Verification of Compartmental Epidemiological Models using Metamorphic Testing, Model Checking and Visual Analytics

Arvind Ramanathan, Chad A. Steed, Laura L. Pullum
Computer Science and Engineering Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37830
Email: {ramanathana,pullumll,steedc}@ornl.gov

***Abstract*—Compartmental models in epidemiology are widely used as a means to model disease spread mechanisms and understand how one can best control the disease in case an outbreak of a widespread epidemic occurs. However, a significant challenge within the community is in the development of approaches that can be used to rigorously verify and validate these models. In this paper, we present an approach to rigorously examine and verify the behavioral properties of compartmental epidemiological models under several common modeling scenarios including birth/death rates and multi-host/pathogen species. Using metamorphic testing, a novel visualization tool and model checking, we build a workflow that provides insights into the functionality of compartmental epidemiological models. Our initial results indicate that metamorphic testing can be used to verify the implementation of these models and provide insights into special conditions where these mathematical models may fail. The visualization front-end allows the end-user to scan through a variety of parameters commonly used in these models to elucidate the conditions under which an epidemic can occur. Further, specifying these models using a process algebra allows one to automatically construct behavioral properties that can be rigorously verified using model checking. Taken together, our approach allows for detecting implementation errors as well as handling conditions under which compartmental epidemiological models may fail to provide insights into disease spread dynamics.**

I. Introduction

A. Compartmental Models in Epidemiology

Compartmental epidemiological models are used by a wide community of epidemiologists, public health officials and scientists to model how diseases spread and what strategies of control are most likely to succeed in case an epidemic occurs within a given population. Compartmental models segregate a population into distinct groups, namely, **S** - susceptible (part of the population previously unexposed to the pathogen); **I** - infected (part of the population affected by the pathogen); **E** - exposed (part of the population that is infected by the pathogen but not infectious); and **R** - recovered (part of the population that has successfully been cured of the infection). Although a number of different compartmental models exist such as SIR, SEIR, SIS and others, the selection of a particular disease model is inherently dependent on the disease, pathogen and the population it affects. Several studies have previously shown how compartmental models can successfully capture the behavior of a disease outbreak and what strategies of intervention may be most effective to combat the disease.

Compartmental models use ordinary differential equations (ODEs) to model the aspects of disease spread and control. We will first describe the SIR model, which is the simplest of the models. The SIR model uses three divisions in the population, namely (S, I, R). ODEs describe the dynamics of how individuals move from one group to the other, for example: $S \rightarrow I$ or from $I \rightarrow R$ and so on. Intuitively, once an individual is infected (I), s/he can move to the recovered (R) compartment, when s/he has successfully been cured of the infection. Once a person is infected, there is a mean time for which the person remains infected, which can be estimated from clinical data. Thus, the probability of an individual moving from I to R ($I \rightarrow R$) is inversely dependent on the mean infectious period (represented as γ, and also referred to as the recovery rate). The progression from S to I ($S \rightarrow I$) is dependent on three factors, namely, the number of infected people, the underlying population structure/demography (i.e., how people are connected to each other, number of deaths/births, etc.), and the infection rate of the disease (denoted as β). In the simplest possible model, it is possible to forget the underlying demography and just capture a closed population with uniform mixing probabilities. The transmission term would then be a product of the infection rate along with the S and I compartments. Based on these observation, it is possible to describe the disease spread dynamics of the SIR model using the ODEs shown below:

$$\frac{dS}{dt} = -\beta SI; \quad (1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I; \quad (2)$$

$$\frac{dR}{dt} = \gamma I. \quad (3)$$

The SIR model can be further refined to include demographics, where we introduce a mortality rate (μ) in the population. The ODEs describing the SIR model with demographics will then

