HAPLOS: Human Population and Location Simulator

# Abstract

Growth in human population and increasing urban densities pose numerous challenges in developing and enacting public policies and emergency processes for dealing with a broad spectrum of issues ranging from infrequent natural catastrophes, seasonal communicable diseases to routine city planning. Comprehensive analysis is necessary to identify and address problems or shortcomings with proposed policies, procedures, and emergency response strategies. Analysis of such large, complex systems is often conducted using simulation-based methods due to their advantages over traditional analytical or statistical approaches. Simulation-based analysis requires the use of high fidelity models for conducting in-depth studies of different scenarios and to ensure that crucial scalability issues do not dominate during validation of simulation results. However, generating large-scale human models from raw demographic, population geographic, and other statistical data is a challenging task. Often these models are tailored for a specific purpose and cannot be used beyond a specific application. Consequently, this research proposes the development of a Human Population and Location Simulator (HAPLOS). The primary objective of HAPLOS is to provide an effective way to create generalized human models.

# Introduction

Currently it is projected that the world’s population will reach 8.1 billion by 2025 and 9.6 billion by 2050(United Nations, Department of Economic and Social Affairs, & Population Division, 2013). As world’s population continues to grow it becomes critical that responses to significant natural or manmade catastrophes, such as evacuations and epidemic containment, be successful in order to minimize loss property or life. However, due to the growing density and size of human populations it is difficult to determine how effective these responses are due to consequences may not manifest until the response is well underway. For example, a policy to implement travel restrictions in the event of a pandemic may not help prevent the pathogen from spreading due to the difficulty in enforcing the policy in the real world (Ferguson et al., 2006). In addition, some catastrophes are rare and responses to them quickly degrade because human populations rarely remain static for any extended period of time. For example, response degradation occurs in evacuations polices which highly rely on road networks and population densities that quickly change over time.

## TYPICAL ANALYSIS METHODS

When it comes to creating responses to particular catastrophes they are often based on historical examples of similar events or by experts in the area in question. However, once candidate responses have been identified, there is often no effective approach to test them before implementing them. Statistical analysis is often conducted to assess the effectiveness of a response. However, statistical analysis only supplies aggregate information about the aftermath of a response to a catastrophe. Without information about what occurs during the response, there is no way to accurately determine the root cause of problems or shortcomings with a response strategy. The lack of comprehensive understanding of a response can result in increased costs, delays, and loss of lives. An example of such a issue was witnessed during the 2005 hurricane Katrina when it was tragically discovered the evacuation plans were not adequate to evacuate the impacted area (Daniels, 2007).

## DRAWBACKS OF CURRENT METHODS

The lack of prior testing of an evacuation plan before its use lead to people having to remain in the dangerous areas as the hurricane made landfall. On the other hand, it is clear that there was no viable way to test the policy in the real world without major disruption of daily activities and extreme cost. Moreover, aggregate statistical analysis of the evacuation plan may not uncover the root causes of flaws in plans. By only knowing the end results of the policy, researchers and policy makers would have to retroactively examine the policy manually for flaws. The task of locating these flaws is especially difficult and time consuming for an area as large as the Gulf Cost. The shortcomings of the aforementioned approaches have motivated extensive use of simulations for detailed study and analysis of policies and emergency management (X. Chen & Zhan, 2008; Ferguson et al., 2005; Hupert, Mushlin, & Callahan, 2002; Ma, Zhang, & Wang, 2013)**.**

### Modeling and Simulation – Alternative Methodology

Simulation is a imitation of a real world process or system over time (Banks, Carson, Nelson, & Nicol, 2010). Through the use of simulation, researchers and policy makers can observe exactly how the population reacts to a catastrophe, which allows for corrections to the responses to policies even before it is ever used in the real world. (Bonabeau, 2002). In addition, a simulation can be run multiple times with different responses in order to determine which one provides the best results without wasting resources or time. Due to these benefits, simulation will help design faster responses and help make them more effective once put in use in the real world. However, a simulation requires suitable underlying model in order to operate.

### Motivation: Modeling and current shortcomings

A model is usually a set of assumptions about a currently existing system in place (Banks et al., 2010) . Models provide the rules with which the simulation functions as it progresses. Thus, without a model that accurately portrays the human population, a simulation will provide inaccurate results. Typically models, especially when concerning humans, tend to be one use and only for a very specific event (Bonabeau, 2002). Single use models stem from the fact to the human populations can change rapidly. Furthermore, catastrophes affect different areas of the society with different population geographies requiring the use of highly specific models. The lack of constancies between population and catastrophes leads to the remaking of the human model every time the population changes. Although models have been successfully developed which can be easily adjusted for different populations and regions, they lack the ability to be used for multiple purposes.

Model adaptability to diverse population geography is typically accomplished by the use of easily modified data sources, parameters, and exchangeable submodels (Carley et al., 2006; Ferguson et al., 2005). These three factors have shown to allow the model to be quickly adapted to the new region and population. Through the use of these adaptable models researchers and policymakers can reuse them and save valuable time and resources when faced with a specific problem. However, if they are faced with a new problem they often have to start over completely. Significant research has been invested into modeling human populations so that the models can be used for more than one type of application (e.g. evacuations and epidemiology). Application adaptability is accomplished in a similar way to that of the population adaptability. By separating the policy from the population model will allow for greater adaptability to various applications.

## PROPOSED RESEARCH

The key component for application adaptability is the creation of a generalized human model. However, generating large-scale human models from raw demographic, population geographic, and other statistical data is a challenging task (Carley et al., 2006; J. Chen et al., 2010)**.** Consequently, this research proposes the development of a Human Population and Location Simulator (HAPLOS). The primary objective of HAPLOS is to provide an effective way to generate these types of human models. Specifically, applying HAPLOS to events that will affect both responses to epidemics and evacuations. The reason these two event types were selected is they address completely different areas of the human model, the social network and transport networks respectively. These areas of the model are often where responses attempt to change human behavior in order to get a desired reaction and are often a source of conflict among human population models that are used for policy making. In order to avoid incompatibility between models, the general human population model will be broken down in to several submodels. Each submodel will take on the same easily modified parameters and data sources. In addition, a different submodel will be added to accommodate changes to the general human model from the response. The separation of the problem from the human model will avoid incompatibility by allowing researchers to change the parameters and data sources of each model to fit the given real world human population and allow the changes needed by the response. HAPLOS will be used to test the hypothesis that the proposed modeling and simulation based methodology will be applicable to analysis of different policies as applied to different population geographies.

