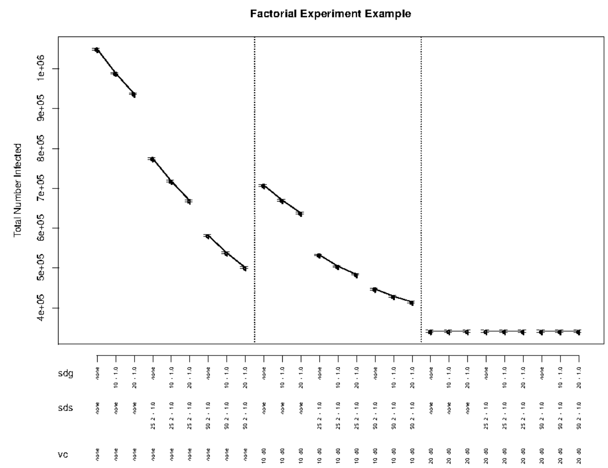

Figure 6: The attack rates (number of infected individuals) for each of the 27 intervention combinations are shown as the circles in the plot. The lines in the plot have no meaning other than to place the circles in logical groups. The x-axis has one row for each of the interventions with the most important intervention along the last row.

4. HPC MODELS FOR DISEASE DIFFUSION AND BEHAVIORAL ADAPTATION

The overall approach consists of four distinct models: (*i*) a model for creating a set of synthetic individuals, (*ii*) a model for generating (time varying) interaction networks, (*iii*) a model for simulating the epidemic process, and (*iv*) a model for representing and evaluating interventions and public policies [1, 13, 2]. Mathematically, these steps can be represented by a combination of: a co-evolving graphical discrete dynamical system (CGDDS) that captures the co-evolution of disease dynamics, social network and individual behavior (first three models) and a partially observable Markov decision process (POMDP) that captures various control and optimization problems formulated on the phase space of this dynamical system. Formal descriptions on CGDDS and POMDP can be found in [6, 5, 9].

Interventions implemented in our HPC simulations can be described with the combination of a set of trigger conditions, a set of subpopulations, and a set of actions. They are usually in the form of: "when *trigger conditions* are satisfied, apply *actions* to *subpopulations*". Table 1 describes how an intervention can be represented. There may be multiple trigger conditions, subpopulations, and actions in an intervention. There may be multiple different interventions in a simulation.

Representing and analyzing disease dynamics over large unstructured and time-varying social contact networks requires new work in high performance computing as well. Table 2 shows the data sizes of a few examples of synthetic populations and networks and simulation results supported by Isis. In an effort to address the scaling problem typical in algorithms implemented for high performance computing, our group has developed three different parallel algorithms and their implementations over the last ten years: EpiSimdemics, EpiFast, and Indemics [3, 31, 7]. These differ in the tradeoff they provide between computation speed, model realism and sophistication, and ease of introducing new behavior and interventions. A comparison between the simulations can be found in [5]. We briefly describe them below.

4.1 EpiSimdemics

EpiSimdemics [3], and its predecessor EpiSims [13], is an interaction-based, highly resolved modeling and simulation system for representing and reasoning about contagion diffusion across large networks. The EpiSimdemics algorithm is based on contagion diffusion across a social network. The network is represented as a bipartite graph, with people and locations as the nodes, and edges between them representing a person's presence at a location. We refer to this as a person-location graph. This representation allows location based interventions such as school closure to remove some contacts (e.g., contacts at a specific school) and replace them with contacts at a replacement location (e.g., home). EpiSimdemics can represent virtually all the existing models of between-host disease propagation. It supports fully dynamic social networks (nodes and edges can be added and removed in response to disease propagation). Through the use of its scenario scripting language [9], it also has the ability to represent a large collection of behavioral and public policy specifications, including the use of pharmaceuticals (e.g., vaccines or antivirals) and change of planned activities (e.g., school closure, quarantine). Contagion and behavior are modeled as coupled Probabilistic Timed Transition Systems (PTTS). Written in C++, EpiSimdemics initially used MPI but has recently been modified to use the Charm++ parallel framework [17, 4], which has led to improved scaling and execution time [31]. It can simulate the entire population of the United States, a person-location network with 300 million people and 1.5 billion edges. On the 352,000 core NCSA BlueWaters system, 120 days of an epidemic on this scale can be simulated in 12 seconds, as shown in Figure 7. To the best our of knowledge, this is the largest agent-based epidemic simulation, in terms of scaling.