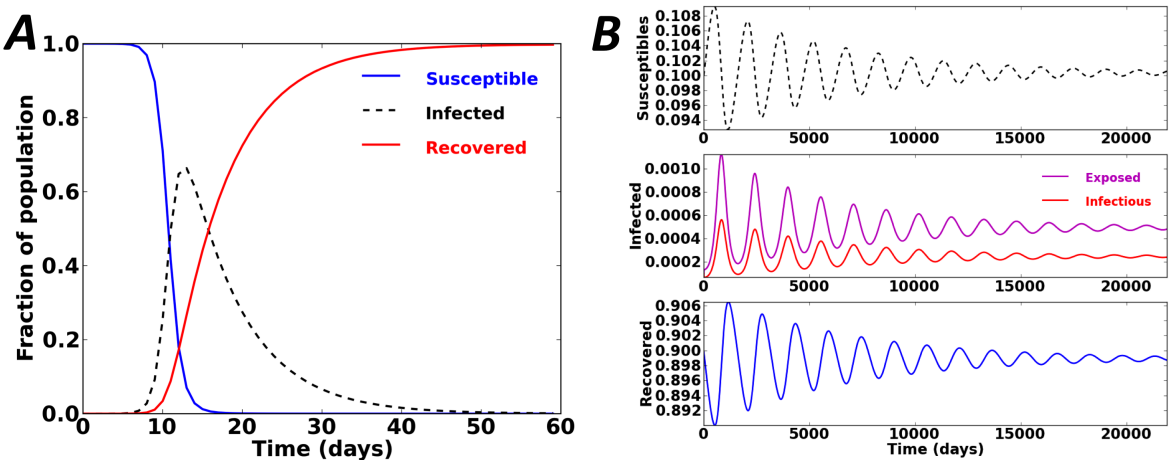

Fig. 1. **Illustration of SIR and SEIR models in epidemiology**. (A) The simple SIR model without demography, with the following parameters $(\beta, \gamma) = (1.4247, 0.1429)$ and initial conditions as: $[S(0), I(0), R(0)] = [1 - 1 \times 10^{-6}, 1 \times 10^{-6}, 0]$. Note the peak in the proportion of infected patients to that of the healthy population. (B) Illustration of the SEIR model with the same initial conditions, but with $\mu = 1/70 \times 365$. Note the oscillatory dynamics exhibited by the disease spread, as a consequence of introducing mortality rate in the population.

include:

$$\frac{dS}{dt} = \mu - \beta SI - \mu S; \quad (4)$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I; \quad (5)$$

$$\frac{dR}{dt} = \gamma I - \mu R. \quad (6)$$

From the specification of the SIR model, it is quite clear that the sum of (S, I, R) at every instant (t), must be 1. A second behavior that we also observe is regarding when the spread of the disease will fail: if the initial number of susceptible $(S(0))$ individuals is less than γ/β, then there will be no infection spread in the population. It is also interesting to note that the long term (asymptotic) behavior of the SIR model will lead to a situation where, in spite of having some susceptible population, the transmission of the disease will die down due to the lack of infected people. A third property that we can observe is that the SIR model with demography can exhibit oscillatory behavior (see below). An illustration of the simple SIR model is shown in Fig. 1A.

While the SIR model can be further refined to include demographics, heterogeneous interacting networks of individuals and different models of disease transmission, the simple SIR model cannot describe the dynamics of disease spread in the case of pathogens which need to be incubated to rapidly reproduce in the host before becoming infectious. Thus, there is a latent period (represented by $1/\sigma$) prior to which no infections can spread. This additional constraint on how the pathogen spreads adds an additional class of population, referred to as exposed (E), who are infected, but cannot transmit it to the susceptible population. Taking these additional model parameters, the SEIR model can be described using the ODEs shown below:

$$\frac{dS}{dt} = \mu - \beta SI - \mu S; \quad (7)$$

$$\frac{dE}{dt} = \beta SI - \mu E - \sigma E; \quad (8)$$

$$\frac{dI}{dt} = \sigma E - \mu I - \gamma I; \quad (9)$$

$$\frac{dR}{dt} = \gamma I - \mu E. \quad (10)$$

Behavioral properties similar to that of the SIR model can be defined for the SEIR model. However, an important distinction from the original SIR model, is that the SEIR model exhibits oscillatory behavior as the amplitude of the I values fluctuates and declines over time as the system equilibrates. For example, as illustrated in Fig. 1B, the SEIR model shows a distinct oscillation over every ~ 1200 days.

The behavioral properties described here represent a small subset of properties that can be obtained from the mathematical models. For a more comprehensive list of behavioral properties for different models, the interested reader is referred to [8].

