VI. DESCRIPTION OF AGENT-BASED MODELS

AGENT-BASED EXPLICIT SPATIAL AND TEMPORAL (A-BEST) MODEL

Overview

The NIH engaged the State University of New York at Buffalo to develop a spatially (location) explicit temporal (over time) model that would build upon traditional epidemiologic approaches such as the SEIR (Susceptible, Exposed, Immune, Removed) model but that would use advanced simulation techniques to predict 1) who would become infected; 2) where they would be infected; 3) when the infection would occur; and 4) the outcome of the infection. A team of medical, population, social and transportation geographers and specialists in individual epidemiologic modeling provided the Agent-Based Explicit Spatial and Temporal (A-BEST) simulation model to study the potential movement of exotic disease agents, over time and space, through a community.

A model is a *simplified representation* of a system or set of circumstances intended to promote understanding of the *actual* system or circumstances such as the underlying drivers of the system dynamics; and causal factors and pathways (Hartley et.al., 2005). Simulations are the computational results of a detailed model that enable the prediction of system behavior in a given set of circumstances. Modeling and simulation are used together to gain an understanding of interactions within the system. These techniques are often used when actual experiments are not possible. The A-BEST model and subsequent simulations allow study of the interactions within a community (the system) that may lead to infectious disease transmission subsequent to a laboratory-acquired infection in a worker or other hypothetical scenario.

NIH developed hypothetical scenarios depicting situations defined by the public and subjected the scenarios to the complex social, work and home interactions and networks of daily life through A-BEST modeling and simulation. See Section VII for risk and accident scenarios subjected to A-BEST simulation. Combining A-BEST modeling with the infectious disease release or accident scenarios allowed simulation of real-world movement of individuals through daytime, nighttime and pastime activities; at home and at work and while in transit. A-BEST simulation provided millions of interactions during explicit individual movement simulating daily life activities thereby optimizing exposure opportunities resulting in worst case modeling.

The A-BEST simulation is constructed based on a hypothetical situation of a laboratory worker who sustains a laboratory-acquired infection and subsequently leaves the laboratory and enters the community. Potential diseases caused by these infectious agents are usually transmitted through percutaneous injury or close contacts. The simulation employs an individual-based and spatially explicit modeling approach for predicting possible health outcomes in three communities: Boston-BUMC, Tyngsborough, MA, and Peterborough, NH. The simulation is conducted for four infectious diseases – Ebola Hemorrhagic Fever, Monkeypox, Brazilian Hemorrhagic Fever (caused by Sabia virus),

and Rift Valley Fever. See Figure VI-1 for a Simplified Process Flow Chart of the A-BEST Model. See Appendix 3 for a step by step explanation of the model construction.

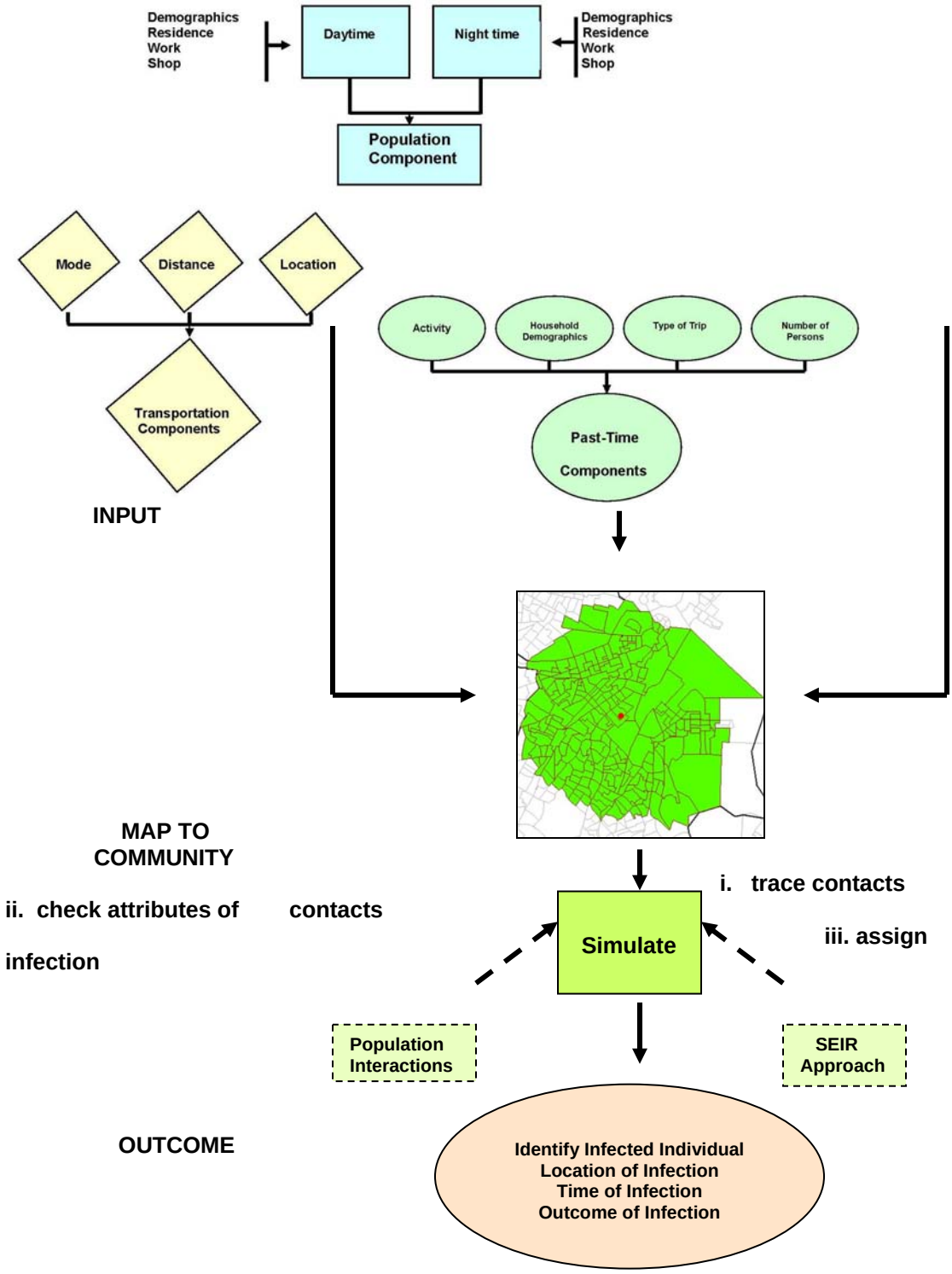


Figure VI-1. Simplified Process Flow Chart for the A-BEST Model

Model Design

The A-BEST modeling approach used in this simulation is built on a conceptual framework developed by Bian (2004) and an analytical framework by Bian and Liebner (2007). The conceptual framework identifies the basic assumptions underlying the model and the essential components for the modeling. Briefly, the conceptual framework focuses on the modeling of discrete individuals, individualized interactions, and changes in the interaction with location and time. Based on this conceptual framework, the analytical framework further defines a multi-population and two-scale network. The multiple populations include the nighttime population at homes and the daytime population at workplaces. The two scales include a local-level and a population-level network. The interactions between individuals at home or in workplaces form a local network, whereas the mobility of individuals between homes and workplaces links these local networks into the population level network. In this network, each individual is represented as a node and the interaction between any two individuals is a link. Collectively, these individuals and their individualized interactions portray a heterogeneous spread of disease through the multi-population and two-scale network. See FigureVI-1.

Using these conceptual and analytical frameworks, three populations are modeled in this simulation, namely, a nighttime population at home, a daytime population in workplaces, and a pastime population at service-places.

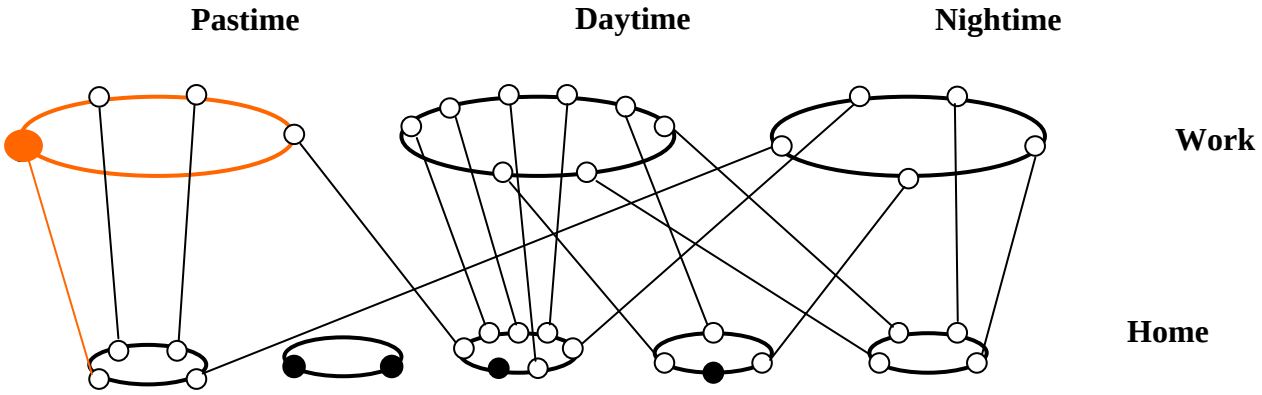


Figure VI-2. Graphic depiction of the multi-population, two scale network (Bian 2004).

A given population travels between home and work or home and a service place as depicted in Figure VI-3. The yellow "hot spots" depict foci of disease. Disease can be transmitted from one location to another and spread through individual or "agent" interactions consistent with daily life activities.

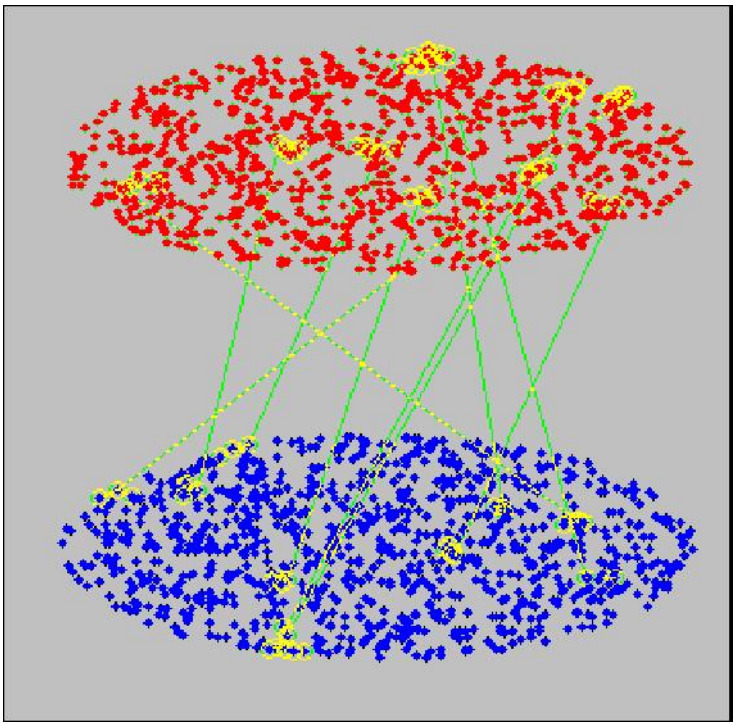


Figure VI-3. Population of the conceptual and analytical framework with identified attributes allows tracking of disease transmittance from one location to another through individual interactions consistent with daily life activities. (Bian and Liebner 2007)

These three populations (pastime, daytime, and nighttime) represent the same set of individuals at different locations and time periods of a day. The contact between individuals at homes, workplaces, and service-places represents the local network. The travel of individuals between homes, workplaces, and service-places links the three populations into the population-level network. All simulated individuals are assigned a unique identification and many other attributes. Each individual is assigned to a family, a workplace, and a number of service-places to create the three populations, and to establish the links between the populations. The spatial locations of homes, workplaces, and service-places and the travel routes for each individual between these three sets of locations are explicitly represented. After an initial infection is introduced into a community, the model simulates the first generation of infection. That is, the model identifies the first set of individuals who are in direct contact with the first infected individual through contacts at home, in the workplace, and at service-places. Each of

these first-set individuals is assigned an infection status, either healthy or infected, according to the primary infection rate of a specific disease. Next, the model simulates the second generation of infection by identifying the second set of individuals who are in direct contact with those infected in the first generation of infection. An infection status is assigned to these second-set individuals according to a secondary infection rate. Thirdly, the model identifies the third set of individuals who are in direct contact with those infected in the second generation of infection. These third-set individuals are assigned an infection status according to what is known as a tertiary infection rate. See Tables VI-1 through VI-4 for the assumptions used to inform the simulations. **It is critical to note that the assumptions overstate the likely rate of infection dramatically.** The real likelihood of infections and the spread of any resultant disease would be much lower or even zero.

TableVI-1. A-BEST Simulation Assumptions – Ebola virus.

Agent: Ebola virus **Representative virus group: *Filovirus***	**Modeling Assumptions**	**Relevant Scientific Data**
Primary Infection Rate	100%	Natural infection rate unknown. Laboratory infection rate unknown.
Primary Route(s) of Human Infection		Person-to-person (most likely through contact with blood or body fluids); Percutaneous through reuse of needles or accidental needlesticks;[1] contact with cadavers during preparation for burial; Direct contact with infected nonhuman primates (eg, chimpanzees, gorillas); Contact of oral mucosa or conjunctivae through infectious droplets or direct contact.
Secondary Transmission	Medical personnel at high risk of secondary transmission if barrier precautions are not used. 0.4% infection with accidental needlestick for HCW.	Medical personnel at high risk because they may also have direct contact with highly infectious body fluids, soiled clothing or bedding, aerosols (coughing), etc., in addition to accidental needlesticks and sharps injuries. (0.3% - 0.5% for HIV)[2,3] No relationship has been shown between rates of HIV and Ebola transmission in blood; used for simulation purposes only
Tertiary Transmission	0.003% casual contact 30% is used for simulating sexual transmission on first day of mild symptoms.	Theoretically possible but extremely low probability in U.S. No direct evidence of sexual transmission of Ebola virus, even though virus or viral nucleic acid has been found in seminal fluid or vaginal mucosa of 9 convalescent patients.[4,5,11] There is no evidence of sexual transmission during the incubation period or acute stages of disease. One report of Marburg virus transmission via semen 3 months after patient had recovered from disease.[6]
Incubation Period	7 days	2-21 days[7]
Infectious Period	7-69 days	Not before onset of febrile illness; increases with stages of illness.
Fatality Rate	85%	57-88%[8] Fatality rate in humans depends on the virus strain: Zaire 55-90%; Sudan 53-65%; Reston 0%[9]
Reservoir (if known)		Unknown (although asymptomatic infection has been found in several species of fruit bats)[10]
Vector (if any)		Unknown

1. CDC 2004
2. CDC 2001
3. CDC 1987, 1988
4. Emond et al. 1977
5. Rodriquez et al. 1999
6. Martini and Schmidt 1968
7. APHA 2004
8.CDC 2002
9. Mahanty and Bray 2004
10. Leroy 2005
11. Emond 1977

Table VI-2. A-BEST Simulation Assumptions – Recombinant Monkeypox.