4.2 EpiFast

EpiFast [7] differs from EpiSimdemics in the following ways: (*i*) The underlying person-person social network is explicitly given, whereas EpiSimdemics builds it implicitly using the people-location network [13]; (*ii*) EpiFast implements a simple linear multi-stage within host disease progression model, e.g., Susceptible Exposed Infectious Recovered (SEIR) disease model for epidemic simulations; (*iii*) EpiFast runs in bulk-synchronous mode; (*iv*) Interventions in EpiFast are coded as structural changes in the network (e.g., non-pharmaceutical interventions) or nodes' properties (e.g., infectivities and vulnerabilities). EpiFast is implemented in C++, using MPI for communication. It is similar to EpiSimdemics in that it can run on any distributed memory system, so long as the total of all available memory can hold the whole network and modeling parameters. The computing architecture is comprised of one master process and multiple worker processes. The system runs in a bulk synchronous parallel mode.

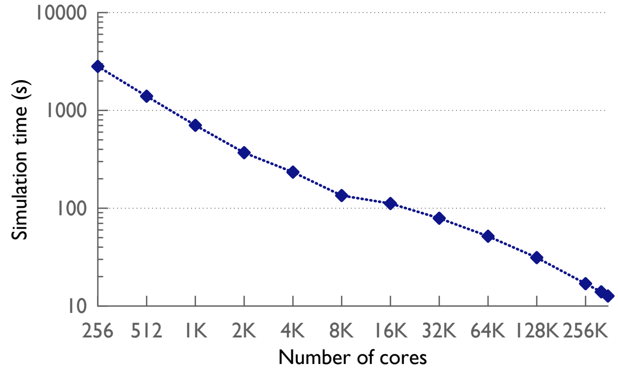

Figure 7: Execution time of EpiSimdemics on Bluewaters for United States population (300 million people) for 120 simulated days.

4.3 Indemics

Indemics [6] is an interactive, database driven HPC framework for epidemic simulations. It consists of (*i*) a simulation engine for computing epidemic dynamics, (*ii*) a situation assessment and intervention simulation engine supported by a relational database, (*iii*) a client (Indemics client) for user interactions, and (*iv*) a middleware platform (Indemics server) that connects the above components to provide online epidemic dynamics to the user for making decisions on interventions and to provide adaptive interventions to the simulation engine. Indemics uses EpiFast or EpiSimdemics as its underlying diffusion engine but is specifically designed to be interactive: a user can stop the simulation at any point, measure system state, and then proceed based on this information. The simulation is also allowed to roll-back to a previous time point if needed. Another notable feature of Indemics is its rich intervention set. In addition to a number of pre-defined interventions, a user can use SQL-based scripting language to define complex interventions [6]. The interventions can be based on the state of the disease and the dynamic, labeled people-location network (including individual attributes). Designing Isis to expose this rich set of interventions is a part of ongoing research.

5. ARCHITECTURE

Isis is hosted on the Simfrastructure service oriented middleware system [8], as shown in Figure 8. Simfrastructure serves as a multiplexer for four types of components that make up the integrated pandemic preparedness platform:

Table 1: The class of interventions provided through Isis.

component	specification	example
trigger	scheduled threshold	winter break Dec. 21–Jan. 3 prevalence > 0.05
subpopulation	demographic geographic social health combination of above	seniors people in Fairfax county, Virginia critical workers diagnosed (flu) people diagnosed seniors in Fairfax county
action	 vaccination antiviral school closure work closure isolation combination of above	effects: enhance immunity reduce symptoms or infection probability remove school activities and in-school contacts remove work activities and at-work contacts remove all contacts combined effects

