

## **Generating Human Terrain for Agent-Based Models**

**William D. Wheaton, Matthew J. Westlake, James I. Rineer, Geoffrey A. Frank, Brandon Bergenroth**

**RTI International**

**Research Triangle Park, NC**

**wdw@rti.org, mwestlake@rti.org, jrin@rti.org, gaf@rti.org, bbergenroth@rti.org**

### **ABSTRACT**

Agent-based models are a powerful tool for understanding the ways in which human populations may react to changes in their environment, social structure, or ruling body policies. Agent-based models allow researchers to measure the impacts of policies such as (1) closing schools, workplaces, and churches, (2) travel restrictions (e.g., roadblocks, border closings), and (3) distribution of relief supplies.

Agent-based models assign behaviors and activities to agents (i.e., individuals) within the population being modeled and then allow those agents to interact with the environment and with each other in complex simulations. To accurately predict population responses, agent-based models depend upon geospatially and demographically accurate population databases.

This paper describes lessons learned in generating a synthesized population database that represents the entire U.S. household population. An iterative proportional-fitting method was used to generate a synthesized, geospatially explicit human-agent database. The synthesized agent database locates individuals within households, places the households relative to census geography, and provides demographic attributes consistent with the census data. The demographic data supports modeling of interactions through social networks. This database was used for “what-if” studies of government responses to influenza epidemics.

This paper describes extending this methodology and these processing steps to generate synthesized agent databases in other parts of the world. We present the results of an effort to create a realistic synthesized agent database for Mexico, which could then be used in agent-based models to study the effects of potential economic and health policy on cross-border migration. The resulting synthesized population database is geospatially explicit and accurately represents the actual household population of Mexico. As more census microdata become available, synthesized population databases such as the one described in this paper become possible for more countries.

### **ABOUT THE AUTHORS**

**William Wheaton** is the director of the Geospatial Science and Technology program at RTI International<sup>1</sup> (RTI). Mr. Wheaton leads the development of RTI’s U.S. synthesized human agent databases and supports the integration of synthetic population models used to study pandemic flu mitigation strategies.

**Matthew Westlake** is a statistician at RTI, where he applies his knowledge of genetics, protein data, and bioinformatics to solving data analysis tasks, including ensuring that the synthetic population models are statistically consistent with multiple demographic data sources.

James Reneer is a geographic information systems (GIS) application developer and water resources engineer. He specializes in the design and development of Internet, desktop, and field tools and services using a variety of GIS and database technologies, application frameworks, and libraries.

**Geoffrey Frank** is a principal scientist at RTI. He has been applying Semantic Web technology ontology to assist in constructing epidemiological models and simulations. He has served on IEEE and ISO standards working groups developing standards for learning technologies.

---

<sup>1</sup> RTI International is a trade name of Research Triangle Institute.

## Generating Human Terrain for Agent-Based Models

William D. Wheaton, Matthew J. Westlake, James I. Rineer, Geoffrey A. Frank

RTI International

Research Triangle Park, NC

wdw@rti.org, mwestlake@rti.org, jrin@rti.org, gaf@rti.org

### INTRODUCTION

Agent-based models are a powerful tool for understanding the ways in which human populations may react to changes in their environment, social structure, or ruling body policies. Agent-based models assign behaviors and activities to agents (i.e., individuals) within the population being modeled and then allow the agents to interact with the environment and each other in complex simulations.

To accurately predict population responses, agent-based models depend upon geospatially and demographically accurate population databases. These population databases are the “human terrain” needed for agent-based model simulations.

This paper discusses the methods used to develop the U.S. synthesized agent database, how agent-based models are built on top of this database, an example of a synthetic agent database in Mexico, and how these models can be used in training applications.

### Agent-Based Models

An agent-based model is a computational model for simulating the actions and interactions of autonomous individuals (hereafter referred to as agents) in a network to assess their effects on the system as a whole. Agent-based models have been applied to such disparate fields as business (e.g., supply chain optimization [Lin and Lin, 2006], logistics, consumer behavior [Sallach and Macal, 2001]), traffic congestion (Hobeika, 2009) infectious agents (Eubank, 2005; Ferguson et al., 2005; Longini et al., 2007; Epstein and Axtel, 1996), and social and financial policy (Orcutt, 1986). In these and other applications, the system of interest is simulated by capturing the behavior of individual agents and their interactions. Agent-based models can be used to test the ways in which changes in the behaviors of individual agents will affect the system’s overall behavior.