Agent: Recombinant Monkeypox **Representative virus group:** *Orthopoxvirus*	**Modeling Assumptions**	**Relevant Scientific Data**
Primary Infection Rate	100%	0.11% in males; 0.09% in females (from affected African villagers)[1] Laboratory workers are required to be vaccinated with the smallpox vaccine, which also protects against monkeypox. No documented monkeypox infections in laboratory workers.
Primary Route(s) of Human Infection		Direct contact with infectious animals, materials or exudates; nosocomial potential low in U.S.[2]
Secondary Transmission	0.4% health care worker by accident 8.3%	Unvaccinated medical personnel are at high risk of secondary transmission if barrier precautions are not used. In the U.S. outbreak, secondary contacts were monitored carefully and none acquired infection. From the Congo (1996-1997): estimates of approximately 8.3 per 100 total contacts; 7.5% in unvaccinated contacts, 0.9% in vaccinated contacts[3,4]
Tertiary Transmission	0.03%;	Possible but low risk in U.S. Up to 6 successive transmissions suggested in one study of an African outbreak[5]; sustained transmission in the human population unlikely[6]
Incubation Period	4-24 days	7- 21 days Infected individuals are highly unlikely to be infectious during the asymptomatic incubation period[7]
Infectious Period	45 days	Individuals are infectious from the development of symptoms until resolution of all skin lesions. [8]
Fatality Rate	14%	1-10% in Africa; would be much lower in U.S. due to better nutrition and access to health care; 1-14% in unvaccinated children[9] In the 2003 monkeypox outbreak in the U.S., there were no fatalities among 47 confirmed cases.
Reservoir (if known)		Known reservoir: African rodents (*Funiscuirus* spp., *Heliosciurus* spp., *Cricetomys* spp., Atherurus spp., Graphiurus spp., and Hybomys spp.) In the U.S.- infected, ill Prairie Dogs (*Cynomys Ludovicianus)* There is no evidence that hamsters are susceptible to monkeypox virus infection; but the CDC warns animal handlers exposed to monkeypox not to handle 'pocket pets' such as gerbils and hamsters [10]; There is no evidence that human to animal transmission occurs but CDC warns veterinarians infected not to handle animals because possibility exits[10]
Vector (if any)		None known

1. Jezek et al. 1986
2. Fleischauer 2005
3. Hutin et al. 2001
4. Mwanbal etal. 1997
5. Learned 2005
6. Jezek 1987
7. Fenner et al 1988
8. Jezek and Fenner 1988
9. APHA 2004
10. CDC 2004

Table VI-3. A-BEST Simulation Assumptions – Sabia virus

Agent: Sabia virus Representative virus group: *Arenavirus*	**Modeling Assumptions**	**Relevant Scientific Data**
Primary Infection Rate	100%	Undetermined; only 3 identified cases, one naturally occurring, two laboratory-acquired
Primary Route(s) of Human Infection		Unknown, although laboratory-acquired cases appear to have been contracted through aerosols; in nature presumed to be through inhalation of infective rodent feces or urine similar to other arenaviruses [1,2]
Secondary Transmission	human-to-human infection	No secondary or nosocomial transmission documented with Sabiá virus. Yale case had 139 secondary contacts; all underwent clinical surveillance for 6 weeks. No contact tested positive for either viral antigen (by PCR) or antibody.
	0.3% by family contact 0.4% by health care contact 0.03% by casual contact	There are reports of Machupo and Junin virus transmission via close person-to person contact during illness or convalescence[3,4]
Tertiary Transmission	0.03% by family contact 0.04% by health care contact 0.003% by casual contact	No tertiary transmission documented for Sabiá virus.
Incubation Period	5-16 days; A random number generated (5 to 16) used for each simulation	8 day incubation period for 2nd lab case (Yale)
Infectious Period	90 days	Unknown; Lassa fever virus can be detected in semen up to 3 months after acute infection No reports documenting transmission of arenaviruses by infected persons during the incubation period[5]
Fatality Rate	0.003%	Unknown; only 3 known cases; 1 fatality The 1 fatality occurred in the naturally acquired infection (index case)
Reservoir (if known)		Not known. Rodents are reservoirs for most arenaviruses, although Tacaribe virus was initially isolated from a bat.
Vector (if any)		none

1. APHA 2004
2. Peters 2002
3. Douglas 1965
4. Briggilier et al. 1987
5. Kilgore et al. 1997
6. Borio et al.

Table VI-4. A-BEST Simulation Assumptions –Rift Valley Fever virus.

Agent: Rift Valley Fever Virus **Representative virus group:** *Bunyavirus(plebovirus)*	**Modeling Assumptions**	**Relevant Scientific Data**
Primary Infection Rate	100%	No data exists for the U.S.
Primary Route(s) of Human Infection	Aerosol and bite of infected mosquito	Bite of infected mosquito or aerosols are primary routes of infection; direct contact with hides or tissues of infected animals (during slaughter, births or abortions, etc)[1]
Secondary Transmission	1% infection rate for 90% of infected humans (1 in 100 infected mosquito bites will transmit disease) 0.03% infection rate for 10% infected humans (bloodborne transmission or aerosol transmission resulting from occupational exposure) **Assumptions for density of infected mosquitoes:** **BUMC:** (Surface water area *10,000 mosquito/acre) * 1% /3 = RVF infected mosquitoes' bites. **Peterborough:** [(Surface water area * 10,000 mosquitoes/acre) + (347 animals * 271 mosquitoes/head * 10)] * 1% /3 = RVF infected mosquitoes' bites. **Tyngsborough:** [(Surface water area * 10,000 mosquitoes/acre) + (369 animals* 271 mosquitoes/head * 10)] * 1%/3 = infected mosquitoes' bites.	Virus titers in infected humans are high enough to infect mosquitoes,[2] so there is a theoretical possibility that after biting infected humans, those mosquitoes can, in turn, transmit virus to other vertebrate hosts or introduce virus into new areas.[3,4] It is unknown whether this can occur via interhuman transmission mediated by arthropods or if viremic domestic animals are required[7].
Tertiary Transmission	none	none
Incubation Period	5 days	2-6 days[5]
Infectious Period	N/A	No direct human-to-human transmission documented or suggested.
Fatality Rate	0.9% of infected individuals will develop hemorrhagic disease, of whom 50% die; 0.003% fatality rate for the remaining infected individuals	Overall <1%; 0.9% of infected individuals will develop hemorrhagic disease, of whom 50% die

Reservoir (if known)	Ruminants (e.g., cattle and sheep) as an amplification host	RVFV is transmitted by mosquitoes and can lay dormant in mosquito eggs for at least 10 years in dry seasons in shallow wetlands in Africa. When these shallow wetlands fill during rainy season, the eggs hatch and infected mosquitoes look for a meal.[6] Cattle, sheep, other ruminants and many other vertebrates (including humans) are highly susceptible.
Vector (if any)	*Aedes canadensis*	*Aedes, Culex, Anopheles*, and other species of mosquitoes

1. Schmaljohn et al. 2007
2. Meegan et al. 1979
3. Hoogstraal et al. 1979
4. Mebus 1998
5. APHA 2004
6. Davies et al. 1985
7. Peters 2005

The Three Communities

The three communities simulated (BUMC, Tyngsborough, and Peterborough) are defined as the block groups within a certain distance from a University of Boston (UB) property. A block group is one of the three spatial units used by the US Census Bureau to represent population. The three spatial units include census track, block group, and block. A census track contains approximately 4,000 to 5,000 individuals, a block group contains approximately 1,000 individuals, and a block is closely related to a street block. The census statistics data are available in an aggregated form at the block group and census track levels (www.census.gov/geo/www/tractez.html, www.census.gov/geo/www/cob/bg_metadata.html). The 2000 census data are used to define the communities.

The BUMC community is defined as the block groups within a 2-mile distance from the BU Medical Center. The race, education, and income information provide by the 2000 census for the general BU Medical Center area are analyzed. Based on this analysis, the block groups within the 2-mile distance seem to be most representative of the area and contain an adequate population size for the simulation. This delineation results in a total of 227 block groups, which contain 104,000 households and 245,842 individuals. Similarly, 17 block groups within a 2.5-mile distance from a BU property in Tyngsborough, MA are defined as the Tyngsborough community that includes 11,000 households and 30,022 individuals. Six block groups within a 3-mile distance from a BU property in Peterborough, NH are defined as the Peterborough community that includes 3,455 households and 8,941 individuals. The three communities collectively represent 250 block groups, 118,455 households, and 284,805 individuals.

Simulation of the Nighttime Population. Two sets of data are used to simulate the nighttime population at homes for the three communities, the 2000 census data and a household dataset purchased from ReferenceUSA, Inc. ReferenceUSA is an internet-based reference service providing comprehensive databases of business and consumer information products. They maintain a proprietary database of virtually all U.S. and Canadian businesses and residents. Since the US census statistics are created based on the residential locations of the population, they are used as the basis for the simulation of the nighttime population. Specifically, 2000 Census Summary File 3 (SF3) “2000 Census

Population and Housing" (http://factfinder.census.gov/servlet/DatasetMainPageServlet) provides most relevant information. SF3 contains a large number of tables, each of which contains a list of variables that describe the characteristics of population (P#) and households (H#) in a block group. The following 11 tables in SF3 are used for the simulation.

P1. Total Population
P8. Sex by Age
P9. Household Type (Including Living Alone) by Relationship
P10. Household Size by Household Type by Presence of Own Children under 18 Years
P11. Household Type (Including Living Alone) by Relationship for the Population 65 Years and Over
P12. Households by Age of Householder by Household Type (Including Living Alone) by Presence of Own Children under 18 Years
P14. Household Type by Household Size
P15. Family Type by Presence of Own Children under 18 Years by Age of Own Children
P16. Own Children under 18 years by Family Type by Age
P17. Family Type by Presence of Related Children under 18 Years by Age of Related Children
P18. Sex by Marital Status for the Population 15 Years and Over

Two sets of information are extracted from the tables, one for aggregated individuals and another for aggregated households in a block group. The information for aggregated individuals includes statistics of age range (children: ≤5, 6-15, 16-17, adults: 18-64, senior: ≥65), gender, and the relationship to the householder (householder self, spouse, own children, other children, other). The age range is determined based on the needs of subsequent simulations and the level of detail of the information provided in the tables. The information for households includes statistics of household size and household type. The latter includes family households (married-couple, single-father, and single-mother families) and non-family households (living alone and not living alone).

Using the extracted information, the individuals in a block group are assigned to individual households in the following sequence. The size of each household is first determined. Householders of all households, the spouse and children of the householder for family households, members of the non-family households, remaining seniors, and remaining adults are subsequently assigned to a household. For a given type of individual (by gender, age, and relationship) and a given type of household (by size and type), the assignment of specific individuals to a specific household is based on a Monte Carlo random draw process. (Monte Carlo methods are a widely used class of computational algorithms for simulating the behavior of various physical and mathematical systems, and for other computations. They are distinguished from other simulation methods (such as molecular dynamics) by being stochastic, that is nondeterministic in some manner – usually by using random numbers (in practice, pseudo-random numbers) – as opposed to deterministic algorithms. Because of the repetition of algorithms and the large number of calculations involved, Monte Carlo is a method suited to calculation using a computer,

utilizing many techniques of computer simulation.) The information used to create households and individuals is in turn used as attributes for the households and individuals for subsequent simulations.

The households and individuals data created based on the 2000 census data are then merged with a dataset purchased from ReferenceUSA in order to obtain additional attributes. The ReferenceUSA dataset provides information at the household level. Most critical of the additional attributes are the location and purchasing behavior of households. Both are necessary for subsequent simulations. The merger is based on the attributes common to both data sets, such as household size, characteristics of the householder (marital status, gender, and age range), and the presence of children. This combined information completes the creation of the nighttime population for the three communities.

Simulation of the Daytime Population. Three sets of information are used to simulate the daytime population at workplaces for the three communities, the 2000 census SF3 data, a "business" dataset purchased from Environment Systems Research Institute Inc. (ESRI), and the attributes associated with previously simulated individuals and households. ESRI is a data acquisition service identifying suitable data sets from among thousands of worldwide data vendors. While the ESRI data provide information about workplaces, the SF3 data and the previously created attributes of the individuals and households help assign the individuals to these workplaces. Collectively, these datasets help create the daytime population and establish the links between the nighttime population and the daytime population.

The workplaces are defined as those businesses that are within a one-hour driving distance from the three communities. The driving distance is computed using Geographic Information Systems (GIS) software and the speed limit of the road network provided in the Census Topologically Integrated Geographic Encoding and Referencing (TIGER) data. The one-hour driving distance is chosen because, first, according to SF3, it encloses approximately 99% of travel-to-work distances of all workers in a block group and, second, this distance includes an area that is relevant to this simulation. This definition results in a total of 310,400 workplaces. Each contains the following attributes: name, location, number of employees, estimated sales, and the type of business that is coded by the North America Industry Classification System created by the United States Census Bureau (NAICS) http://www.census.gov/epcd/www/naics.html).

The following 14 tables in SF3 are used to assign individuals in households to a workplace:

P30. Means of Transportation to Work for Worker 16 Years and Over
P31. Travel Time to Work for Workers 16 Years and Over
P32. Travel Time to Work by Means of Transportation to Work for Workers 16 Years and Over Who Did Not Work at Home
P36. Sex by School Enrollment by Level of School by Type of School for the Population 3 Years and Over

P39. Sex by age by Armed Forces Status by Veteran Status for the Population 18 Years and Over
P42. Sex by Age by Disability Status by Employment Status for the Civilian Non-institutionalized Population 5 Years and Over
P43. Sex by Employment Status for the Population 16 Years and Over
P44. Family Type by Employment Status
P48. Family Type by Number of Workers in family in 1999
P49. Sex by Industry for the Employed Civilian Population 16 Years and Over
P53. Median Household Income in 1999
H44. Tenure by Vehicles Available
H45. Tenure by Vehicles Available by Age of Householder
H46. Aggregate Number of Vehicles Available by Tenure

Four sets of statistics information are extracted from these tables: (1) workers who work outside the home, (2) the transportation means (walk, bus, car, and subway) used by these workers to travel to work, (3) the time needed for the workers to travel, and (4) the type of employer of the workers that is coded by NAICS. According to this extracted information, certain individuals in each household are assigned as workers. Then, a number of vehicles are assigned to each household. Next, a transportation means and a travel time are assigned to each worker in order to derive a travel distance according to: travel distance = travel speed x travel time, where the speed is estimated from the transportation means. Lastly, each worker is assigned to a work place based on three constraints: the travel distance to work, the match between the employer type of a worker and the type of a workplace (by NAICS), and the number of employees at a workplace. The service-oriented workplaces are open in both daytime and pastime, thus two shifts of workers are assigned to these workplaces. For a given type of worker and a given type of workplace, the assignment of a specific worker to a specific workplace is based on the Monte Carlo method.

The school-aged individuals and other individuals in each household are assigned to a school according to three constraints: an age range previously simulated for the nighttime population, the school type (nursery and preschools, elementary and secondary schools, colleges and universities) appropriate for them, and the nearest schools identified in the ESRI data. This completes the creation of the daytime population at workplaces and links the nighttime population at homes to the daytime population in workplaces.

The information used to assign individuals to workplaces and schools are in turn used as attributes for workplaces and schools, and are added to those attributes previously created for the individuals and households during the simulation for the nighttime population. The attributes for both the daytime and nighttime populations are then used for subsequent simulations.

Simulation of the Pastime Population. The simulation of the pastime population at service-places is based on three sets of information. The first set is a 1991 travel diary survey data obtained from the Central Transportation Planning Staff of Massachusetts. The second is a subset of workplaces that is identified as service-places. The third is the

attributes for individuals and households created in the previous simulations for the nighttime and daytime populations. The travel diary survey provides information on the type and frequency of services needed, the service-place data describe the characteristics of the service-places, and the attributes data help assign individuals to service-places. Subsequently, the three sets of data collectively help create the pastime population and link it with the nighttime and daytime populations.

The travel diary data is a survey of travel behavior of all individuals in selected households in the Boston area for one day during 1991. This is the most recent survey of this kind for the study area. The survey involves a total of 39,934 trips taken by 9,256 individuals in 3,906 households. This data provide three sets of information most relevant for this simulation. The first set is six types of services recorded as the purpose of the trips: pick up or drop off, shopping, eating out, recreation, personal, and social. A trip to a hospital is created as a special seventh type for this simulation, according to the infection Scenarios (see Section VI). The second set is three types of service trips in terms of the origin and destination. These include home-to-service, work-to-service, and service-to-service trips. The third set is the frequency statistics, by number of individuals, by number of households, and by number of trips, for each of the six types of services and for each of the three types of service trips. The statistics also include the frequency for no-trips, which represents the number of individuals and households that did not travel on the surveyed day.

A total of 89,159 workplaces are identified as service-places from the 310,400 workplaces. They belong to the six aforementioned types of services based on the NAICS attribute of the workplaces. The individuals in households are assigned to these service-places. The assignment is based on the attributes of individuals (gender and age range) and households (the number of workers in a household, the number of vehicles in a household, and the household size). Using the frequency statistics as constraints, the total number of trips for a household is first determined. Workers in a household, then other members of the household, are subsequently assigned to one of the three types of service trips (workplace-to-service, home-to-service, and service-to-service), then to one of the six types of services. For a given type of service trip, a given type of service, and a given type of individual in a given type of household, the Monte Carlo method is used to randomly assign an individual to the nearest service place whose NAICS matches the service type.

Trips to service-places occur during both the daytime and pastime periods of the day. Each service-place has two sets of individuals, those who need services and those who provide services. Both sets of individuals are used as the attributes of service places. The service and trip information is used as additional attributes of individuals and households. This completes the creation of the pastime population and links the three populations into a single social network.

Simulation of Disease Transmission. With the three populations created and the links between them established, the first case for each of the five diseases is introduced into each of the three communities according to the Scenarios (see Section VI). Using the A-

BEST model (see the section on Model Design), the transmission of a disease may follow all possible paths in the three-population and two-scale network.

The probability of an individual becoming infected after being in contact with an infected individual is based on the primary, secondary, and tertiary infection rate of a disease (see the section on A-BEST Model Design). The three infection rates for specific diseases are specified in the disease tables (Tables VI-1 through 4). An infection rate of 0.1 is used for adults and a rate of 0.15 for children and seniors. For a given infection rate, a given disease, and a given set of individuals who are in direct contact with an infected individual, the Monte Carlo method is used to assign the infected status to certain individuals. Each infected individual experiences three stages of disease development: an incubation period, an infectious period, and an immune period. The length of each period for each disease is specified in Tables VI-1 through 4. Infected individuals may recover or die. The probability of the latter is simulated based on the fatality rate of each disease (see Section VI). If they survive, the individual obtains a life-long immunity, except for Influenza because new strains may be introduced.

The transmission occurs at three time periods of a day. These include daytime at workplaces or homes, pastime at service-places, workplaces, or homes, and nighttime at homes. The contact between individuals follows a full mix pattern at homes, and a partial mix with an average of 10 contacts at workplaces and service-places.

Simulation Output. The results of the simulation, also known as the simulation output, includes the day of the infection, the identification of the infected individuals, the time period and location of the infection, and the consequence of the infection (recovery or death). See Appendix 1 for an example of simulation outputs for each of the four diseases modeled (Ebola HF, Brazilian HF, Monkeypox and Rift Valley Fever) in the simulated communities of Boston-BUMC, Tyngsborough and Peterborough. Results of each simulation overstate the actual risks to the communities, to a large degree, as a result of extremely conservative assumptions (general and scenario-specific) made for each disease scenario.

The Multilayer Agent Based-Simulation Tool (MLAB-ST) is a set of algorithms and data visualization techniques for understanding the interactions between large numbers of agents in an environment. The problem addressed with MLAB-ST in this project is the spread of Rift Valley Fever virus in the New England region as a result of a transportation accident. The simulation starts with one human infected as a result of an aerosolized exposure to the virus in a dry ice explosion during transport. The assumption is that this human becomes viremic (viruses in the blood) and is bitten by a single mosquito, which in turn infects one ruminant in the vicinity. The ruminant amplifies the virus and is subsequently bitten by other mosquitoes which become infected, transmitting the virus to other ruminants and humans.

Agent-Based (A-B) Simulation

MLAB-ST is an A-B simulation tool, and as such, supports the basic components of the A-B modeling environment – environment, agents, and a scenario.

1. **Environment**: The environment acts as a container for the simulation, both computationally and in terms of visualization. It usually represents a specific physical space or geography and a defined time span. The environment has the ability to manage messages across itself to support global housekeeping chores and communication to agents about environmental events. The environment manages events, which are timed processes that have global scope and can be executed by the environment on startup, shutdown, or asynchronously throughout the simulation based on predefined conditions. Our simulation environment supports the three regions of Boston, Peterborough, and Tyngsborough. The simulation time span is eighteen months beginning April 1st, and was chosen to represent the full lifecycle of the Aedes canadensis starting with seasonal high density, through over wintering, and through another high density period.

2. **One or more agents**: An agent is an entity that interacts with its environment and with other agents based on rules and constraints. The notion of an agent is based in biological principals as an agent is created (instantiated) from a base class or template. The base class defines the most important features (properties) and behaviors (methods) of the agent at birth, many of which can be modified by the environment or other agents over time. Our agents are mosquitoes, ruminants, and humans, and are created in the number specified by the simulation scenario at startup.

Each individual agent is created with a corresponding "state chart" that defines its initial state and how it will respond to its environment over its life span. Figure VI-4 represents the state chart of a ruminant agent.

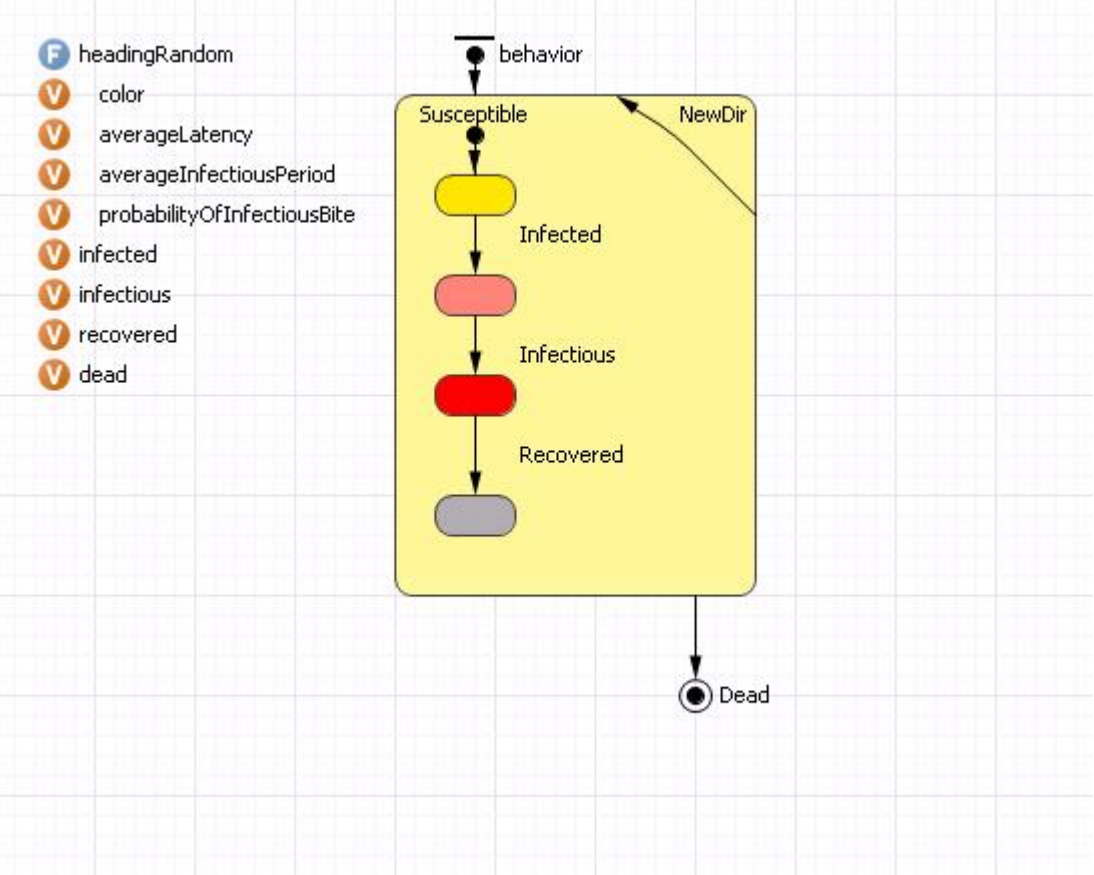


Figure VI- 4. Each individual agent is created with a corresponding "state chart" that defines its initial state and how it will respond to its environment over its life span.

All ruminant agents are created *susceptible*. If and when they are bitten by an infected mosquito, the transition from *susceptible* to *infected* occurs. After some time randomly defined in a "latency" range at creation, the agent makes the transition to *infectious*. After some time randomly defined in a "infectious" range at creation, the agent makes the transition to *recovered*, or at some lower frequency, *dead* as a result of the disease. In any case, the agent will end up dead after its normal life span.

During the agent's lifespan, it wanders randomly within its constraints, and either by reaching a boundary or a random timer event, picks a new direction and heads off.

This agent is representative of all agents in the model, and separately created state charts for every one of the millions of individuals in the simulation are managed by the simulation environment as they interact.

3. **A simulation scenario with rules and parameters**: The simulation scenario with rules and parameters is sometimes referred to as the *experiment*. The experiment sets the stage for the environment and the agents, and is typically run hundreds or thousands of times to collect simulation output. Experiments based on similar starting values are used to validate the constancy of the model. Experiments with widely differing starting values are used to build predictive capacity models.

Model Design

The interactions are described by MLAB-ST using three different agent models: Human, Ruminant, and Mosquito. These agents are instantiated in predefined numbers for the region under study and then interact according to a set of mathematical relationships and geography-based constraints.

A unique feature of MLAB-ST is the use of layers of maps, in this case representing physical regions, which reflect pixels inaccessible to the agent. In our use, roads, water, and fixed regional boundaries are marked on maps in pure red, and used to block agent movement. An agent "looks down" at each time step to see if the next movement will be over one or more red pixels in layers directly beneath it. If so, another course is plotted.

During agent movement, a variety of interactions can occur depending on what other agent is in the vicinity. The essence is, for example with a mosquito, a 1-pixel movement is made at which time all other agents in the area are examined. It there is a ruminant or human in the same space, the mosquito bites with some probability and infects with another probability (see chart below). The states can change at each time step as the scenario unfolds. Although probability and temporal ranges are predetermined, the progression of the simulation is stochastic (random).

Model Flow

1. Simulation is started and all constraint maps are loaded

2. Agents are instantiated in the specified numbers and placed randomly wherever they are allowed to be, and then set in random motion, avoiding any barriers.

3. At each time step, each mosquito examines its vicinity.

    a. If there is a ruminant present, it gets bitten
        i. if the mosquito is infectious the virus is transmitted to the ruminant
        ii. if the ruminant is infectious the virus is transmitted to the mosquito.
    b. If there is a human present, it gets bitten
        i. if the mosquito is infectious the virus is transmitted to the human
        ii. if the human is infectious the virus is transmitted to the mosquito

4. Transmission of virus depends on sufficiently high titer, and once transmitted, the response of the infected organism is determined by its individual state chart.

The infection process continues until all ruminants are recovered or dead.

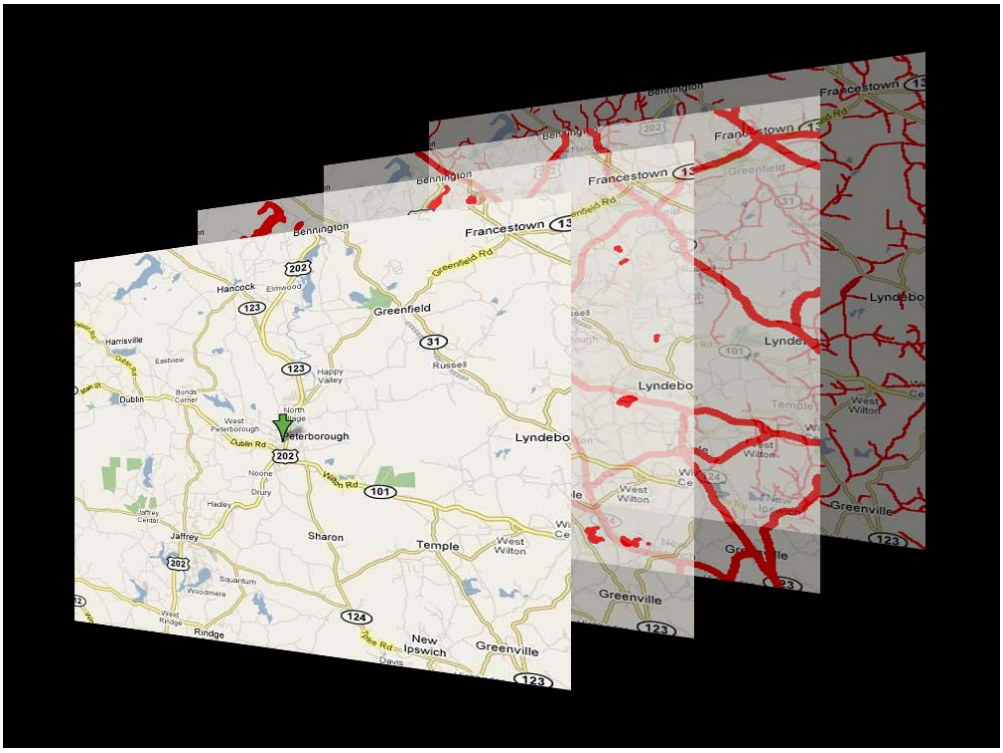


Figure VI-5. MLAB-ST Constraint Layers. **MLAB-ST distributes agents randomly in a highly-defined constraint space.**

MLAB-ST distributes agents randomly in a highly-defined constraint space. In the case of geography-based simulations like this particular one, the constraints are built up from area maps. The graphic above illustrates how layered maps are stacked with various features marked in red, in this case, water features, big roads, and small roads. We provide additional constraint layers as described below, the most dramatic of which is the ruminant layer. Constraints can be added into the model in minutes by making a copy of the master map and drawing with true red (255,0,0) in areas to be masked out. This layer reference is simply added to the agent event loop. This graphical approach provides layperson access to complex models using familiar tools.

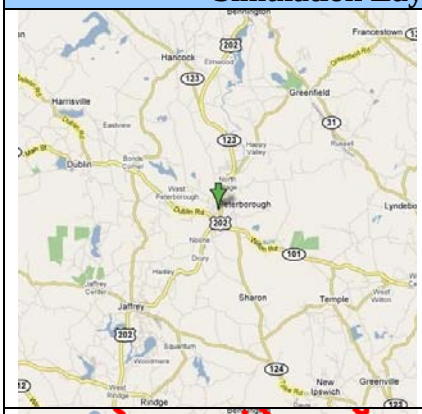
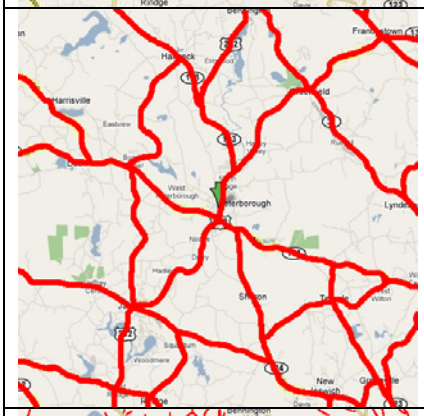
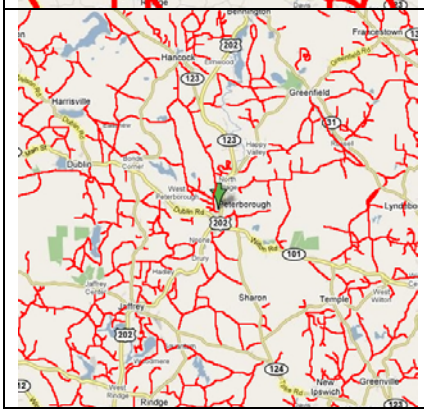
At start-up, agents are randomly distributed into available (non-constrained) positions in the simulation space and then start moving about. At any given pixel, a randomly moving agent can look down through the layers of constraints and move based on the presence or absence of red. Each class of agent (here, Human, Ruminant, and Mosquito)

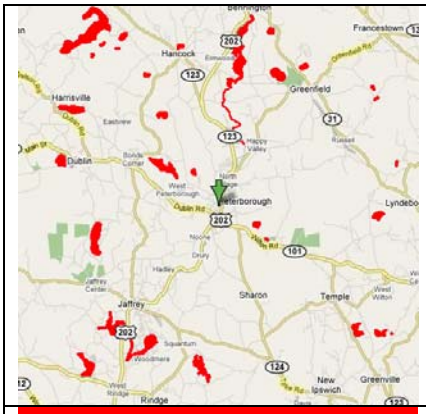
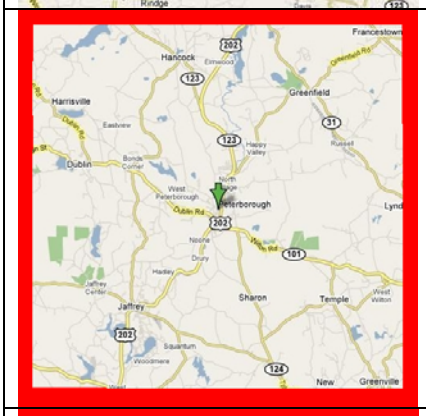
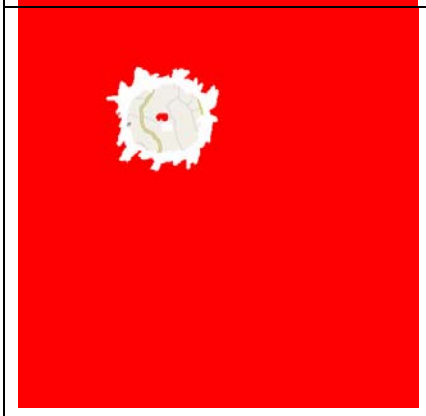
has behaviors that depend on different map layers. For example, in this simulation humans cannot walk on water and they have no boats, so when they randomly arrive at the edge of water, they turn around and continue moving in a different direction. They "know" they have arrived at the edge of water because the next step puts them over a red pixel on the water layer.