Table 2: Examples of big data in Isis.

region	population size (million)	contacts (million)	simulation output size (per replicate)
Miami	2.2	55	15MB
Chicago	9.0	267	50MB
Delhi	13.8	210	180MB
California	33.6	962	190MB

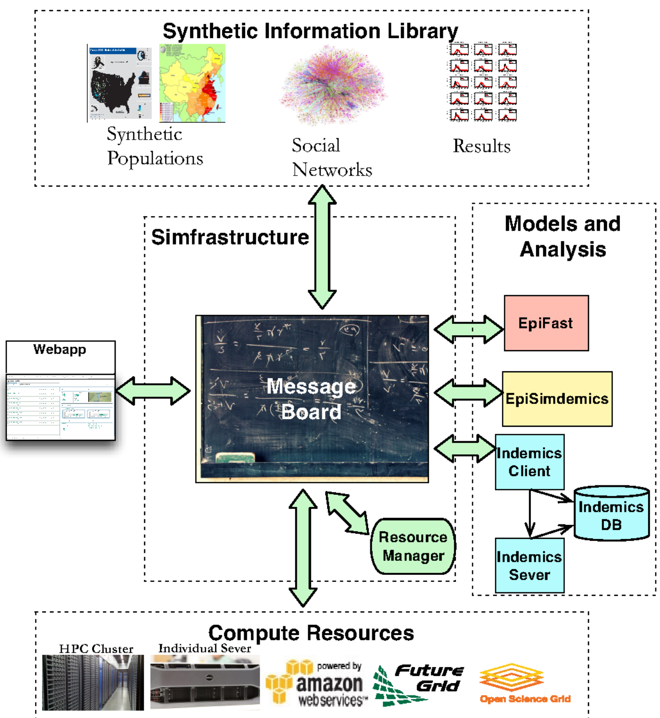

Figure 8: Architecture of Isis.

the web app, the synthetic information library, HPC models, and HPC resources. It is designed to allow any component to easily access the services provided by any other component without needing to know about the existence of any particular service provider, through the use of an asynchronous *Message Board*. The Simfrastructure *Resource Broker* matches service requests to appropriate service providers and computing resources, taking into account required timeliness and accuracy, as well as security, cost and other factors.

Simdemics is comprised of: (*i*) high-resolution scalable models of disease transmission and behavioral adaptation; (*ii*) a service-oriented architecture and delivery mechanism for facilitating the use of these models by domain experts; (*iii*) a distributed coordinating architecture for information fusion, model execution and data processing; (*iv*) a scalable data management architecture to support model execution and analytics; and (*v*) scalable methods for visual and data analytics to support analysts [1, 2].

The challenges faced when designing and developing Simfrastructure include:

Scalability The cyber infrastructure must be globally scalable. The scalability comes in three forms: (*i*) allowing multiple concurrent users; (*ii*) processing huge quantities of distributed data; and (*iii*) executing large, national-scale models.

Coordination The cyber infrastructure should allow computational steering of experiments. The systems needed by stake holders are geographically distributed, controlled by multiple independent, sometimes competing, organizations and are occasionally dynamically assembled for a short period of time.

Data and Information Processing The cyber infrastructure should facilitate efficient data and information fusion and analysis. The Internet has enabled data sharing in a simple and cost effective way, from the producers' side. Consumers of the data must still locate the appropriate data and deal with multiple incompatible data formats.

User Support The cyber infrastructure should provide an appropriate analysis framework for users, including user interfaces, high level formalisms to set up experiments, and visual and data analytics that include methods for integrating heterogeneous databases to support multi-view visualization (e.g., disease spread in a geographic region and epidemic curves).

SIMFRASTRUCTURE is designed to address many of these challenges. For example, disease diffusion models have been optimized so that they can be used to undertake studies over large urban regions. Multiple concurrent users are also currently supported but further scaling is required.