# Background and Literature Survey

There has been significant research conducted on modeling human populations, each with its own benefits and issues. The following subsections will summarize the approaches that were considered for HAPLOS and how the final methods were selected. This section is divided up into three subsections. The first section will discuss the different strategies that have been used to model human populations in the past and why simulation was selected over them for the problem of policy making. The next section will provide a list of criteria that was used to judge the models that will be used in the construction of the simulation. The last section will discuss the models that will be used for each submodel used in this thesis. These are as follows: the population model, the social model, the geographical model and the problem model. The problem model is further broken down into sections specifically for on epidemiological and evacuation policy models.

## Strategies for determining the effects of policies on human populations

There are multiple strategies for determining the affects of policies on human populations. These include simulation modeling, statistical analysis and real world testing. Statistical analysis is often included with simulation due to it used during the process of creating simulation model (Bonabeau, 2002). It is possible to just have a simulation based off of statistical analysis alone which is refereed to as mathematical simulations (Iwami, Takeuchi, & Liu, 2007). Real world testing will not be discussed at length, since they require extensive logistics and costs in order to accomplish them. Due to the excess cost and time real world testing not suitable.

As mentioned previously a simulation is an imitation of a real world system or process over a given period of time. (Banks et al., 2010). Simulation often coincides with an underlying model due to simulation is the process used to study the model as it evolves over time. A model is the actual representation of the underlying system(Banks, 1998). A system is a collection of objects that interact or function on their own to accomplish a purpose (Banks et al., 2010). In the case of HAPLOS the system is a human population. When discussing simulations there are specific terms used in order to describe each component of the model. A entity is the object that is often trying to be modeled by the system (Banks et al., 2010). In the case of human populations, the entities will always be humans. Entities are referred to as individual or agents as well. Every entity has properties, which are referred to as attributes. Some examples of attributes in a human population are age, gender, or current location. As the simulation progresses often these attributes can change or could stay static. The last definition refers to the actions taken by entities that span over a period of time, which are referred to as activities (Banks et al., 2010). It is often that activates are represent of the period of time where changes in attributes occur, such as when an entity would be traveling.

Statistical analysis uses mathematics in order to determine the outcome of a scenario on a population. Statistical analysis is often done though what is known as a mathematical simulation. These differ from the classical simulations by focusing only on the input and output of the simulation, not what actually determines the output (Banks, 1998). Mathematical simulations consist of a series of equations that all entities follow through out the simulation period (Iwami et al., 2007). Statistical analysis when compared to simulation requires less time to implement and to run because they are not require to run through all the internal interactions that must take place as in a real world. As mentioned earlier, due to the inability of statistical analysis to show the internal interactions it cannot provide the necessary information needed for policy development. Since statistical analysis cannot show what occurs when the population is adapting to a given policy, it is impossible to know exactly where the policy is failing without manually analysis the process.

It is clear that statistical analysis is not adequate for policy development, due to the information gap between when the policy is implemented and the policy is complete. Which is why the problem of observing a information, organizational or environmental change on a given population is a tackled through the use of simulation (Banks et al., 2010). Simulation can produce data about what is occurring during this gap, which allows researchers and policy makers to locate the sources of flaws in their policy.

## Criteria for models

There are many approaches to choose from when modeling a real world system, however some provide more accurate results for modeling human populations than others. In order to select models for HAPLOS a standard set of criteria have been selected to determine a suitable modeling approach. Criteria for model selection will be based off three elements: realism, computational cost, and ease of calibration. The reason these criteria were selected is due common problems encountered when developing models and can lead to poor results and lack of reusability.

### REALISM

Realism is how well the model mimics the real world system. It is important that a model is as realistic as possible to allow elements in the model to map back to real world counterparts. Realism is key in order to transfer any results found during the simulation process to the real world. In order for the model to be as realistic as possible, the elements of the model must match both the behavior and attributes of the real world counterparts. Specifically the attributes of the elements must be statistically similar to those of the real world elements and interactions between entities should be realistic. For example, a realistic model must ensure that the age distribution of the simulation population statistically matches the age distribution of the existing population.

### CALIBRATION

In order for a model to remain accurate, the model must be amenable to calibration in order to match real world data. Without the ability to calibrate the model it would be impossible to maintain realism of the model over long-term use. Human societies can change rapidly with the introduction of newer technology and transportation systems, which in turn must be reflected in the model. Changes in real world populations are not just limited to long-term changes because distributions of sub-populations can significantly change from year to year. Without calibration a model can quickly become outdated. In addition, models that are easily calibrated enable them to be repurposed for different data sources and different uses. Quick adaptation of models is frequently accomplished by constructing them modularly.

### COMPUTATIONAL COSTS

The last criterion for model selection is the computational cost of the model. Computational cost often finds itself in opposition with the realism requirement (X. Chen & Zhan, 2008). If the model is extremely realistic it is often requires many more calculations, which in turn increases the computational cost. Thus if a model is very computationally costly it can lead to the model not being feasible to use due to time constraints or computing resources needed to run the simulation. Therefor a balance between realism and computational cost must be met.

### SUMMARY OF CRITERIA FOR HAPLOS

It is clear that a balance between the four criteria will need to be met, but at the same time each must be maximized in order to enable effective use of HAPLOS. Without realistic models, HAPLOS’s results would not be able to be mapped back to the real world. If HAPLOS cannot be calibrated for a different region or population, then the system will quickly become outdated. Finally if HAPLOS requires too much computational cost it would become inefficient to be used on larger populations. Thus each one of the four criteria must be considered when selecting the final models.

## Models

The human population model for HAPLOS will be divided into submodels model as shown in Figure 1. HAPLOS will be using three submodels in order to simulate a human population and a policy that will be tested. The submodels used to simulate the human population are the population model, social model and geographical model. The population model will be responsible for the creation of the synthetic population of entities and assigning their attributes. The social model will handle how each entity will behave when in contact or not in contact with other entities. The last sub model is the geographical model, which will determine how the entities will move in the region. There will be two different policy models used to test HAPLOS. These polices are for evacuations and epidemic containment.