### Agent-Based Models for “What If?” Analysis

The primary reason for building agent-based simulations is to answer “what-if” questions. We distinguish the following two forms of “what-if” questions:

1. *Exploration of possible interventions by the civil authorities.* An example of this type of “what-if” question is, “Which geographic and demographic subgroups should be given anti-viral treatments to prevent an epidemic?” Allowing trainees to understand the consequences of their decisions on large and diverse populations under various conditions is the predominant form of training application for agent-based models.
2. *Exploration of the possible consequences of different initial conditions.* An example of this type of “what-if” question is, “What is the likelihood of an influenza pandemic starting in Southeast Asia or Central America?”

### Data Requirements for Agent-Based Models

Agent-based models are built on microdata (i.e., data on individual agents), not on aggregated population summary data. Microdata—in which each data element is associated with a specific agent—allow researchers to specify and model important population characteristics and structures, such as family structure, that are derived from data on individuals. For example, microdata on a single household of four would contain information about age and sex and the relationship of each individual to the household and to each other. This level of detailed structure could not be derived from aggregated population data.

### An Agent-Based Model of the U.S. Population

RTI International (RTI) developed a U.S. agent database that was synthesized from existing national datasets. This database contains more than 104 million U.S. households and more than 270 million individuals representing the U.S. household population in 2000 (Wheaton et al., 2009). It supports agent-based simulations by providing the necessary agents that form the basis of a simulation. (Cooley et al., 2008). For example, using the synthesized population database in a model studying the spread of seasonal influenza in North Carolina, the simulation model predicted the spread of influenza by calculating probabilities of disease spread between individuals. The synthesized population database, along with the associated school and workplace assignments, provided the agents for the model and the baseline data for how individuals may come into contact with each other at school or at work. Figure 1 illustrates a portion of the model results, showing the percentage of individuals who would be ill on Day 85 of a simulated influenza outbreak in central North Carolina. Yellow, orange, and red dots represent low, medium, and high percentages, respectively. As shown in the figure, such an outbreak would be concentrated in the western part of the study area.

### SYNTHESIZED AGENT DATABASE FOR THE U.S. POPULATION

The U.S. synthesized population database contains a record for each household and separate records for each individual in the United States. These records identify individual occupants for any given household, as well as maintain characteristics of each individual, and links between individuals and their respective households. If all of the synthetic households in a geospatial area (e.g., census tract, county, state) were aggregated, the resulting synthetic population would be consistent with the census data for that area.

Figure 2 shows a map of a portion of the geospatial database representing an area about 2 miles wide; the black dots represent the synthesized households in that area. Figure 2 also shows a table presenting characteristics of a single household, as well as the links between the household and its four occupants (i.e., individuals). The database contains information on geographic location, household attributes (i.e., number of individuals in the household, individuals under the age of 18, household income, and number of workers in the family), age, and sex for these four individuals.