5.1 Synthetic Information Library

The Synthetic Information Library (SIL) contains all of the information needed to create, run, and analyze experiments in ISIS, and well as the results of past experiments. SIL enables structured and efficient management of data sources which may be located on a collection of hardware and stored in multiple formats. A registry maintains an index of all the data sources as well as the specifications of the data. Computational methods can map existing data sets to new data sets, and the registry records these processes to track data dependencies and also to support reproducible science. The registry may be compared to a library index providing central access all methods, data sets and their full specifications along with metadata.

SIL also provides services, from very generic to application specific, that allows high level operations on the data. Some examples of these services include provenance tracking, data curation, browsing and searching. One particularly interesting service is called *memoization*. When a request for computation arrives, it is looked up to see if that computation has already been performed. If so, the stored result is returned immediately. If the request has not previously been seen, it is run and the results stored for further use.

6. CASE STUDIES

ISIS has been deployed and used in dozens of user defined studies, all mainly focused on specific pandemic planning studies and exercises. Specifically users from DHS, DoD and DHHS have used ISIS to study the impacts of a variety of interventions in response to outbreaks of influenza or other respiratory diseases, assess the characteristics of emerging epidemics, and to optimize the allocation of existing medical resources. The studies have guided the continued evolution of ISIS both in terms of its usability for the design and execution of simulations as well the delivery and consumption of the simulation results themselves. The studies also helped us identify new research questions at the interface of multi-agent modeling, data mining, network science, and high performance computing. Notable studies include:

1. **Emergence of H1N1 Influenza**. Supported modeling efforts during H1N1 pandemic. One of the only three groups nationwide that was able to provide near real time support (2009 DHHS and DTRA).
2. **Unified Combatant Command Pandemic Study**. Combined national spread of influenza with high resolution local spread for several military installations to support decision making and planning (2010 Department of Defense).
3. **Socio-Economic Impact of Pharmaceutical and Non-Pharmaceutical Interventions**. Studied impact of economic and social influence on public health mitigation strategies during epidemics (2009 NIH Modeling of Infectious Disease Agent Study).
4. **Pandemic Tabletop Exercise Support**. Modeling support for Asst. Cabinet Secretary level exercise for response to emerging pandemic (2013 National Security Council).
5. **National Planning Scenario 1**. The role of individual behavior and communications in response and recovery in the aftermath of an improvised nuclear device explosion: health effect were studied using ISIS technology (2012 Department of Defense, DTRA).
6. **Military Response to Novel Virus**. Modeling support for fervent virus two day command tabletop exercise (2011 US NORTHCOM & DTRA).
7. **Military Pandemic Preparedness**. Modeling support for military and guard preparedness in the event of an epidemic outbreak (2007, 2008 Alabama National Guard, US MEDCOM and DTRA).
8. **Non-Pharmaceutical Targeted Layered Containment Strategy**. Evaluating sensitivity to timing and compliance level of non-pharmaceutical mitigation strategies (2005, 2006 National Security Council).

See `http://ndssl.vbi.vt.edu/apps/` and `http://ndssl.vbi.vt.edu/supplementary-info/vskumar/cacm2012/` for additional examples. In the subsections below we elaborate on three of the studies.

6.1 Emergence of H1N1 Influenza

This effort was in direct response to the initial reports of the emergence of the H1N1 influenza virus in April 2009, which eventually caused a global pandemic. In the early days of the outbreak, infections were confined to Mexico, California, and Texas and then spread to New York. The rapid spread combined with initial overestimates of its mortality rate raised serious concerns of a repeat of the 1918 influenza pandemic.

Initial reports about the disease characteristics were unreliable, with wide variations placed on important disease parameters like the proportion of symptomatic individuals and the duration of infectious periods. Having developed the ISIS tool for just this purpose, we were able to quickly run a series of studies exploring the impact of the variation in these parameters in a large US population. A quick report was drafted about the impact of disease characteristics on the size and shape of the expected epidemic curve. Several variants of disease models were added to the ISIS tool.

As H1N1 influenza continued to spread in the US, the Department of Health and Human Services teamed up with the Defense Threat Reduction Agency to place the ISIS tool in the hands of US government analysts to provide day to day modeling results. This integration inside the 24-hour decision cycle running the federal government's response to this emerging crisis would not have been possible without the development of highly optimized modeling software as well as the web-enabled interface [16]. The analysts were able to perform course-of-action analyses to estimate the impact of closing schools and shutting down workplaces. Better situational awareness was also enabled by calibrating the model to newly available data from the real world about the disease's characteristics, giving a clearer picture of how many infections were likely being missed by the existing surveillance systems.

This experience demonstrates the importance and feasibility of placing sophisticated modeling tools in the hands of public health decision makers and highlights the role that highly detailed modeling can play during a response to an emerging crisis.

6.2 Unified Combatant Command Pandemic Study

This work was conducted as part of an exercise hosted by a Unified Combatant Command to prepare defense decision makers for the information and response environment likely to be encountered in future influenza pandemics. This table-top exercise sought to provide the participants with a realistic course of events, information flows and stakeholder interests that will be involved during nationwide spread of influenza. Specifically we constructed and ran a multi-scale simulation of the nation-wide spread of influenza which served as the "ground truth" of the event, guided the exercise scenario and informed the exercise white team. The results of the spread based on a national model are shown in Figures 9 and 10. The simulation was run for a total of 300 days. The figures show how influenza spreads from the west coast to the eastern parts of the United States. Figure 9 shows the state of cumulative spread three months after the initial introduction of the disease and Figure 10 shows the state of cumulative spread after it has reached all major metropolitan areas. Our high resolution models enabled enhanced situational assessment, and leveraged surveillance information available to local public health authorities.

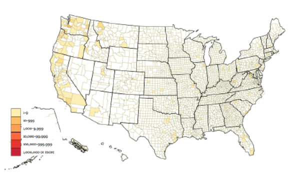

Figure 9: This is a snapshot of cumulative infections on day 91 of the influenza epidemic. The color scale shows the number of people infected at each location.

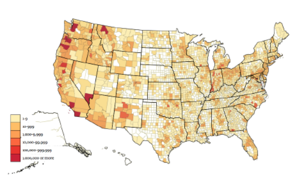

Figure 10: This figure shows cumulative infections on day 193 of the epidemic. The color scale shows the number of people infected at each location.

For realism, since no surveillance system can capture all possible infections, we used synthetic surveillance system techniques to estimate a realistic surveillance system signal for the epidemic, thus providing a realistic "partial" set of information to the different stakeholders participating in the exercise. Additionally the nation-wide simulation was used to run simulations on a more localized (and at a highly-detailed) scale enabling more realistic estimations of the epidemic's impacts on areas of particular interest (Ft. Lewis area, Ft. Sam Houston area, and Ft. Carson / Peterson AFB / USAF Academy area). Figure 11 shows attack rate with and without interventions for the cities of San Antonio, Texas and Seattle, Washington where Fort Sam Houston and Fort Lewis are located respectively.

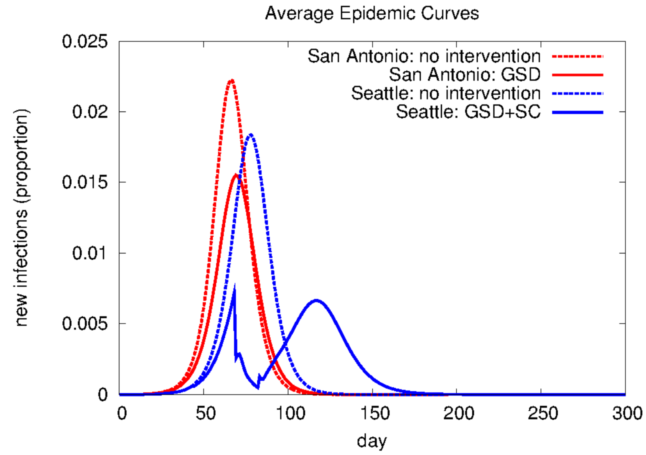

Figure 11: Mean epidemic curves for the city of San Antonio, TX and Seattle, WA. These cities were chosen because of the presence of military bases there. The "no intervention" case refers to the base case, GSD implies generic social distancing, and SC refers to school closure. The figure shows the effect of generic social distancing on the attack rate in San Antonio compared to the base case. In Seattle, generic social distancing is combined with school closure; their joint effect in controlling the attack rate is much larger.

The actual state of the epidemic, surveillance reports, detailed assessment of the impacts on the areas of interest, the interpreted actions of both public health and DoD officials, and simulation support events were all organized on a single timeline. This timeline was used to drive the table-top exercise. The timeline's interactive display concisely conveys multiple levels of information tailored to the particular interests and influence of different stakeholder communities while still enabling a vision of the overall progress of the exercise.

Requirements defined during support of this exercise led to several novel uses of existing analytical tools and techniques and to significant advances in very large scale, highly detailed social epidemiological simulation technology. This marks the first time that knowledge of existing surveillance systems was applied to simulation results to provide a realistically obscured situational awareness. This enhanced the realism for the table-top exercise for the participants. The exercise required that we integrate simulations occurring at different scales (nation-wide and regional). Technologically, this had not been done before and led to the development of a framework to enable this linkage. The use of the timeline to organize data from different sources and actions of different stakeholders while presenting a unified view of the events was important for the success of the exercise.

6.3 Socio-Economic Impact of Pharmaceutical and Non-Pharmaceutical Interventions

The goal of the project was to determine the economic and social impact of typical interventions proposed by public health officials and preventive behavioral changes adopted by private citizens in the event of an acute respiratory virus epidemic. We applied an individual-based simulation model to the New River Valley area of Virginia to address this critical problem. The economic costs included not only the loss in productivity due to sickness but also the indirect cost incurred through disease avoidance and caring for dependents.

People with different socio-economic constraints follow different behavioral strategies to avoid getting infected. These strategies are based on how individuals perceive both the state of the larger society around them and the partially observable actions taken by their immediate peer group or demographic class. The general principle being that high income individuals, young children and seniors have a lower tolerance of risk as compared to the rest of the population. Once the number of infected people in the population or in a person's class reaches the personal threshold value, the individual is triggered to modify his behavior. For members of affluent households, the modified behavior is reflected through the purchase of antivirals. The members of the middle income class eliminate non-essential activities such as shopping trips and recreational activities. Those in the poorest income class people find it too expensive to purchase antivirals, or reduce contacts and hence take no direct actions.

Public strategies include the distribution of antiviral kits and school closures. The trigger threshold for the public intervention is set at 1% of the total population becoming infected. The public stockpile of the antiviral kits is limited to 10,000. These kits are distributed to the individuals based on one of four selection criteria: randomly selected individuals, poorest individuals, first sick individuals, and the most vulnerable individuals. Each time the "close school" strategy is used, the schools are closed for a period of two weeks.

In order to assess the social and economic impact of the intervention strategies, we developed eleven distinctive scenarios based on individual and governmental actions. A base case was included to determine the size of the epidemic in the absence of interventions. The results show that the most important factor responsible for preventing income loss is the modification of individual behavior; it drops the total income loss by 62% compared to the base case. The next most important factor is the closure of schools, which reduces the total income loss by another 40%. Studies showed that the preventive behavior of private citizens is the most important factor in controlling the epidemic. The most effective interventions require school closures, public distribution of antivirals to the most vulnerable, and behavior modification by private citizens. The work appeared in [1] and was also mentioned as a part of the CDC report on school closures.

The requirements to complete a study of this nature required the extension of one of the core simulation engines used by Isis. Specifically the need to efficiently assess the state of disease and actions across several scales, as well as delivering fine control of thresholds for different classes of individuals. This extension brings the expressive power of SQL queries to the specification of both triggering conditions and targeted populations of interventions, greatly enhancing the realism of the interventions Isis can represent.