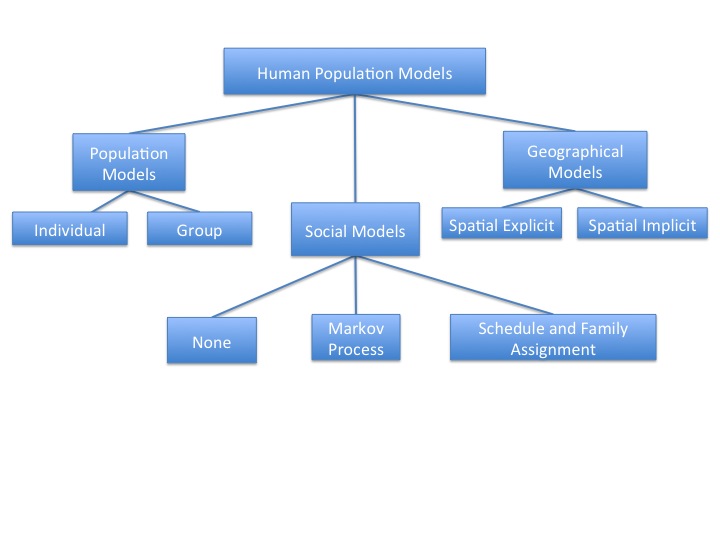


Figure 1: Diagram of the parts of a human population model and the variations of each submodel.

### Population Model

The population models are specific to the creation of a synthetic population. The creation of the synthetic population involves statistical analysis of real world data and creating a population with similar demographics. A major decision to make with population models is deciding what an entity will represent. An entity can represent either an individual or a group. Both can be used to model human populations since humans can be represented as a collective groups (e.g. Households) or a single person.

Group models represent a collection of members as an entity. An example of group modeling is having entities represent an entire farm instead of single cow in order to observe the spread of Foot and Mouth Disease through a area (Keeling et al., 2001). The benefit of using a group model is it reduces the computational cost because there are a smaller number of entities. By having fewer entities it reduces the number of computations that must be preformed at each time step. However, the reduction of entities comes at the cost of loosing resolution of the data produced. Information about each individual member of the group is lost and this may be important for certain policies. As with the farm example, only information about how many farms that are infected is available. Lack of more detail information can lead to an inaccurate estimation of livestock that is infected and can lead to a misallocation of resources in response to the outbreak.

Individual models will model every member as an entity. Individual models are often used for modeling the spread of disease from person to person in a specific area or on a global scale (Edlund, Davis, & Kaufman, 2010; Epstein, 2009). Modeling disease spread on an individual basis allows for more accurate data on how the disease spreads and about how well preventive techniques are at minimizing the spread (Nsoesie, Beckman, & Marathe, 2012). In addition, by modeling on the individual level it allows the population to be heterogenetic which is often more representative of real life human populations (Smieszek et al., 2011).

### Social Model

The social model is responsible for determining how entities in the system will interact and is referred to as a social network. The social models are often left out because the geographical models can provide a simplistic version of a social model. Using a geographical model instead of a social model is often done in systems that deal with very large-scale populations. An example of such a system is STEM which involves the population of the entire world (Edlund et al., 2010). However, without implementing a social model, interactions between entities may not be realistic. For example, in STEM, unrealistic interactions are caused by entities modeled to have equal probability of coming into contact with each other in geographical model. However, in reality entities do not have equal probability of interacting with each other. An example of entities having unequal chances of interacting is when two people work in different buildings on different sides of town. It is very unlikely that they will cross paths during any of their day-to-day activities. Polices highly reliant on social interaction can be skewed by unrealistic interactions between entities(J. Chen et al., 2010). In order to minimize unrealistic interaction it is important HAPLOS have a social model. There are currently two ways of modeling social model; Markov processes and schedules and family assignment.

#### Markov Processes for Social Models

Markov processes are dependent on the Markov property. The Markov property states that the next future state of a entity is solely based on the entity’s present state (Lerman & Galstyan, 2003). The major benefit of these models is there is no need to keep track of an entity’s history in order to determine its next action within the system. However, it has been shown that Markov processes are not sufficient to simulate complex social systems (Carley et al., 2006). They are unable to simulate complex social because the Markov assumption does not always apply due to human interactions. For example, a person who is currently at work can choose to go home or to a store. By the Markov process both of these could potently occur. However, lets say that this person had already gone to the store before coming to work. Normally a person would have a less likely chance to return to the store that day. Due to interactions relying on what was done prior means that interactions of entities must be planned out in advance.

#### Schedule and Family Assignment for Social Models

The schedule and family assignment models gives each entity a family assignment and daily schedule to preform through out the course of a given time frame (J. Chen et al., 2010; Nsoesie et al., 2012; Smieszek et al., 2011). These schedules are usually based off of data collected on the most common daily schedules seen in human populations. These schedules are then assigned either based off of percentages or demographics information. Assigning these schedules allows an entity to model realistic interactions and doesn’t require maintaining history of the entity’s interactions. There is no need to keep a history unlike the Markov processes due to their locations are predetermined. However, these schedules are only possible if there is data about the types of schedules used in the population.

Without information on the types of schedules the interactions will not be realistic. In reality, humans do not always follow a set schedule. One solution to the lack of a dynamic schedule is schedules can be set for the whole week period or more to account for changes during shorter periods of time. However setting the schedule to go over longer periods of time still doesn't take into account that people take vacations or have life transition (e.g. going from being a student to being an employee). These life events are not as predictable and thus are cannot be planned for ahead without losing realism. However, these types of life events can be still incorporated for longer term simulations by allowing for a random or set change of the daily schedules for either a short period of time or permanently.

### Geographical Model

The geographical model is responsible for determining how entities can move through out the region. The geographical models serve to move the entity to the location where the social model dictates, which helps to maintain realistic interactions even when the entity is moving. It is often the case that a network model is used for creation of the geographical model (Carley et al., 2006; Eubank et al., 2004; Smieszek et al., 2011). A network model consists of a graph with weighted edges, where each edge represents a valid path to travel between the two locations (Eubank et al., 2004). The main difference between geographical models implementations is if they keep track of the physical location entity or do not. These are referred to as spatially explicit and spatially implicit respectively.