B. Verifying Behavioral Properties of Epidemiological Models

The compartmental models described above provide an overview of how disease spread models can capture the diverse behavior of infectious diseases within a given population. The traditional view of validating these models has relied on the availability of clinical data from which a number of parameters are estimated and then fit to describe the underlying dynamics of disease spread. However, with increasing complexity in the nature of emergent diseases and population models, compartmental models have been augmented with other behaviors including stochastic models and spatio-temporal dependencies to understand the nature by which diseases may spread. These models are now being used by policy and public health officials to define intervention strategies for various diseases. Making such critical decisions based solely on models, without rigorously evaluating how and where these models might fail can have drastic impact on the social and economic systems in

case an intervention strategy fails. The increasing complexity of models also increases the number of parameters to include, which in turn results in a complex landscape to be modeled and understood. Further, with the development of agent-based models for epidemiology, the potential state-space for such systems can be quite enormous and complex, leading to a greater difficulty in validating and verifying if the disease spread models can be used to predict emergent behavior as well as strategies that can contain a disease outbreak. Therefore, there is a need to develop automated approaches to verify and validate such epidemiological models.

An important aspect of checking the behavioral properties of the models described above is to ensure that they are correctly implemented. Thus, formal testing techniques are necessary to obtain insights into implementation issues. In this paper, we use metamorphic testing as a tool to verify if the implementation of the SIR/SEIR models are correct. We input the results from simulating the disease spread model as well as metamorphic testing to a visualization toolkit, namely multi-dimensional data explorer (MDX) [10] to identify implementation errors within the models. Then, we verify several aspects of disease spread, including quantities such as maximum number of infected individuals (epidemic peak value), the time at which the maximum infection occurs (epidemic peak time), total number of infected individuals at any given time using model checking to reveal behavioral errors within the compartmental epidemiological models.

The rest of the paper is organized as follows: in the next section, we describe the methods used for verifying disease spread models. We provide a brief overview of related work in the areas of metamorphic testing, visualization and model checking and how they have been applied to verify epidemiological models. We then briefly summarize our experience in using these tools to verify and validate the SIR/SEIR compartmental models under three different scenarios. Finally, we conclude with a perspective of how our results shed light on creating a verification and validation framework for epidemiological models and how we can further improve these approaches.

II. Methods

The research described in this paper used a series of tools including metamorphic testing, model checking and visual analytics to verify compartmental epidemiological models. A schematic illustrating our workflow is shown in Fig. 2.

A. Metamorphic Testing of Epidemiological Models

Metamorphic testing was introduced by Chen and co-workers [12] to overcome the problem of the lack of a test oracle. The general idea behind metamorphic testing is to use well defined relations for a particular input such that the outcome can be tested. For example, consider a computer program which calculates the function $cos(x)$. While standard tests will include the values of $x = \{0, 60, 90\}$, for which $cos(x) = \{1, 0.5, 0\}$, one approach to metamorphic testing uses the identity $cos(2x) = 2cos^2(x) - 1$ to determine if

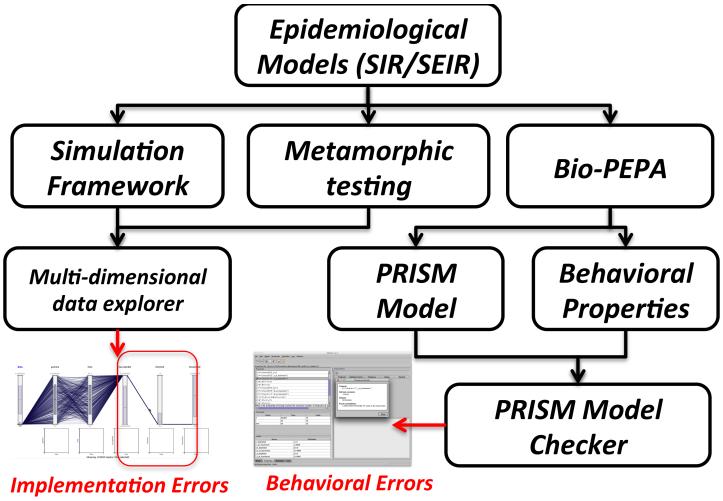

Fig. 2. **Overview of workflow to verify epidemiological models.** Using metamorphic testing, a novel visualization toolkit and process algebras, we were able to specify the behavioral properties of epidemiological models and verify them using a model checker. The outputs enable us to evaluate if and under what conditions these models fail.

the program is working correctly. This approach has been used to successfully test several bioinformatics applications and heat-transfer models. It must be noted that metamorphic testing requires definitions of relations for which the output can be known *a priori*; however, in epidemiological models, such relations can be challenging to define.

Instead of explicitly defining metamorphic relations on the differential equations itself, we use the constants β, γ and μ (Models 1, 4 and 7) to evaluate how the models behave with different values. For example, note that the SIR model with demography (Model 4) is a simple extension of the SIR model, with an additional parameter, namely, μ. By setting both μ to zero in this SIR model, one must obtain back the simple SIR model (Model 1). Thus, this 'morphing' of the SIR model should constitute the same results from the SIR model. In a similar manner, setting (μ, E) to zero in the SEIR model, should result in the simple SIR model above.

We also performed a scan of different β and γ values for both the models and tested if the decrease/increase corresponds to the expected decrease/increase in the proportion of susceptible, infected and recovered population. These parameter scans allow us to evaluate under what conditions an epidemic may manifest, especially, since from the differential equation model, we know that the behavioral specification for an epidemic to spread is $\beta/\gamma > 1$. This allowed us to measure changes in $(S, E/I, R)$ and trace through the runs to obtain a behavioral landscape of the models. The scanned parameters were then input into the visualization tool to examine the different models.

B. Visualizing parameter sweep and metamorphic testing

An important aspect of epidemiological model analysis is the determination of relationship between variables and the identification of the most significant associations. With

conventional tools, unexpected discoveres are nearly impossible. We have applied the Multidimensional Data eXplorer (MDX) system to analyze the parameter sweeps. MDX has been successfully applied to sensitivity analysis and more general exploratory data analysis of climate model simulation data [9] and long term climate studies [10] but this is the first application of the tool to empidemiology model analysis.

MDX is built around a popular multivariate information visualization technique called parallel coordinates. The parallel coordinates technique was initially popularized by Inselberg [6] as a novel method for representing hyper-dimensional geometries, and later demonstrated in the direct analysis of multivariate relationships in data by Wegman [11]. In general, the technique yields a compact two-dimensional representation of even large multidimensional data sets by representing the N-dimensional data tuple C with coordinates $(c_1, c_2, \ldots, c_N)$ by points on N parallel axes which are joined with a polyline [7]. The interactive parallel coordinates display is an effective method for analyzing the parameter sweeps because of the multidimensional display capabilities. A common set of parallel coordinates capabilities such as reconfigurable axes, details on demands, and axis inversion are available in MDX. The parallel coordinate plot has also been extended with a number of capabilities that facilitate exploratory data analysis and guide the user to the most significant relationships in the data. These features are highlighted in the following results but the reader is encouraged to explore prior publications for a more detailed explanation. [10]

C. Model-checking Epidemiological Models

Metamorphic testing provides a means to check whether the compartmental models are implemented correctly, and elucidate certain behavioral constraints imposed by scanning the respective parameters used in the compartmental epidemiological models. However, metamorphic testing may not fully explore all possible behaviors that are explored by ODE models. For example, consider the behavioral property where we want to estimate the probability that the number of infected patients (I) will be less than susceptible (S) people. Using metamorphic testing to examine all possible values of (S, I, R) and (β, γ) to estimate this probability can be cumbersome. Hence, formal methods and model checking can be complementary means to verify the correctness of these models and identify (automatically) behavioral properties that can fail, apart from characterizing them in the context of runtime behaviors of such systems.

Developed initially to capture faulty behavior in both hardware and software, model checking has now routinely been used to verify biological applications. Formal methods have been used to understand and characterize the behaviors of epidemiological models in the context of several applications, including modeling avian influenza and treatment strategies [3], describing conditions of disease outbreak and control across populations [4], [5]. These models have focussed on the use of formal methods to capture the behaviors of SIR/SEIR models and examine different mechanisms of intervention (such as isolation or vaccine treatment) and to study their outcomes. In this study we propose to examine the model checking of SIR/SEIR models under a variety of scenarios.

The models for our study were constructed using Bio-PEPA [2], a process algebra specifically developed for modeling biological/biochemical systems. These models were then input into PRISM, a model checker. We considered three common scenarios for compartmental models that are used by epidemiologists: (1) SIR/SEIR models (with and without mortality); (2) host-heterogeneity within the SIR model (children/adults) and (3) multi-pathogen models. The differential equation models were initially input into Bio-PEPA and automatically converted into a PRISM model using the Bio-PEPA Eclipse plugin. The Eclipse plugin also provides a set of behavioral properties to check. A list of these properties are discussed further in the Results section.

III. Results

A. Metamorphic Testing and Data Visualization

We begin by examining the results of metamorphic testing on the SIR/SEIR models outlined in the introduction section. First, we consider the setting $\mu = 0$ in the SIR model with demography (4) must result in the original SIR model (1). Indeed, when we simulate the system by replacing $\mu = 0$ in model (4), we find that the SIR model with demography reduces to the simple SIR model. A similar observation can also be made, when $E = 0$ in the SEIR model, which reduces to the SIR model with demography. A subsequent test of setting (μ, E) to zero, results in conversion of the SEIR model (with demography) to the simple SIR model. We also observe that when $\mu = 0$, it results in destroying the oscillatory behavior in the SIR and SEIR models. In our tests, we found that this approach of morphing allowed us to verify that the implementations of the more complex models reduced to the simple SIR model under special conditions mentioned above. In the context of metamorphic testing, the inherent relationships between the SIR $\rightarrow$ SIR-demography $\rightarrow$ SEIR-demography models allows us to examine the relationship between the addition of newer parameters and change in the behavioral properties of the system depending on these parameters.

A second test that we performed included a scan of (β, γ) for the SIR and SEIR model. In scanning through these parameters, we found under what conditions an epidemic may prevail. For example, we illustrate this by setting $\beta = 1$ and varying γ from a minimum of 0.2 to maximum of 0.5. The parameter settings are collectively visualized using the MDX tool. It is quite evident from the figure that when the value of γ is higher than β, there is no epidemic, as evidenced by the top lines in the susceptible column of the plot. However, as γ values tend to rise, there is a subsequent increase in the number of infected and recovered population. Thus, this simple scan of parameters allows us to visualize how β and γ are dependent on each other and further allows us to examine the behavioral properties of the simple SIR model in terms of when an epidemic may prevail in the population.