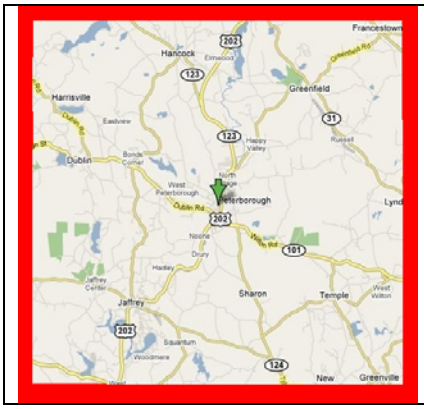
Ruminants are constrained more heavily in this simulation since in the real world they are normally kept within fenced areas. By definition, we model 10% of the regional ruminants and randomly place them within the boundaries of the ruminant layer (see Table VI-5). They can neither leave the area nor walk on roads or water. Mosquitoes, on the other hand, as in reality can go anywhere they want to within their defined lifespan and range.

The specific layers used for this simulation are shown in Table VI-5, using Peterborough, NH as an example.

Table VI-5. The specific layers used for the Rift Valley Fever MLAB-ST simulation are shown below using Peterborough, NH as an example.

Simulation Layer	Purpose
	Master Map Layer is the visual piece presented to the user. It is not used for computations.
	Big Roads Layer is used for agents that cannot be on big roads.
	Small Roads Layer is used for agents that cannot be on small roads.

	Water Feature Layer is used for agents that cannot be on water. It is used in this simulation only to constrain placement or movement, not to determine standing water for purposes of mosquito population calculation.
	Human Layer is used to place and constrain humans. Big Roads, Small Roads, and Water Feature layers are used along with this map.
	Ruminant Layer is used to place and constrain ruminants. For this example, the facility in Peterborough was pinpointed. Using Google Earth's "Line" measurement utility, a one mile radius was established around the facility. To add to the realism, and account for environmental factors such as wind and precipitation that might affect the vector's range, +/- .35 miles were added to the established one mile radius. The resulting ruminant layer is a rough circular area. Big Roads, Small Roads, and Water Feature layers are used along with this map.

		Mosquito Layer is used to constrain mosquito placement and movement. Unlike humans, they are not further constrained by roads or water features.

The following assumptions and constraints were used to populate the MLAB-ST for the purpose of this simulation.

Mosquitoes Species.

The mosquito species Aedes canadensis is used because it

- is an aggressive biter (> 260 bites/minute)
- is long lived (up to 90 days)
- is univoltine
- can overwinter as eggs or live females
- can transmit virus transovarially (to its eggs)

This is a worst-case vector for our scenario, and all of the above factors are incorporated into the simulation.

Population. Only female mosquitoes were modeled and the assumption was made that, at the peak population around April 1st, there were 100,000 females/acre wetland or other suitable water body area. (±10% for each simulation run for realism), with area given for Boston at 199 acres, Peterborough at 6,309 acres, and Tyngsborough at 2,654 acres. Additionally, the graph below articulates the seasonal variation of the mosquito population.

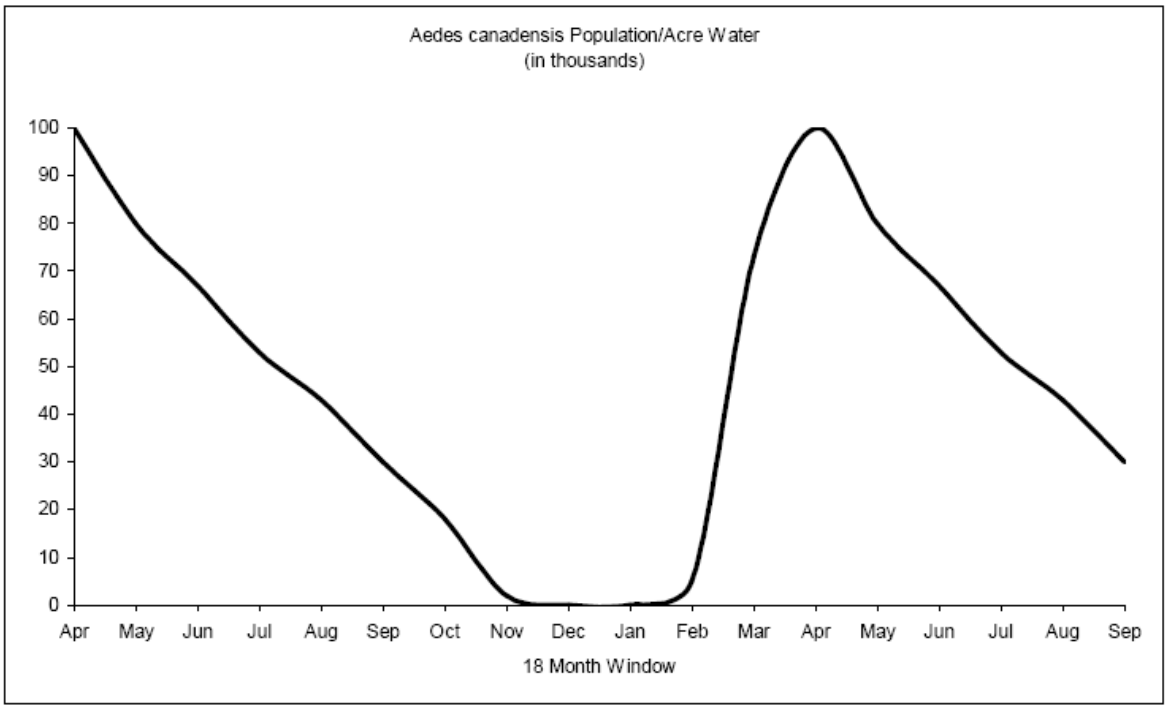


Figure VI-6. Seasonal Variation in *Aedes canadensis* Population

The density of mosquitoes in our simulation is controlled to match the known seasonal variation, which is accomplished by simulating deaths and hatchings as needed. Newly-hatched mosquitoes may or may not be infected depending on the status of the female who produced the eggs.

Infectivity Rate. The rate of viral dissemination, and therefore the chance of passing on the infection, is time dependent. The longer a mosquito lives, the more the virus replicates and is able to migrate to reservoirs, increasing the chance that a bite will transmit the virus. We assume mosquito → ruminant infectivity is 21% at 7 days, 58% at 14 days, and that mosquito → human infectivity is static at 1%.

Range and Movement. The simulation assumes that at startup, mosquitoes in the appropriate numbers are distributed randomly and uniformly across the simulation space. Mosquitoes are not constrained by roads or water features. As the simulation progresses, the mosquitoes change heading randomly ensuring a reasonable likelihood of encountering a source or a target of infection. Affinity for standing water is not modeled here.

Human Population. The human populations assumed in the MLAB-ST simulation were Boston at 600,000, Peterborough at 6,000, and Tyngsborough at 11,000. The number and characteristics of humans in the simulation is based on the unadorned numbers available from census resources. Given the population and the household number, age range data,

household (night time) networks and work/school (daytime networks) can be created. How important these networks are for this particular simulation is questionable since the disease is not spread human to human. However, these types of networks would be critically important for other disease models, and it is part of the basic design of the MLAB-ST model so that it may be used in other vector-borne disease scenarios.

Age. Age is not used as a parameter in this simulation. For Rift Valley Fever, it was not important to assign each agent an age because everyone in the United States was assumed to be equally susceptible, but for simulations in general, age is a powerful variable. Stay at home spouses, young children, and retirees would have daytime networks different from individuals working outside the home and from each other. School age children would have school networks with other children and some adults in a ratio of ~20 to 1. In some disease models there would also be age variability in susceptibility.