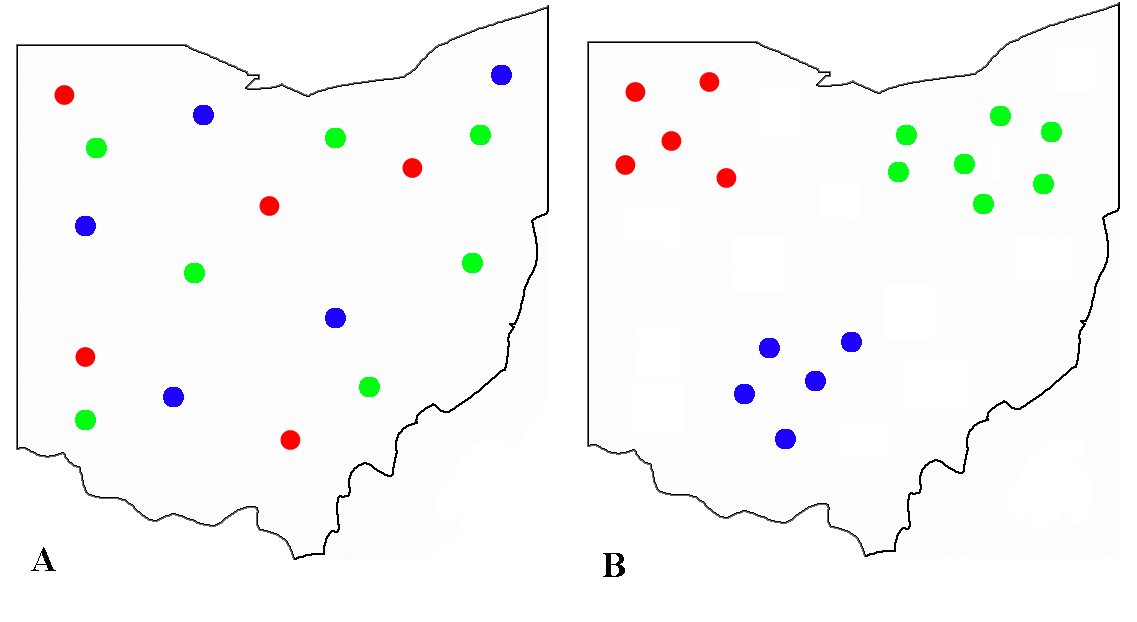


Figure 2: For both figures, dots of the same color represent entities that are in the same contact network. Figure A shows what a spatially implicit model contact networks would look like, well Figure B shows what spatially explicit model contact networks would look like.

A spatially implicit model is where each entity interacts with a set number of other entities, but they may not be collocated in the same area (Bian, 2013). An example of spatially implicit models an been seen in the Figure 2 A above. Spatially implicit models are often used when interaction between entities is high, such as groups of friends. Spatially implicit models rely on the heterogeneous mixing of the population. If the population is not heterogeneous then interactions will no longer be unique. Lack of unique interactions can lead to a sub population being walled off from the rest of the population, which in many modern human populations is impossible to achieve. The main benefit of a spatially implicit model is it helps reduce the computational overhead due to it is does not need to keep track of where each entity is current. In spatially implicit models the interactions solely relying on the social networks and not their physical location in the system. However, the lack of keeping track of their physical location can lead to problems when dealing with populations with a high transportation rate. As entities move their social network will alter, which often occurs as people commute from one location to another throughout the day (Bian, 2013).

The spatially explicit models can handle these cases as they keep track of the actual location of where the entity is, thus allowing for the creation of a dynamic social network. Figure B above shows an example of how entities are located in a spatially explicit model. By using a spatially explicit model it can easily be used to model an actual existing location using transportation and population density data (Ferguson et al., 2005). Spatially implicit models are unable to model existing locations due to they solely focuses on the social networks and not the actual location of entities.

### Policy Models

The main goal for HAPLOS is to provide a human population model for use in policy development. It is important that possible polices be explored in order to confirm the hypothesis that HAPLOS can be used for a range of different polices without having to modify the human population submodel. Two polices that will be explored are for use in evacuation and epidemic containment. These polices were selected because they must modify the human population model in different areas in order to successfully simulate the implementation of the policy. Modifying different areas of the human model will confirm the versatility of HAPLOS on different polices.

#### Epidemiologic Models

Epidemic containment models alter the social model of the human population. The modification of the social model is often done to mimic the spread of a pathogen in a population. In order to model the spread of a pathogen a compartmental model, cellular automata model or a network model is often used(Bian, 2013; Carley et al., 2006). The simplest of these models is the compartmental models.

Compartmental models are also referred to by the stages that each entity can go through during simulation, which are SIR (Susceptible, Infected, Recovered) or SEIR (Susceptible, Exposed, Infected, Recovered). Based on whether the individual is near an infected individual, a set of ordinary differential equations will determine when the individual moves from one stage to another(Kermack & McKendrick, 1991). However, compartmental models highly rely on the homogeneous mixing of the population in order to simulate an outbreak accurately. In addition, they are very deterministic, which means they can only provide averages and not show extreme cases (Keeling, 2005). Compartmental models lack any built in connection to a spatial environment and the actual contact network. Due to these problems they are often used in combination with other methods, such as network or cellular automata.

Cellular automata models consist of individual finite state machines that are connected together, usually in a grid(Epstein, 2004). Each of these finite state machines will change state based on the states of the neighboring machines (Kari, 2005). Cellular automata models are often coupled with the compartmental models because they provide a both a spatial and contact structure that compartmental models lack. In addition, compartmental models provide the distinct states required by the cellular automata model (Epstein, 2004). Combining cellular automata with compartmental models allows them to be used with heterogeneous populations, which normal for the majority of human populations (Bian, 2013). However, cellular automata models are often setup in a 2D grid and therefore can lead to a lack of realistic interactions between entities and have a hard time being fit to a real world space (Carley et al., 2006). Unrealistic interactions can lead into inaccurate measure of how quickly a pathogen can spread through a large environment.

A network model is one way to avoid unrealistic interactions. A network model differs from the cellular automata model by using a graph structure instead of a 2D grid. An example of such systems are the EpiSims and BioWar (Carley et al., 2006; Eubank et al., 2004). The graph structure allows for a more accurate representation of a geographical area due often it is hard to fit geographical areas to a grid. In addition, graph structure allows for more realistic interactions between entities by allowing for more entities to be able to interact with each other if they are in a close proximity.

#### Evacuation Models

Evacuation model will often alter transportation network, so it is highly reliant on how the geographical model is setup in the human population. Many of the examples of evacuation simulations tend to involve the use of cellular automata or a network model (Duanmu, Taaffe, Chowdhury, & Robinson, 2012; Ma et al., 2013). Both of these types of models are very similar to that of the epidemiology models, however they are applied to the geographical model instead of the social model.

Cellular automata models are often used as an easy way to break up a region (X. Chen & Zhan, 2008). As previously mentioned, breaking up a region into a grid can pose a problem because not all regions can be broken down into a grid format(Carley et al., 2006). In addition, the finite state machines of the cellular automata model can ignore factors that could potentially affect the evacuation route an individual may take. An example of such an affect is often during evacuations people will return home to collect items or collect other members of their family before they evacuate the area. Events that cause mass evacuations are often extremely time restricted (e.g. hurricanes, wildfires, etc.) and these detours can prove costly if they are not taken into account in the evacuation policy. Due to these factors, it is clear that network models provide a more realistic representation of evacuation scenarios.

Most network models used for evacuation simulations are based around a group population model. These group population models are based around the number of vehicles that are able to get out in time (X. Chen & Zhan, 2008; Duanmu et al., 2012). The issue with using a group model is it can give an inaccurate estimate of the actual number of people successfully evacuated from the area. During an evacuation transportation out of the area is done on a family level. Families tend to evacuate together which means they are more likely to use one vehicle and thus the number of people per car is normally higher than normal. One of the factors used to judge evacuation policies is how many people are successfully evacuated, so without an accurate judge of how many people have evacuated a possible policy may be overlooked. So it is key to still study polices at the individual level.

## Summary

In order to maximize the realism and calibration while minimizing computational costs, HAPLOS will select specific set of submodels. However, the more realistic a model is the higher the computational costs thus more importance must be place on some criteria than others. With policy development it is critical that they be as realistic as possible and be able to be calibrated easily. For the population model an individual model will be used in order to maintain the detail needed by many polices. The social model will be implemented as a schedule and family assignment in order to maintain realistic interactions between entities. The geographical model will be implemented as an explicit network model in order to maintain realistic movement of the entities. Each of these submodels will be then be modified by an additional policy model. Of the epidemic models examined a combination of a network model and a compartmental model will be used. A network model will be used for the evacuation model. The main reason for these models to be selected is they maintain the most realistic result.

# Proposed Research

The goal of this research is to create an effective way to generate generalized human based models in order to address the problem of having to recreate a human population model every time one is needed to test a protocol or policy. HAPLOS will address the problem of having to remake the human population model by using multiple submodels to allow for easy adjustment of the human model for a new purpose. HAPLOS will be made up of a population, social and geographical submodels as was discussed in the background section. Once HAPLOS is completed it will then be used in combination with additional policy models in order to test event responses on a variety of different populations.

## Models

### Population Model

The population model of HAPLOS will consist of an individual based model. Each individual will consist of a person from the simulated population. The number of individuals to be created will be based off of the latest counts from the U.S. Census (U.S. Census Bureau, 2011a). Only age and gender will initially be set for each person in the population in order to keep the model as general as possible and these are needed in order to set the schedule and family assignments. Age will consist of an integer value with a range from 0 to 100, with 0 representing a child under the age of 1. Gender will consist of a binary value representing male or female. Other attributes can be added through the policy model if they are needed. Each individual will be assigned a age and gender based on the distributions retrieved from the latest U.S. Census data (U.S. Census Bureau, 2011a).

### Social Model

The social model of HAPLOS will consist of a schedule and family assignment model. Families will be created and then each member of the family will receive a schedule based on their age and family makeup. Schedules will consist of the location and the start and end time the individual will be at that location before moving to the next location. For simplicity locations will be limited to homes, daycare, schools, businesses, and medical facilities. In addition to spending time at a location time for transportation time will be included in the schedule.

#### Family Assignment

Family assignment will be done so in order to approximate the family sizes and compositions based on U.S Census Data (U.S. Census Bureau, 2011a). Each individual will be assigned a family, which will consist of one to seven individuals. In order to maintain realism, the only regulation on family assignment is that individuals who are age 17 or younger must be assigned to a family with person who is 18 years or older.

#### Schedule Templates

Schedule templates serve as a way to help create realistic schedules for common lifestyles of a particular age group. Age groups selected for HAPLOS are young children (0-4 years old), school age children (5-17 years old), and adults (18-100 years old). Template schedules will span for an entire week and then repeat. In order to allow for changes in schedules due to life events, schedules will be altered temporarily or permanently at set times (e.g. when a person turns 18) or randomly (e.g. person goes on vacation). If there is a child in the family age 13 or younger and not at a school or daycare location, it will be assumed an adult from their family must be present with them and the adult cannot take the child to work.

##### Young Child template (0-4 years of age)

If an individual is age 4 and under they will be assigned the same schedule as an adult in their family provided they are not assigned to attend work. If no adult is available an adult will be assigned to take the individual to a daycare and then leave that location for work. The individual will be assigned to the closet daycare to their home location. If that daycare is at capacity then the next nearest will be selected. Capacity of a daycare facility capacity will range from 10-30 children following a exponential distribution. When the first adult in the family is no longer at work, they will then retrieve the individual from the daycare. The individual will then follow the adult’s schedule till it is no longer possible and must return to daycare or will be passed off to another adult in the family.

##### School Age Child Template (5-17 Years of Age)

If the individual is between the ages of 5-17 they will be assigned to a school. The schools they will be assigned is based on their age and the nearest school to their home that provides the proper grade level. If the nearest school to their home is at capacity the next school will be selected. Capacity and grades available at each school will be determine from the total enrollment information from the Common Core of Data (CCD) and Private School Survey (PSS) for the 2010-2011 school year (National Center for Education Statistics, 2011).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade | Lunch | Recess | Number of Classrooms | Length of Day |
| Kindergarten | Yes | Yes | 1 | 4 hours |
| 1-5 | Yes | Yes | 1 | 8 hours |
| 6-12 | Yes | No | 7 | 8 hours |

Table 1: Summary of school time schedule restrictions for specific grades.

The individual will be assigned a series of room numbers in order to mimic actual movement through the school during a typical day. For attending those 1st through 5rd grade they will be restricted to a 4 room schedule (representing their classroom, lunch, and recess). Time spent in lunch, and recess will all last 60 minutes each, while the remaining time will be spent in the classrooms. Students from 6th-12th grade will be assigned a 7-9 room schedule (representing the common use of departmentalized classrooms). If the individual is of age 5 (kindergarten) they will be assigned a half school day for the school they are attending. Half will be assigned a morning half day and the other half will be assigned a afternoon half day. The kindergarten schedules will consist of a three room schedule (representing recess, classroom and lunch). Lunch will last for 60 minute time period. After school they will return home for the rest of the day.

Each room assigned an equal amount of time during the school day. The distribution of rooms assigned will follow a discrete expediential distribution with 7 rooms being the highest probability. The number of individuals assigned to a classroom will match a discrete normal distribution with the mean set at the average pupil/teacher ratio for the school (retrieved from the CCD and PSS datasets) and range of 10 to 30. During the weekday they will attend the school for a set number of hours based on the hours provided from CCD and PSS datasets. If school day hours are not listed it will be assumed to be a 8 hour school day.

|  |  |  |  |
| --- | --- | --- | --- |
| Age Range | Mimic Adult | Radius of Travel | Curfew |
| 5-13 | Yes | N/A | N/A |
| 14-15 | No | 2 miles from home | 5:00PM |
| 16-17 | No | 10 miles from home | 8:00PM |

Table 2: Summary of free time restrictions for specific age ranges of school age children.