7. DISCUSSIONS

Data challenges. The data involved in Isis and the backend simulations are of very large scale, and dynamic, ambiguous, and heterogeneous in nature. The experiments on Isis usually consist of dozens of simulations, each of which simulates dynamics in synthetic contact networks of millions of nodes and billions of edges, and produces dynamics of each of the nodes during each of hundreds of simulations days. The size of the input and output data files of each simulation are in gigabytes. The output data is increasing as more experiments are run; the input data also keeps increasing because users are often interested in calibrating synthetic population, disease model, initial conditions, and interventions to do sensitivity analyses. It is impossible to consider all possible configurations users could come up with and pre-load them in Isis.

We addressed these challenges related to data volume, velocity, variety, and veracity with the Synthetic Information Library in Isis. This is different than the other systems, e.g., VecNet, EpiC and Gleamviz do not model person level details; DiCon does not have a GUI; FRED navigator allows one to access a database of previously computed scenarios. FRED Mobile App is the most similar to Isis but works only at a county level. Furthermore the current version of FRED allows limited number of interventions [27].

Lessons learned. The use of Isis in real world studies yielded important lessons and helped improve the system. First, early versions of Isis (including the dynamical models) were slow and somewhat cumbersome to use – the current version is significantly faster and easy to use improving the overall time it takes to design and analyze a case study.

Second, what we learned in designing and using Isis is the tradeoff between functionality and usability. While the backend HPC simulation tools (EpiSimdemics, EpiFast, and Indemics) can handle very complicated scenarios, defined by their scenario description languages, only limited features can be accommodated in a web application. We have worked with Isis users to include only the most useful features, while allowing expert users to access a few more complex functions.

Third, we learned to distribute the computation on multiple clusters. The use of cloud computing paradigm is quite appropriate for this purpose. the Simfrastructure middleware achieves this goal by seamlessly mapping jobs on various clusters available for the study.

Finally, as Isis becomes available to the broader community, there is a need to develop a scalable data management system including a digital library to support collaborative decision making. While the current system has a basic data management system, future versions will need to provide a rich set of services to achieve this goal.

Future work. Currently in Isis, experimental data is managed by an Oracle database server and output data is written in disk files. We are exploring using parallel databases to manage the simulation output data. We are also working on integrating Isis with FluCaster, a web app we developed for epidemic forecasting [26]. Finally, we are working on integrating optimal intervention strategies that would allow users to evaluate the efficacy of strategies that are currently implemented.

8. ACKNOWLEDGMENTS

We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments. This work has been partially supported by DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NIH MIDAS Grant 2U01GM070694-09, NSF NetSE Grant CNS-1011769 and NSF SDCI Grant OCI-1032677.

9. REFERENCES

[1] C. Barrett, K. Bisset, J. Leidig, A. Marathe, and M. V. Marathe. An integrated modeling environment to study the coevolution of networks, individual behavior and epidemics. *AI Magazine*, 31(1):75–87, 2010.

[2] C. Barrett, S. Eubank, and M. Marathe. Modeling and simulation of large biological, information and socio-technical systems: An interaction based approach. In *Interactive Computation*, pages 353–392. 2006.

[3] C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. V. Marathe. EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the ACM/IEEE conference on Supercomputing*, pages 290–294, 2008.

[4] K. Bisset, A. Aji, T. Kamal, J.-S. Yeom, M. Marathe, E. Bohm, and A. Gupta. Contagion diffusion with EpiSimdemics. In *Parallel Science and Engineering Applications: The Charm++ Approach*. 2013.

[5] K. Bisset, J. Chen, C. J. Kuhlman, V. S. A. Kumar, and M. V. Marathe. Interaction-based HPC modeling of social, biological, and economic contagions over large networks. In *Proceedings of the Winter Simulation Conference*, 2011.

[6] K. R. Bisset, J. Chen, S. Deodhar, X. Feng, Y. Ma, and M. V. Marathe. Indemics: An interactive high-performance computing framework for data intensive epidemic modeling. *ACM Transactions on Modeling and Computer Simulation*, 24(1), 2014.

[7] K. R. Bisset, J. Chen, X. Feng, V. Kumar, and M. V. Marathe. EpiFast: A fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In *Proceedings of 23rd ACM International Conference on Supercomputing*, pages 430–439, 2009.

[8] K. R. Bisset, S. Deodhar, H. Makkapati, M. V. Marathe, P. E. Stretz, and C. L. Barrett. Simfrastructure: A flexible and adaptable middleware platform for modeling and analysis of socially coupled systems. In *Proc. IEEE CCGRID*, pages 506–513, 2013.

[9] K. R. Bisset, X. Feng, M. V. Marathe, and S. M. Yardi. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In *Proceedings of the Winter Simulation Conference*, pages 2020–2031, 2009.

[10] J. Brownstein, C. Freifeld, and L. Madoff. Digital disease detection – harnessing the web for public health surveillance. *N. England J. Med.*, pages 2153–2157, 2009.

[11] P. Chakraborty, P. Khadivi, B. Lewis, A. Mahendiran, J. Chen, P. Butler, E. O. Nsoesie, S. R. Mekaru, J. S. Brownstein, M. Marathe, , and N. Ramakrishnan. Forecasting a moving target: Ensemble models for ILI case count predictions. In *Proceedings of the 14th SIAM International Conference on Data Mining*, 2014.

[12] M. Dredze. How social media will change public health. *IEEE Intelligent Systems*, Vol. 27(4):81–84, July–Aug 2012.

[13] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modeling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004.

[14] N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452, 2006.

[15] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[16] N. Hupert, J. Olsen, and A. Kircher. Personal communication, 2010.

[17] L. Kalé and S. Krishnan. CHARM++: A portable concurrent object oriented system based on C++. In *Proceedings of OOPSLA'93*, pages 91–108, 1993.

[18] M. V. Kerkhove and N. Ferguson. Epidemic and intervention modelling — a scientific rationale for policy decisions? Lessons from the 2009 influenza pandemic. *Bull World Health Organ.*, pages 306–10, 2012.

[19] M. Lipsitch, L. Finelli, R. T. Heffernan, G. M. Leung, and S. C. Redd. Improving the evidence base for decision making during a pandemic: The example of 2009 influenza A H1N1. *Biosecur Bioterror.*, 9(2):89–115, 2011.

[20] I. M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. T. Cummings, and M. E. Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

[21] M. Marathe and N. Ramakrishnan. Recent advances in computational epidemiology. *IEEE Intelligent Systems*, 28(4):96–101, 2013.

[22] M. Marathe and A. Vullikanti. Computational epidemiology. *Commun. ACM*, 56(7):88–96, July 2013.

[23] L. A. Meyers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of The American Mathematical Society*, 44:63–86, 2007.

[24] L. A. Meyers and N. Dimitrov. Mathematical approaches to infectious disease prediction and control. In *INFORMS, Tutorials in OPerations Research*, 2010.

[25] L. A. Meyers, N. Dimitrov, and S. Goll. DiCon: Disease control system. http://www.bio.utexas.edu/research/meyers/dicon/.

[26] Network Dynamics and Simulation Science Laboratory. FluCaster. http://ndssl.vbi.vt.edu/apps/flucaster.

[27] U. of Pittsburgh. FRED – framework for reconstructing epidemiological dynamics. http://fred.publichealth.pitt.edu/tutorials.

[28] U. of Pittsburgh. The geospatial area and information analyzer (GAIA). http://midas.pitt.edu/gaia.

[29] RTI International. Synthetic population viewer. https://www.epimodels.org/midas/synthpopviewer_index.do.

[30] VecNet. Vector-borne disease network. https://www.vecnet.org.

[31] J.-S. Yeom, A. Bhatele, K. R. Bisset, E. Bohm, A. Gupta, L. V. Kale, M. Marathe, D. S. Nikolopoulos, M. Schulz, and L. Wesolowski. Overcoming the scalability challenges of epidemic simulations on Blue Waters. In *Proc. IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, 2014.