Gender. This parameter is not used for this simulation. This variable is possibly significant for susceptibility in some diseases, but not relevant for RVF.

Susceptibility. Susceptibility is defined as the percent of the population which has never been exposed to the disease and is therefore not immune. There may be other conditions which convey protection in other disease processes, e.g., in diseases where men are more susceptible than women, younger individuals more susceptible than older individuals, etc. For the purpose of this simulation, the susceptibility is set at 100% of the population in each community which is far above what the susceptibility would be for most agents in reality.

Percent chance of getting disease after mosquito bite. For this simulation, mosquito → human infectivity rate is static at 1%, which is the probability of a successful bite with a high enough viral titer to produce disease.

Percent chance of getting disease after ill ruminant contact. In this simulation, a non-vector route of infection occurs, but requires significant exposure to abortion products or other animal tissues and body fluids (e.g., through broken skin or aerosols) in humans who work closely with ill animals.

Latent period. Latency represents the time after an exposure occurs but before symptoms appear and individual becomes ill and infectious (2-6 days). The simulated agents continue the normal network connections during this timeframe to assure the most conservative approach. A more realistic model would have individuals take “sick days” and interact with fewer people during the symptomatic and infectious period.

Infectious period. The viremic or infectious period of the human corresponds to the timeframe when a mosquito could feed on a human having a high enough titer in the blood such that the mosquito acquires the virus. The range for RVF virus and Aedes canadensis, in this simulation is 3-4 days.

Range and Movement. The simulation assumes that initially humans, in the appropriate numbers for each community, are distributed randomly and uniformly across the simulation space. Humans are constrained by roads and water features. As the simulation progresses, humans change heading randomly ensuring a reasonable likelihood of encountering a source or a target of infection. As stated above, networks are not modeled in this situation.

Mortality. Assume 0.9% of infected humans suffer the hemorrhagic form of the disease and 50% of those cases die (APHA 2004).

Ruminants

Population. Ruminant numbers were derived from estimated total ruminants in the counties where the communities of interest are found (Massachusetts and New Hampshire State and County Data-2002 Census of Agriculture -Vol. 1, Geographic Area Series, Part 21); 10% of each total ruminant number was placed within approximately one mile radius of proposed facility site in each of the three communities. These numbers were set for Boston at 0, Peterborough at 347, and Tyngsborough at 369.

Number of human contacts. See discussion for humans above.

Percent chance of getting disease after mosquito bite. For this simulation, we assume mosquito → ruminant infectivity is 21% at 7 days, 58% at 14 days (Gargan, et.al., 1988).

Susceptibility. The literature shows that susceptibility of sheep to RVF is ~100%, cows ~85%, other ruminants not applicable for this simulation. Assume 100% for all types of ruminants present in U.S. to be conservative.

Latent period. Shorter than for humans, based on animal age/type. The range used in this simulation is 12 hours to 3 days.

Infectious period. This viremic period corresponds to the timeframe when a mosquito could feed on a ruminant and acquire the virus and the range for RVF virus and the *Aedes canadensis* mosquito is assumed to be 1-7 days. Many young animals may die in the first 3 days after infection, so this period would end soon after they died, though a mosquito could feed on a dead animal for a period of time.

Range and Movement. The MLAB-ST simulation assumes that initially, ruminants in the appropriate numbers are distributed randomly and uniformly across the roughly-circular region surrounding the proposed facility in the simulation space. Ruminants are further constrained by roads and water features. As the simulation progresses, ruminants change heading randomly ensuring a reasonable likelihood of encountering a source or a target of infection.

Mortality. For simplicity it was assumed that 10% of infected ruminants die. In nature, RVF is nearly 100% fatal for all ruminant fetuses, up to 90% mortality occurs in young sheep (a few weeks old), 20% mortality in adult sheep, and roughly 10% fatal in adult cows.

Simulation Output

Sample simulation output is provided below. In this simulation:

1. Gray dots are infected, then recovered ruminants
2. Blue dots are recovered humans
3. There were no human deaths at this low population number (6,000), and only 2 ruminant deaths. This does not count ruminant abortions.
4. The steep ruminant graph shows that a very low number of infected mosquitoes is required to infect the herd. The rate of infection is constant across a wide range of mosquito densities. Even at day 200, the low density can still infect remaining ruminants.
5. The flat human graph is not sensitive enough show the <1% recovered humans.
6. The mosquito graph follows the hatching graph assumed above.
7. Mosquitoes that hatch from over wintered infected eggs do not survive the rest of the second season (0 infected at end of simulation) because no amplification host is available (all susceptibles have become infected; recovered and no new stock is added to the population). This represents an artificially closed population for the purpose of controlling variables and comparing results in the three communities of interest.

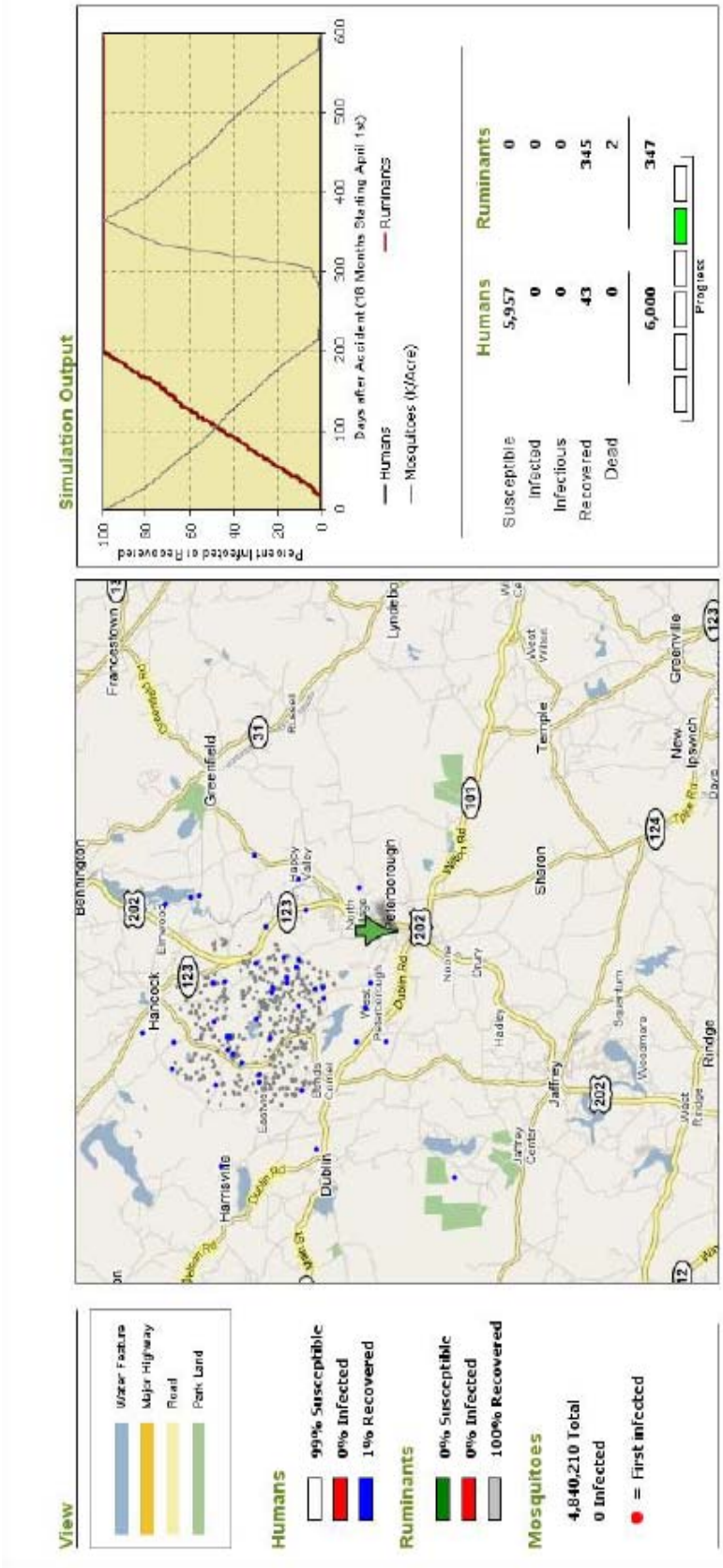


Figure VI-7. Example MLAB -ST Simulation Output