What is done during times not reserved for school will be dependent on the age of the individual. If they are under the age of 14 they will be restricted to following a schedule of an older member of their family. If the individual is age 14-16 they will be restricted to visiting other locations with in a 2 mile radius of their home and required to return home by 5:00PM. Those over the age of 16 will be allowed to travel within a 10 mile radius of their home and will be required to return home by 8:00PM.

##### Adults (18-100 years of age)

###### Working Adult Template

If an individual is above the age of 18 they will be assigned a location of employment. The location will be determined based on closet location to there home. All locations other than homes will be allowed to have employees. If the location closest to their home is at capacity then the next nearest location will be selected. The capacities for locations will be based off of data from the distribution of number of employees per establishment provided by the 2011 Country Business Patterns (U.S. Census Bureau, 2011b). The number of hours of work per week assigned will range from 1 to 60 hours. The distribution of hours will approximate the distribution and average provided by the “Person at Work by Hours Worked” table from the most recent U.S. Census (U.S. Census Bureau, 2011a). Based on the number of hours per week assigned, they will be assigned a number of days to work such that it minimizes the total number of days they will have to work. If a person is working more than 40 hours in a week a maximum number of hours per day will be 10 hours, otherwise the maximum number of hours per day will be 8 hours.

During periods where they are not working, they will be assigned to visit other locations in the area or return home. The amount of time spent at other locations will be determined from the latest Time Usage Survey from the Bureau of Labor Statistics (U.S. Bureau of Labor Statistics, 2012). The capacity for visitors for businesses will range between 1-100 depending on the number of employees in a given hour. Schools, daycares and homes will be restricted to only 1-20 visitors in a given hour. Medical facilities capacities will be allowed 1-20 visitors in a given hour.

###### Non-Working Adult Template

The non-working adult template is meant for adults who are not employed or those who are retired. This differs from the working adult template by removing the need of having to be at a job. Thus their schedules will be much more random in terms of composition. Time spent out of the home will be calculated from the latest Time Usage Survey from the Bureau of Labor Statistics (U.S. Bureau of Labor Statistics, 2012) for each age group. Based on distributions calculated the amount of time spent at other homes or other locations will be assigned to each person. The time spent at each of these locations will then be distributed through out the week randomly.

For those under at or under the age of 64 the unemployment percentage of individuals assigned this schedule will be used from the “Unemployed Workers -Summary” table form the latest U.S Census (U.S. Census Bureau, 2011a). Those who are older than 64 a percentage of individuals will be assigned this schedule will match with the percentage of retired individuals of that age group. The percentage will be calculated based on the number of non-working individuals in a given age group divided by the total number of people in the age group from the most recent American Community Survey (U.S. Census Bureau, 2012).

### Geographical Model

The geographical model will consist of spatially explicit network model. Movement of individuals will be done so in a way to approximate the distribution the transportation methods used for moving to a specific location provided by the National Household Travel Survey (U.S Department of Transportation & Federal Highway Administration, 2009). Transportation methods available will change based on the population density in the area due to cities often will have more public transport than rural areas. In addition, transportation methods will have maximum capacity allowed before it will be unable to take more individuals due to traffic congestion.

#### Buildings

Public buildings will be divided up into four sub-types; daycare, schools, businesses, and medical facilities. Daycares, medical facilities and businesses will be placed based on population density from either data from SEDAC or LandScan (Bright, Coleman, Rose, & Urban, 2012; Center for International Earth Science Information Network (CIESIN)/Columbia University, 2005). The number of business will be created based on information about the number of establishments for each given type from the 2007 Economic Census (U.S. Census Bureau, 2007). Schools will be created in order to match the geolocation of actual schools in the U.S using data collected from Common Core of Data (CCD) and Private School Survey (PSS) for the 2010-2011 school year (National Center for Education Statistics, 2011)

The number of homes created and types of housing (single or multi-family) will be created based the totals provided by the U.S Census Data (U.S. Census Bureau, 2011a). The maximum number of family’s allowed in a multi-family home will range from 2-100 families and will follow a discrete expediential distribution. Each home will be placed based on population density with higher capacity multi-family homes being more frequent in denser areas. Each family will be assigned a home number, which will correspond to a home. The home number will range between one and the number of homes created for the simulation. More than one family can be assigned to a home if the home is considered to be a multi-family home, otherwise only one family will be assigned to a home.

### Policy Model

The policy model for HAPLOS will consist of modifications to the population, social and geographical submodels. The modifications allowed for the population model will the adding additional attributes for to be tracked and updated for each individual. As with the case of a policy model to be used for pandemic response testing an attribute such as infection status would need to be added to the population model. Modifications to the social model will include forced modification to schedules and family assignments. An example of a modification to he social model is during a evacuation schedules would have to be altered in order to have the in order to mimic a person evacuating instead of a normal day to day life. The last submodel to be modified is the geographical model. Modifications can be shutting down of transport methods or buildings, restricting people to their homes, and changes in the capacity of transportation methods. An example of such a modification is the shutting down of schools in response to a pandemic.

In order to test HAPLOS effectively a policy models must be applied. The policy models that will be constructed will be one to simulate a pandemic spread and response though a population. The pandemic spread will use a SEIR model that will mimic a influenza outbreak. Values for the SEIR model will be the same as those of used for Nsoesie et al. for the simulation of a outbreak of H1N1 in 2009 (Nsoesie et al., 2012).

## Development and Prototype

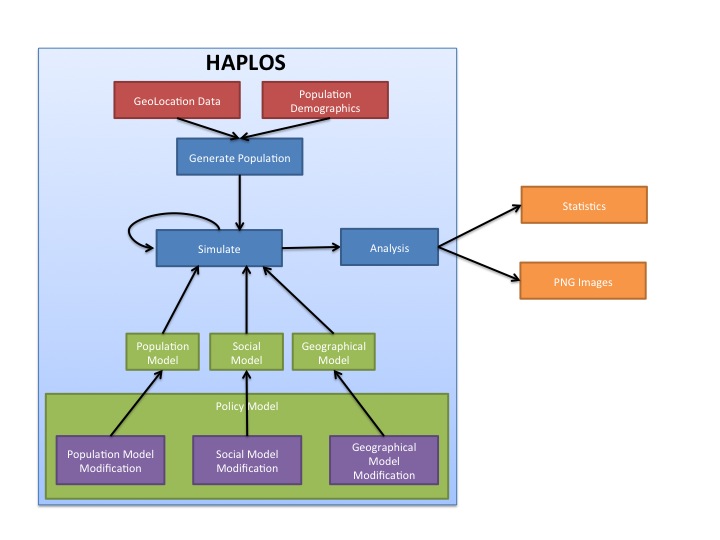
### Statistical Analysis

Statistical analysis will be done using the software R. R is a open source software that can be used to create scripts which can then be used to process large amounts of data quickly. In addition, these scripts can subsequently be reused to process data as it is updated allowing HAPLOS to be updated efficiently. Validation and verification will also be preformed using R.