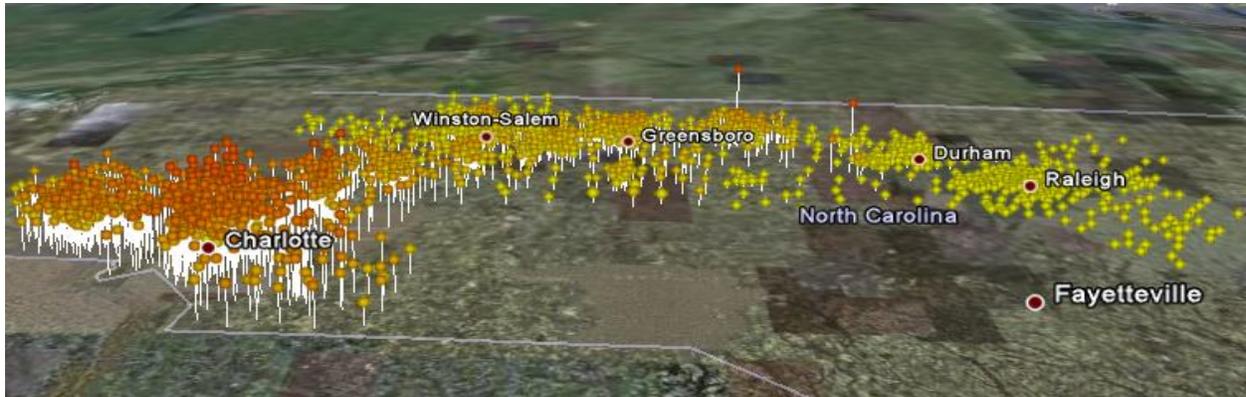


Figure 1. Visualization of Disease Spread Based on Synthetic Population Data

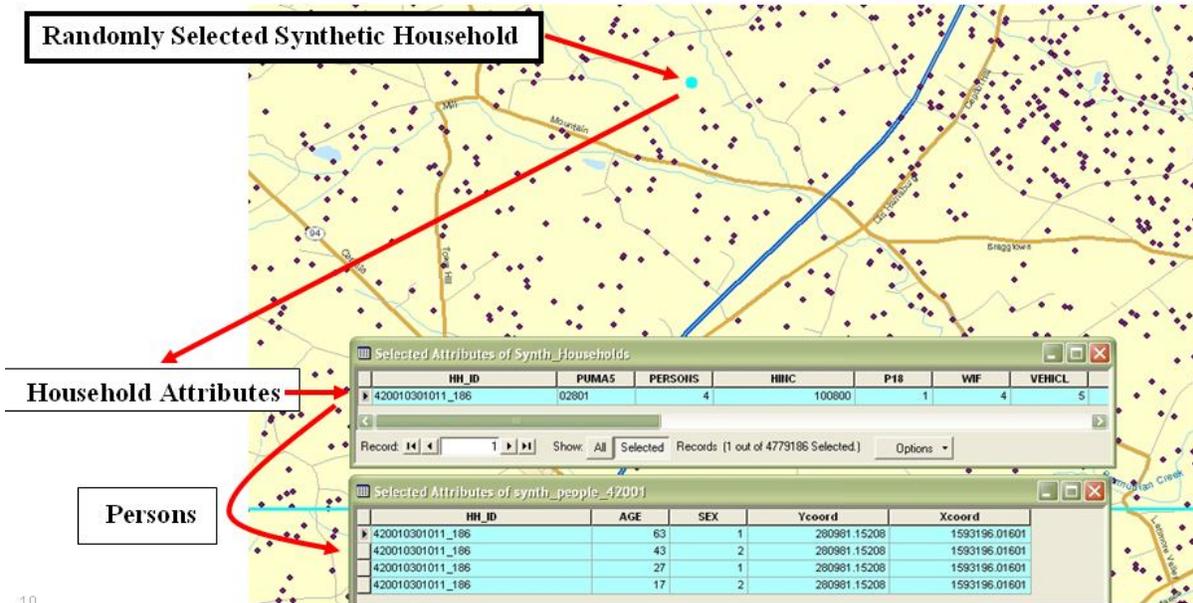


Figure 2. Conceptual Diagram of the Construction of the U.S. Synthetic Database

## METHODS

This section describes the main processing methods that we developed to generate the synthesized agent database. The overall process includes the following four steps:

1. Generating and positioning the basic synthesized households
2. Populating those households with agents
3. Assigning school-age agents to schools
4. Assigning adult agents to workplaces.

### Data Sources

We used the following three primary data sources, all of which are produced by the U.S. Census Bureau, to generate the synthesized agents and households:

#### U.S. Census Bureau TIGER Data

The Topologically Integrated Geographic Encoding and Referencing (TIGER) data provide the spatial context for decennial census data collection (U.S. Census Bureau, 2005a).

#### Summary File 3 Data

The Summary File 3 (SF3) data contain the demographic variables from the census and is organized and aggregated to many different geographic boundaries (U.S. