Early Results from Metamorphic Testing of Epidemiological Models

Laura L. Pullum
Oak Ridge National Laboratory
Computational Sciences & Engineering Division
Oak Ridge, TN USA
1-865-574-4602
pullumll@ornl.gov

Ozgur Ozmen
Oak Ridge National Laboratory
Computational Sciences & Engineering Division
Oak Ridge, TN USA
1-865-241-0009
ozmeno@ornl.gov

ABSTRACT
The research reported in this paper is motivated by the need to validate models of disease spread. To date, reasoned confidence in the results of these models is complicated by the increasing complexity of the models and by their desired predictive use. As part of an overall approach to verification and validation of disease spread models, we investigate metamorphic testing. In this paper, we present early results of initial metamorphic testing on agent-based epidemiological models.

Categories and Subject Descriptors
I.6.4 [Model Validation and Analysis]

General Terms
Algorithms, Reliability, Experimentation, Verification.

Keywords
Epidemiological models, Verification, 1918 Spanish flu, Agent-based models, Metamorphic testing.

1. INTRODUCTION
Models are used to study the potential behavior and impact of disease spread in populations. Historically, epidemiological models have employed various modeling approaches ranging from equation-based models (EBM) to individual-based models (IBM) to agent-based models (ABM). Each type of model has its inherent strengths and weaknesses. EBM partition the population into “compartments” of susceptible, infectious, and recovered, for example in the SIR model. IBM describe behavior at the individual (person) level and report the results at the individual and aggregate levels. ABM have been widely deployed in many fields to study the collective behavior of large numbers of interacting agents and are of particular interest today in application to epidemiology. Like IBM, they model the phenomenon of interest at the individual level, however unlike IBM, ABM explicitly model the interaction between individuals.

Model (and data) validation ensures that the right model has been built and, along with model and data verification, lends confidence to the use of that model to inform critical decisions. In this paper, we describe our preliminary efforts and results in agent-based epidemiological model verification and validation (V&V) using metamorphic testing. The objective of the experiments reported herein is to determine whether metamorphic testing is a useful tool in the V&V of disease spread models, using as test subjects an EBM and an ABM of the 1918 Spanish flu. These are initial findings on a subset of potential metamorphic relations for testing disease-spread models.

In Section 2, we provide background on issues related to the V&V of agent-based models, on metamorphic testing, and on the models used in this experiment. Section 3 describes the experiment design and results, and is followed by our conclusions and notes for future research in Section 4.

2. BACKGROUND
2.1 ABM V&V Issues
In general, ABM belong to a class of software sometimes referred to as “non-testable programs”, described by Davis and Weyuker [1] as “programs which were written in order to determine the answer in the first place. There would be no need to write such programs if the correct answer were known.” Since there are no oracles for testing these programs, it is generally impossible to know a priori what the correct or expected output should be for a given input. Though often used, there are many criticisms regarding the use of ABM to study complex systems [2].

Recently discussed ABM V&V approaches cite several V&V frameworks and techniques as the bases to build upon, e.g., [3, 14]. Depending on access to the actual phenomenon investigated and on model complexity, these techniques include: 1) Compare ABM/system output with real phenomenon. This is a straightforward comparison, with the difficulty being access to sufficient ground truth data on the relevant aspects of the phenomenon under study. 2) Compare ABM/system results with mathematical model results. This approach has the disadvantage of requiring construction of the mathematical models, which may be difficult to formulate for a complex system. 3) Docking with other simulations of the same phenomenon. This approach aligns two dissimilar models to address the same question or problem, to investigate their similarities and their differences, and to gain new understanding of the issue being investigated [4]. These techniques are individually and collectively insufficient to provide requisite confidence in the validity of the models. (No single technique is sufficient.) We investigate metamorphic testing as an additional technique, which, with an overall V&V approach adds to the confidence required to use these models in real-world situations.

2.2 Metamorphic Relations and Testing
Chen et al. [5] introduced the metamorphic testing approach to address the problem of testing programs with no oracle. Metamorphic testing is a testing method that uses expected properties of the function or application under consideration to test programs. The properties, called metamorphic relations (MR), are functions that define the relationships between changes in
program, model, or function input and the expected changes in output, and can provide means to define V&V test cases.

Suppose \( x \) is a test case input and produces output \( f(x) \). The metamorphic properties of the function \( f \) can be used to develop a transformation function, which when applied to the test input produces \( x' \). This then enables prediction of the expected output \( f(x') \) based on the already known \( f(x) \). If the outcome \( f(x') \) is “as expected”, it is not necessarily correct. However, any violation of the metamorphic property indicates that one (or both) of the outputs, \( f(x) \) or \( f(x') \), is wrong. So, though it may not be possible to know with a single test whether an output is correct, we can determine if an output is incorrect. Metamorphic testing has been demonstrated at the function \([6, 7]\), application \([7-11]\) and simulation \([12]\) levels.

As a simple example illustrating the concept of metamorphic relations, consider a function that calculates the standard deviation of a set of numbers. For some transformations of the input set, we expect no change in the result, e.g., if the order of the members of the input set is permuted or if each member of the input set is multiplied by -1. Other transformations of the input set will predictably alter the output, e.g., multiplying each member of the input set by 2 will result in a standard deviation twice that of the original input set.

To date, there has been only a single published effort \([13]\) to apply metamorphic testing to ABM. In \([13]\), Murphy et al. investigate the use of metamorphic testing on a simulation of a hospital. We are developing the metamorphic testing technique and applying it to agent-based epidemiological models, specifically at the various components of ABM for epidemics – that is, the individual agents, a multi-agent system, the disease spread algorithm, the transportation model, the population model, and the agent-based disease spread model itself.

### 2.3 Models

In this study, we use one EBM (based on ordinary differential equations (ODE)) and one agent-based model, both of which implement the dynamics of a Susceptible-Infectious-Incubating-Recovered (SIIR) model. The ODE model was calibrated to match the original data observed in the Influenza outbreak of 1918 \([15]\).

For further understanding of the parameters and the implementation process, \([15]\) compares ABM and ODE models, and the models described in that paper are used in this study. The parameters for both models are provided below, along with their base values. The base values of the model parameters are those values that enable the original ODE and ABM models to mimic the 1918 flu pandemic.

- Probability of death given a person has contacted the disease. \( P(\text{Mortality}) = 0.01 \)
- Time to recover from the disease given that the infected person will not die. \( T(\text{Recovery}) = 2.5 \text{ days} \)
- Time to die from the disease given that the person with the disease will die. \( T(\text{Die}) = 1 \text{ day} \)
- Probability of transmitting the disease to another person upon contact with that person. \( P(\text{Transmission}) = 0.15 \)
- Number of persons encountered by an individual in a day. \( \text{Rate(Encounter)} = 4 \text{ people per day} \)
- Time for the disease to incubate. \( T(\text{Incubate}) = 3 \text{ days} \)
- Number of individuals initially infectious. \( N(\text{Infectious}) = 1,000 \).

### 3. EXPERIMENT

#### 3.1 Experiment Design and Expected Results

Table 1 (shown at the end of this paper) provides the model parameters, the metamorphic relations (MR) investigated for the parameters, and the expected results from running the model with the MR-transformed parameter values. The MR and expected results are identified based on domain knowledge and understanding of the underlying equations (ODE) of the EBM.

The first runs consisted of a) the baseline EBM and b) the baseline ABM. Other than the rate at which the disease spreads, the ABM mimics the ODE model used for ground truth. Given the results of analysis of variance of the ABM, we conservatively set the number of run iterations to 200. Each MR translation listed in Table 1 was investigated and the actual simulation results compared to the expected results are presented below.

#### 3.2 Results

For results reported here, we present the cumulative number of deaths over time. In each plot, the \( x \)-axis represents time in days, the \( y \)-axis represents the number of deaths as a proportion of the total population, and the red line is the data for the baseline ODE model. Figure 1 illustrates the results from the original 1918 Influenza pandemic scenario for both ODE and ABM models. We see that the flu killed 0.55% of the United States population, which was approximately 103 million at the time.

![Figure 1. Cumulative deaths over time - baseline scenario.](image-url)

Subsequently, we conducted runs for each of the metamorphic relations in Table 1 for both ABM and ODE models. Table 2 summarizes the cumulative number of deaths in the ABM and ODE simulations for each MR. The expected results are stated in terms of comparison with the original ODE model results. Below we discuss the response of the ABM model for each MR.
First, consider the MR on original \( P(\text{Mortality}) = 0.01 \). The specific scenarios considered are:

- **MR1**: \( x' = P(\text{Mortality}) \times 0.5 = 0.005 \)
- **MR2**: \( x' = P(\text{Mortality}) \times 2 = 0.02 \)

For MR2, we expect to observe an increase in the number of deaths, however the increase is expected to be less than a factor of 2. The reason is that as the number of deaths increases, the proportion of the susceptible population decreases. For MR1, we expect a decrease in the number of deaths. Figure 2 illustrates the original ODE results (in red), along with the ABM results for MR1 (blue) and for MR2 (black). In these tests, the ABM results meet the expected results, so we can state that the model is “possibly correct for these scenarios” and that these MR did not reveal a defect.

Next we consider MR on original \( T(\text{Die}) = 1 \) day. The specific scenarios considered are:

- **MR5**: \( x' = T(\text{Die}) \times 0.5 = 0.5 \)
- **MR6**: \( x' = T(\text{Die}) \times 2 = 2 \)

An increase in the time to die (MR6) is expected to increase the number of infected and the number of dead because it extends the period of contact between the infected, dying population and the susceptible population. Figure 4 illustrates how the time to die impacts the number of deaths over time. As expected, the increase is observed at low levels since the transmission and the mortality probabilities also highly impact the expected number of deaths. The ABM results meet the expected results for this test. Again, we can only state that the model is “possibly correct for these scenarios” and these MR did not reveal a defect.

Results from MR on the encounter rate, probability of transmission, and time to recover exhibited similar patterns for the resulting number of deaths. As such, we present the results only for \( P(\text{Transmission}) \) (figure 5). The MR applied to the original \( P(\text{Transmission}) = 0.15 \) are:

- **MR7**: \( x' = P(\text{Transmission}) \times 0.5 = 0.075 \)
- **MR8**: \( x' = P(\text{Transmission}) \times 2 = 0.3 \)

In some iterations of the scenario with an initial infectious population of 100, the epidemic never gets started. The results of this case exhibited a constant, low cumulative number of deaths. This raised questions about the veracity of the ABM model. After analyzing the code, an error in the output method was identified and corrected. Figure 3 illustrates how the number of initial infectious people impacts the number of deaths over time (for the revised code). As expected, the number of initial infected population has a very small effect on the cumulative number of deaths.

### Table 2. Average cumulative number of deaths.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>ODE (Base)</th>
<th>ODE</th>
<th>ABM</th>
<th>95% CI (±/4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>589,624</td>
<td>589,624</td>
<td>588,692</td>
<td>470.04</td>
</tr>
<tr>
<td>MR 1</td>
<td>589,624</td>
<td>297,883</td>
<td>295,380</td>
<td>332.28</td>
</tr>
<tr>
<td>MR 2</td>
<td>589,624</td>
<td>1,154,430</td>
<td>1,171,616</td>
<td>630.71</td>
</tr>
<tr>
<td>MR 3</td>
<td>589,624</td>
<td>589,593</td>
<td>582,377</td>
<td>8196.95</td>
</tr>
<tr>
<td>MR 4</td>
<td>589,624</td>
<td>589,784</td>
<td>589,372</td>
<td>509.81</td>
</tr>
<tr>
<td>MR 5</td>
<td>589,624</td>
<td>563,850</td>
<td>586,616</td>
<td>470.04</td>
</tr>
<tr>
<td>MR 6</td>
<td>589,624</td>
<td>602,297</td>
<td>588,686</td>
<td>465.18</td>
</tr>
<tr>
<td>MR 7</td>
<td>589,624</td>
<td>38</td>
<td>35</td>
<td>3.76</td>
</tr>
<tr>
<td>MR 8</td>
<td>589,624</td>
<td>969,036</td>
<td>962,209</td>
<td>690.56</td>
</tr>
<tr>
<td>MR 9</td>
<td>589,624</td>
<td>589,626</td>
<td>583,962</td>
<td>493.53</td>
</tr>
<tr>
<td>MR 10</td>
<td>589,624</td>
<td>585,195</td>
<td>592,586</td>
<td>465.24</td>
</tr>
<tr>
<td>MR 11</td>
<td>589,624</td>
<td>40</td>
<td>40</td>
<td>3.78</td>
</tr>
<tr>
<td>MR 12</td>
<td>589,624</td>
<td>963,570</td>
<td>954,263</td>
<td>599.22</td>
</tr>
<tr>
<td>MR 13</td>
<td>589,624</td>
<td>38</td>
<td>41</td>
<td>4.11</td>
</tr>
<tr>
<td>MR 14</td>
<td>589,624</td>
<td>969,036</td>
<td>960,131</td>
<td>619.68</td>
</tr>
</tbody>
</table>

* confidence intervals provided for ABM runs

### Figure 3. Cumulative deaths over time – N(Infectious) MR3 and MR4.

Next, consider the MR on original \( N(\text{Infectious}) = 1,000 \). The specific scenarios considered are:

- **MR3**: \( x' = N(\text{Infectious}) \times 0.1 = 100 \)
- **MR4**: \( x' = N(\text{Infectious}) \times 10 = 10,000 \)

An increase in the number of infections (MR3) is expected to increase the number of deaths as the number of deaths increases with any given number of infections. As such, we present the results only for \( N(\text{Infectious}) \) (figure 6). The ABM results meet the expected results for this test. Again, we can only state that the model is “possibly correct for these scenarios” and these MR did not reveal a defect.
For these three parameters, when we multiply the base values by 2, the number of dead increases by less than a factor of 2. For those scenarios at which we divide the base values by 2, the epidemic slows down and does not reach the critical mass required to spread the disease among the population (see MR 7 on figure 5). Hence, the number of dead decreases significantly stabilizing at around 40 deaths.

Finally we consider the MR on original $T(\text{Incubate}) = 3$ days. The specific scenarios considered are:

- MR9: $x' = T(\text{Incubate}) \times 0.5 = 1.5$
- MR10: $x' = T(\text{Incubate}) \times 2.0 = 6$

For these scenarios, MR9 and MR10, we expect the timeline to be compressed or extended, respectively. Similar to the case for $N(\text{Infectious})$ runs, some iterations did not see the epidemic start. In Figure 6, we plot the results of the revised ABM model. Runs of the revised code meet the expected results.

4. CONCLUSIONS AND FUTURE WORK

In this study, we investigated two disease-spread models of the 1918 Influenza pandemic. Both models have simple SIIR disease dynamics, without demographic changes, as their underlying framework. After we gained enough confidence that both models mimic the same phenomenon, we conducted preliminary metamorphic testing experiments to evaluate a) the ABM model of the 1918 flu and b) examine the effectiveness of metamorphic testing given the metamorphic relations chosen. The ABM model performed as expected for most of the tests, and for those where discrepancies were identified, MR were effective in highlighting those issues. This set of tests is insufficient to determine the validity of the ABM model.

In these experiments, metamorphic testing is a type of black box testing approach [12]. Since ABM are software systems that inherit non-linear interactions, there may not be easily identifiable exact expected results for testing. We have identified bounds within which the results should reside. The limits were stated in terms of the impact on the cumulative number of deaths and, in some cases, on the scenario timeline. We visually inspected the output to determine if the actual results met the expected results. The metamorphic relations identified in this work are not the complete suite of potential MR. They did meet our goal of determining if
they could be used to detect a defect or to illustrate whether the model appears valid for the scenario identified by the MR used.

The tests conducted cannot verify the ABM model under certain scenarios. Even though the ODE and the ABM models are validated for the original scenario, MR effectively identified the dissimilarities between the ODE and the ABM model especially under the scenarios that decrease the rate of infection and the size of the initial infectious population. Initial conditions are found as the cause of the discrepancy between the models. Furthermore, in ABM for the sake of reality, the agents can only infect integer numbers of other agents while the ODE model is continuous. So, rounding and initial conditions are other processes to investigate in creating future MR. Further analysis will be conducted on ABM model updates and more complex ABM models.

Additionally, the phenomenon of “diminishing returns” needs to be investigated. As aforementioned, P(Mortality) increases the number of dead to a certain extent, then it diminishes the rate of disease spread. At what levels the parameters change behavior can be identified with regard to other parameter values. In this work, we only modify one parameter at a time. If the parameter levels at which the simulation results are altered can be identified, multiple parameter changes and the impacts of these new scenarios can be tested [16].

Our future work includes investigating different aspects of ABM for identifying metamorphic relations (domain, individual behavior rules, implementation, algorithms, etc.). Effectiveness measures for MR other than mutation fault injection coverage will also be investigated.

5. ACKNOWLEDGMENTS

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. 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, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

6. REFERENCES

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MR (where x = Parameter)</th>
<th>Expected Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Mortality)</td>
<td>MR1: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Decreases the number of Deceased by less than $n$</td>
</tr>
<tr>
<td></td>
<td>MR2: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Increases the number of Deceased by less than $1/n$</td>
</tr>
<tr>
<td>N(Infectious)</td>
<td>MR3: $x' = x \times n$, $n &lt; 1.0$</td>
<td>In ODE, this transformation stretches out the start time; in the agent model, the start time is moved to the right.</td>
</tr>
<tr>
<td></td>
<td>MR4: $x' = x \times n$, $n &gt; 1.0$</td>
<td>In both models, this transformation compresses the epidemic spread time (the epidemic will spread faster, more quickly).</td>
</tr>
<tr>
<td>T(Die)</td>
<td>MR5: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Decreases the number of people infected, which in turn influences the number of Deceased (decreases)</td>
</tr>
<tr>
<td></td>
<td>MR6: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Increases the number of people infected, which in turn influences the number of Deceased (increases)</td>
</tr>
<tr>
<td>P(Transmission)</td>
<td>MR7: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Decreases the number of people infected, which in turn decreases the numbers of Deceased and Recovered</td>
</tr>
<tr>
<td></td>
<td>MR8: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Increases the number of people infected, which in turn increases the numbers of Deceased and Recovered</td>
</tr>
<tr>
<td>T(Incubate)</td>
<td>MR9: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Compresses the model timeline, with related changes in other parameters</td>
</tr>
<tr>
<td></td>
<td>MR10: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Stretches out model timeline, with related changes in other parameters</td>
</tr>
<tr>
<td>T(Recovery)</td>
<td>MR11: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Decreases the number of infectious (the number of cases), which in turn influences the rate of Deceased (decreases)</td>
</tr>
<tr>
<td></td>
<td>MR12: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Increases the number of infectious (the number of cases), which in turn influences the rate of Deceased (increases)</td>
</tr>
<tr>
<td>Rate(Encounter)</td>
<td>MR13: $x' = x \times n$, $n &lt; 1.0$</td>
<td>Decreases the number of cases, which in turn influences the rate of Deceased (decreases)</td>
</tr>
<tr>
<td></td>
<td>MR14: $x' = x \times n$, $n &gt; 1.0$</td>
<td>Increases the number of cases, which in turn influences the rate of Deceased (increases)</td>
</tr>
</tbody>
</table>