### Software

The development of HAPLOS will be done in C++ due to the efficiency will be needed in order to handle the number entities. It is key that due to the number of entities that will be needed a server cluster will be needed to insure proper testing of the system. Due to the computational power needed the Redhawk Cluster will be used in order to run the HAPLOS. Results of each simulation will be summarized in both text output and in snapshots of the population provided in PNG format. These results will be used to confirm that the population being generated matches the real world population as closely as possible.

HAPLOS will need to be easily modified by researchers for their needs, it will have to be structured in an object oriented way. As seen in Figure 4, the main HAPLOS system will consist of two main methods, generating population and update population. The generating population method will take in the geo-location data and the demographic data to create a synthetic population based on those parameters. The generated population will be passed to the update population method. The update method will consist of three calls to the three models (population, social and geographical) that are needed to update the population to the next time step. These submodels will call the researcher’s event model, which contains the modification models for each of the submodels in HAPLOS. The modification models will be responsible for modifying the original HAPLOS models to fit the event taking place.

Figure 4: Dataflow diagram of HAPLOS system.

### Prototype

Currently a simple prototype of HAPLOS generates a synthetic population with gender and location to match that of the most recent data from the United States Census and using SEDAC geo-location data(Center for International Earth Science Information Network (CIESIN)/Columbia University, 2005; U.S. Census Bureau, 2011a). A snapshot can be taken at anytime during the simulation as can be seen in Figure 3 for any variable about the population (e.g. population density). In addition basic statistics about the gender and location are generated as a text output.

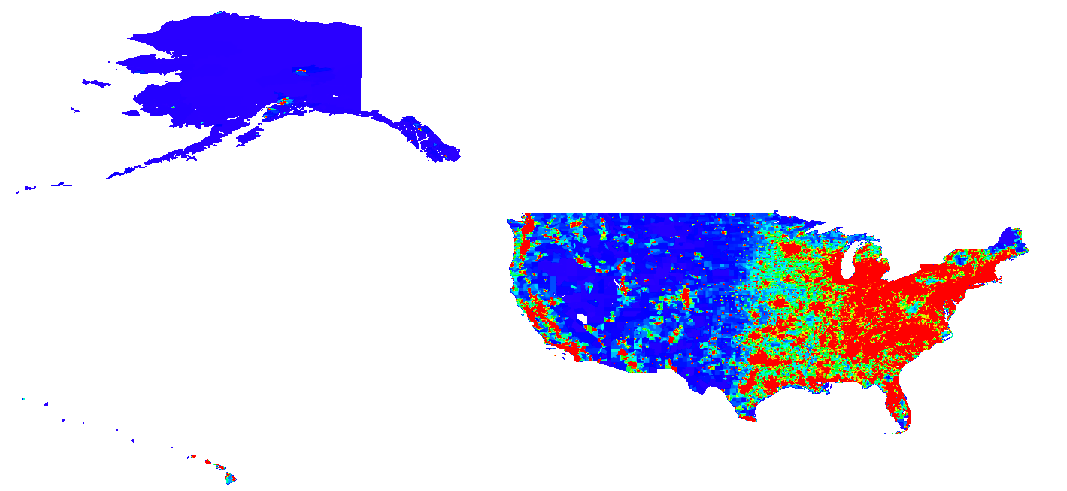


Figure 3: HAPLOS prototype using SEDAC population density from 2000 for the basis of generating a synthetic population with the same density. Red areas represent areas that are highly populated well blue areas represent lightly populated areas.

## Indication of Research Completion

In order for the proposed research to be completed, HAPLOS must be able to generate a realistic human population with or without a policy model being introduced. The human population model will consist of a fully completed population, social and geographical models as described in the Models section of the Proposed Research. As realism is a key requirement, realism will be determined by how well the simulated population matches the existing population in all attributes and distributions. In the case of HAPLOS these attributes are age and gender; distributions are transportation methods used in a year, locations visited in a year, and household composition. The acceptable difference between the simulated population and the existing population for any attribute is ± 5%.

When the policy models are added, since the same SEIR model that was used in the Nsoesie et al, HAPLOS should produce similar disease propagation results. A difference of ±10 % will be considered acceptable. Once the SEIR model has been confirmed an actual policy will be injected into the policy model. These policies will be household quarantine and anti-viral usage. A similar reduction in disease spread should be expected as that of other models used to test the same strategies (Ferguson et al., 2005, 2006). An difference in reduction of ± 5% will be consider acceptable.

# Timeline

(To Be Added)

# Summary

Testing of policies and procedures has become more critical as populations continue to grow and it becomes more difficult to locate problems prior to implementation. Locating problems before implementation can help save resources and lives, especially in developing countries where the populations have increased significantly (United Nations et al., 2013). One of the best ways to test polices and procedures is through simulation. However, the main drawback of simulation is that the human model must be tailored to a specific policy or procedure. HAPLOS will address the problem of having to remake the human model by providing a way to generate a generalized human population that can be used for testing different policies or procedures. HAPLOS will thus save time of researchers and policies makers by allowing them to focus less time on the creation of the human model and more on testing of the policy or procedure. The more time allowed for testing the better policies and procedures will be able to respond to changes in populations, even if the policy or procedure is not used for sometime.

# Work Cited

American Hospital Association. (2012). AHA Annual Survey Database.

Banks, J. (1998). Principles of Simulation. In J. Banks (Ed.), *Handbook of Simulation* (pp. 3–30). New York, NY, USA: John Wiley & Sons Inc.

Banks, J., Carson, J., Nelson, B., & Nicol, D. (2010). *Discrete-Event System Simulation* (Fith.). Upper Saddle River, New Jersey, USA: Pearson Education Inc.