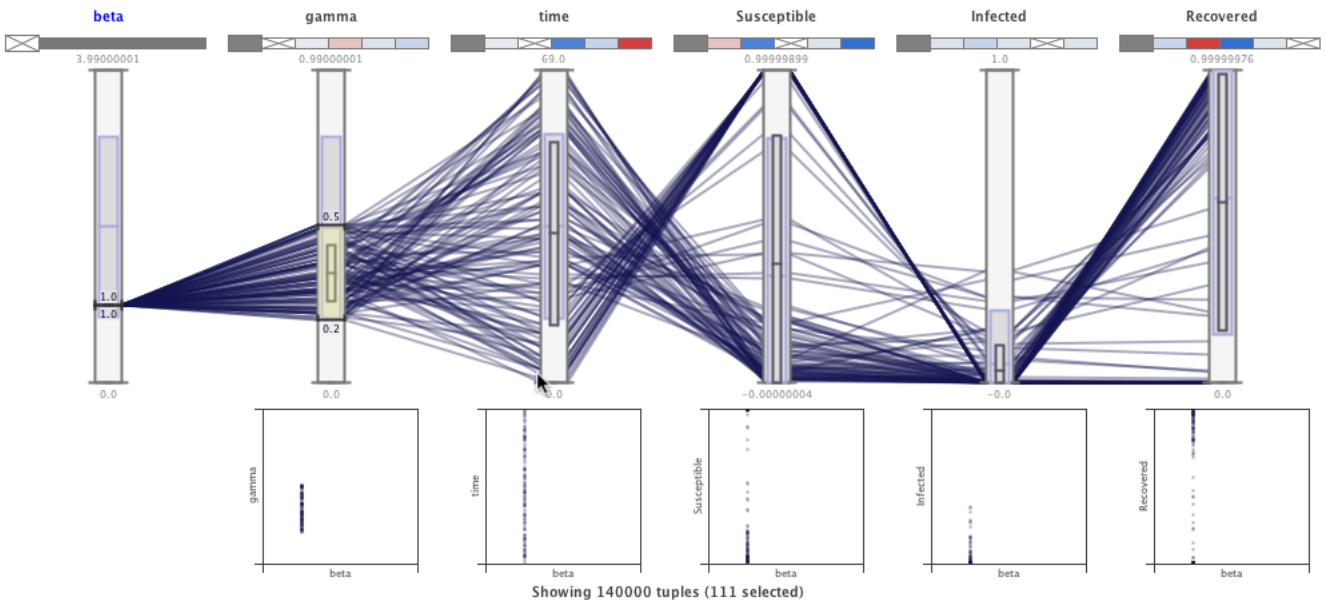

Fig. 3. In this case, we have selected the tuples with $\beta = 1$ and γ is between 0.2 and 0.5 (inclusive) to show sensitivities for the SIR simulation variables.

A similar scan of all parameters used in the SIR/SEIR can be visualized using the MDX tool, though we do not describe them in detail here due to lack of space. However, we must note that by scanning the parameter ranges for each of the models allows us to examine the special conditions under which an epidemic can prevail (or not prevail) as well as provide for a framework to visualize the data depending on certain ranges of the parameters used. In the case of epidemiological models, we found that parameter scanning along with data exploration and visualization provides novel insights into the behavioral properties of these models.

B. Model Checking

In the previous section, we discussed our results from the perspective of using metamorphic testing and visualizing how different parameters of the epidemiological models influence their behavioral properties. In this section, we discuss how one can formally specify and verify the behavioral properties of SIR/SEIR models using model checking. As outlined in the methods section, we used Bio-PEPA as a tool to specify the epidemiological models. The Bio-PEPA specification for the simple SIR model, introduced in model (1) is shown below:

```
beta = 1.4247;
gamma = 0.1429;

kineticLawOf infect : beta*S*I;
kineticLawOf recover : gamma*I;

S = infect <<;
I = infect >> + recover <<;
R = recover >>;

S[100] <*> I[0] <*> R[0]
```

The SIR model is translated into a Bio-PEPA specification by first defining the rate constants and then defining each term of the differential equations as a kinetic parameter in the model. The differential equations are themselves encoded by the different kinetic parameters, with $>>$ indicating an increase and $<<$ indicating a decrease in the different species of the system (S, I, R). Finally, the initial conditions are defined using the $< * >$ term, indicated. The PRISM model generated automatically from the Bio-PEPA specification also specifies a number of behavioral properties that can be verified using the model checker in PRISM. A few examples of the behavioral properties that can be verified in PRISM are illustrated below:

- Expected number of (S, I, R) at any given instant of time t.
- Probability of reaching the maximum number of (S, I, R) before time t.
- Probability of stability in the states of (S, I, R).

Complementary to the tests performed using metamorphic testing that capture initial conditions and implementation issues, model checking can provide any insights into the behavioral properties that we described in the introduction for each of the models considered (SIR, SEIR, SIR with demography and SEIR with demography). The simple SIR model has a total of 7,171 states with $\sim 14,000$ transitions defined by the PRISM model. Each state in the system corresponds to a specific configuration of (S, I, R) at some time instant t and the transitions correspond to the path taken by the program to reach those states. Model checking involves computing over this state-space the probable ways of reaching each state through a connected transition. In the case of the simple SIR model, for example, we found that the expected number of (S, I, R) at an instant $t = 10$ to be $(78.1, 28.4, 12.449)$.

Similarly, we can determine the probabilities of each of the behavioral properties specified above depending on the initial conditions. These properties allow us to know if (and what) behaviors of the SIR model are stable and valid at every time-instant.

IV. Discussion and Conclusion

With the development of complex epidemiological models, large scale simulations of epidemic spread and various intervention strategies are becoming highly popular in the literature. However, there is much speculation whether these complex models would be useful in real-time situations, especially when data collection, assimilation and analysis for these complex models have become a limiting factor in obtaining critical insights and predicting how a disease may spread and how one can control its spread. Therefore, there is an emerging need to develop rigorous approaches to verify and validate such complex epidemiological models.

In this paper, we have proposed and developed a preliminary approach using metamorphic testing, data visualization and model checking techniques to formally verify and validate compartmental epidemiological models. Beginning with metamorphic testing, where more complex models were reduced to simpler models by controlling certain parameters, we were able to obtain insights into how the SEIR model and SIR model with demographic data can be reduced to the simple SIR model. By varying the distribution of the parameters that control the behavior of these models, we were able to obtain insights into the behavioral properties of system, especially in the context of when an epidemic would occur, and how long it would last and so on. While we have found metamorphic testing of behavioral aspects of ODE based systems can be challenging, the testing strategies described here have allowed us to examine whether a particular model exhibits all known behavioral properties that can be obtained by analyzing the ODE systems. The visualization tools proposed here could also serve as a platform for further data analysis, especially for epidemiologists to visually cluster and analyze their large-scale data sets based on individual parameters of interest. The use of model checking, although not entirely new to epidemiological models, has allowed us to explore the various behavioral properties of these models in an automated fashion. Taken together, these tools have provided a framework to examine, verify and validate compartmental epidemiological models.

A future direction in our research would be to address how one can integrate testing and verification tools to study epidemiological models, including compartmental (or population) and agent-based models. One aspect to note is that although model checking can be used to verify simple SIR/SEIR models, exploring larger and complex models can render the state-space (and the behavioral properties) to be verified quite untenable for even efficient model checkers. For example, in the SEIR model with demography and including different age groups, we found that the state-space had exceeded nearly 1 million states. Exploring such large state-spaces is already a concern being addressed by the model checking community [1].

Acknowledgment

The authors would like to thank the Defense Threat Reduction Agency (DTRA) for partial support funded under the interagency agreement with Department of Energy (DOE) for DOE proposal number 2216-V387-11 as authorized by DOE contract number DE-AC05-00OR22725 and the DOE Laboratory Directed Research and Development Program at Oak Ridge National Laboratory (ORNL). The contents of this publication are the responsibility of the authors and do not necessarily represent the official views of DTRA. ORNL is operated by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

[1] R. H. Bordini, M. Fisher, W. Visser, and M. Wooldridge, "State-space reduction techniques in agent verification," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'04)*. IEEE Computer Society, 2004, pp. 896–903.

[2] F. Ciocchetta and J. Hillston, "Bio-pepa: a framework for the modelling and analysis of biochemical networks," *Theoretical Computer Science*, vol. 410, no. (33-34), pp. 3065–3084, 2009.

[3] ——, "Bio-pepa for epidemiological models," *Electronic Notes in Theoretical Computer Science*, vol. 261, no. 0, pp. 43 – 69, 2010, proceedings of the Fourth International Workshop on the Practical Application of Stochastic Modelling (PASM 2009). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S157106611000006X

[4] P. Drabik and G. Scatena, "An application of model checking to epidemiology," in *Proceedings of Applications of Membrane computing, Concurrency and Agent-based modelling in Population biology (AMCA-POP)*, P. Milazzo and M. J. Perez Jimenez, Eds., 2010, pp. 90–97.

[5] S. Huang, "Probabilistic model checking of disease spread and prevention," Master's thesis, Computer Science Department, University of Maryland, 2010.

[6] A. Inselberg, "The plane with parallel coordinates," *The Visual Computer*, vol. 1, no. 4, pp. 69–91, 1985.

[7] ——, "Parallel coordinates: Interactive visualization for high dimensions," in *Trends in Interactive Visualization*, E. Zudilova-Seinstra, T. Adriaansen, and R. Liere, Eds. London, UK: Springer-Verlag, 2009, pp. 49–78.

[8] M. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[9] C. A. Steed, G. Shipman, P. Thornton, D. Ricciuto, D. Erickson, and M. Branstetter, "Practical application of parallel coordinates for climate model analysis," in *Proceedings of the International Conference on Computational Science*, 2012, pp. 877–886.

[10] C. A. Steed, T.J. Jankun-Kelly, and P. J. Fitzpatrick, "Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates," in *IEEE Symposium on Visual Analytics Science and Technology*. Atlantic City, NJ: IEEE Computer Society, Oct. 2009, pp. 19–26.

[11] E. J. Wegman, "Hyperdimensional data analysis using parallel coordinates," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 664–675, 1990.

[12] Z. Zhou, D. Huang, T. Tse, Z. Yang, H. Huang, and T. Chen, "Metamorphic testing and its applications," in *Proc. of the 8th International Symposium on Future Software Technology (ISFST 2004)*, 2004.