Bian, L. (2013). Spatial Approaches to Modeling Dispersion of Communicable Diseases &ndash; A Review. *Transactions in GIS*, *17*(1), 1–17. doi:10.1111/j.1467-9671.2012.01329.x

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(Suppl 3), 7280–7287. doi:10.1073/pnas.082080899

Bright, E. A., Coleman, P. R., Rose, A. N., & Urban, M. L. (2012). LandScan 2011. Oak Ridge, TN: Oak Ridge National Laboratory SE  - July 1, 2012. Retrieved from http://www.ornl.gov/landscan/

Carley, K. M., Fridsma, D. B., Casman, E., Yahja, A., Altman, N., Chen, L.-C., … Nave, D. (2006). BioWar: scalable agent-based model of bioattacks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *36*(2), 252–265. doi:10.1109/TSMCA.2005.851291

Center for International Earth Science Information Network (CIESIN)/Columbia University, and C. I. de A. T. (CIAT). (2005). ridded Population of the World, Version 3 (GPWv3): Population Density Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Retrieved from http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-density

Chen, J., Huang, F., Khan, M., Marathe, M., Stretz, P., & Xia, H. (2010). The effect of demographic and spatial variability on epidemics: A comparison between Beijing, Delhi, and Los Angeles. *Critical Infrastructure (CRIS), 2010 5th International Conference on*. doi:10.1109/CRIS.2010.5617584

Chen, X., & Zhan, F. B. (2008). Agent-Based Modelling and Simulation of Urban Evacuation: Relative Effectiveness of Simultaneous and Staged Evacuation Strategies. *The Journal of the Operational Research Society*, *59*(1), 25–33. doi:10.2307/4622879

Daniels, R. S. (2007). Revitalizing Emergency Management after Katrina. *Public Manager*, *36*(3), 16–20. Retrieved from https://proxy.lib.muohio.edu/login?source=ebsco&url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=27170373&site=eds-live

Duanmu, J., Taaffe, K. M., Chowdhury, M., & Robinson, R. M. (2012). Simulation analysis for evacuation under congested traffic scenarios: a case study. *SIMULATION*, *88*(11), 1379–1389. doi:10.1177/0037549712454688

Edlund, S. B., Davis, M. A., & Kaufman, J. H. (2010). The spatiotemporal epidemiological modeler. In *Proceedings of the 1st ACM International Health Informatics Symposium* (pp. 817–820). New York, NY, USA: ACM. doi:10.1145/1882992.1883115

Epstein, J. M. (2004). *Toward a containment strategy for smallpox bioterror : an individual-based computational approach / Joshua M. Epstein ... [et al.]*. Washington, D.C. : Brookings Institution Press, c2004. Retrieved from https://proxy.lib.muohio.edu/login?source=ebsco&url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00344a&AN=mucat.b3108161&site=eds-live

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature*, *460*(7256), 687. doi:10.1038/460687a

Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V, Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*(6988), 180–184. Retrieved from http://dx.doi.org/10.1038/nature02541

Ferguson, N. M., Cummings, D. a T., Fraser, C., Cajka, J. C., Cooley, P. C., & Burke, D. S. (2006). Strategies for mitigating an influenza pandemic. *Nature*, *442*(7101), 448–52. doi:10.1038/nature04795

Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., … Burke, D. S. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, *437*(7056), 209–14. doi:10.1038/nature04017

Hupert, N., Mushlin, A. I., & Callahan, M. A. (2002). Modeling the Public Health Response to Bioterrorism: Using Discrete Event Simulation to Design Antibiotic Distribution Centers. *Medical Decision Making*, *22*(s5), S17–S25. doi:10.1177/027298902237709

Iwami, S., Takeuchi, Y., & Liu, X. (2007). Avian&ndash;human influenza epidemic model. *Mathematical Biosciences*, *207*(1), 1–25. doi:10.1016/j.mbs.2006.08.001

Kari, J. (2005). Theory of cellular automata: A survey. *Theoretical Computer Science*, *334*(1-3), 3–33. doi:10.1016/j.tcs.2004.11.021

Keeling, M. J. (2005). Models of Foot-and-Mouth Disease. *Proceedings: Biological Sciences*, *272*(1569), 1195–1202. doi:10.2307/30047668

Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., … Grenfell, B. T. (2001). Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*, *294*(5543), 813–817. doi:10.2307/3085067

Kermack, W. O., & McKendrick, A. G. (1991). Contributions to the mathematical theory of epidemics - I. *Bulletin of Mathematical Biology*, *53*(1-2), 33–55. doi:10. 1016/S0092-8240(05)80040-0

Lerman, K., & Galstyan, A. (2003). Agent memory and adaptation in multi-agent systems. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems* (pp. 797–803). New York, NY, USA: ACM. doi:10.1145/860575.860703

Ma, Y., Zhang, F., & Wang, J. (2013). A study on simulation and optimization for personnel and vehicle coordination evacuation in case of emergency. *Journal of Theoretical & Applied Information Technology*, *49*(2), 864–870. Retrieved from https://proxy.lib.muohio.edu/login?source=ebsco&url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=87526003&site=eds-live

National Center for Education Statistics. (2011). Common Core of Data and Private School Survey. Retrieved from http://nces.ed.gov/ccd/elsi/

Nsoesie, E. O., Beckman, R. J., & Marathe, M. V. (2012). Sensitivity Analysis of an Individual-Based Model for Simulation of Influenza Epidemics. (A. Vespignani, Ed.)*PLoS ONE*, *7*(10), e45414. doi:10.1371/journal.pone.0045414

Smieszek, T., Balmer, M., Hattendorf, J., Axhausen, K. W., Zinsstag, J., & Scholz, R. W. (2011). Reconstructing the 2003/2004 H3N2 influenza epidemic in Switzerland with a spatially explicit, individual-based model. *BMC Infectious Diseases*, *11*, 115. doi:10.1186/1471-2334-11-115

U.S Department of Transportation, & Federal Highway Administration. (2009). National Household Travel Survey. Retrieved from http://nhts.ornl.gov

U.S. Bureau of Labor Statistics. (2012). American Time Use Survey.

U.S. Census Bureau. (2007). 2007 Economic Census. Retrieved from http://www.census.gov/econ/census07/

U.S. Census Bureau. (2011a). Statistical Abstract of the United States: 2012. Washington, DC. Retrieved from http://www.census.gov/compendia/statab/

U.S. Census Bureau. (2011b). 2011 County Business Patterns. Retrieved from http://www.census.gov/econ/cbp/

U.S. Census Bureau. (2012). American Community Survey. Retrieved from http://www.census.gov/acs/www/

United Nations, Department of Economic and Social Affairs, & Population Division. (2013). World Population Prospects: The 2012 Revision, Highlights and Advance Tables. Retrieved from http://esa.un.org/wpp/Documentation/publications.htm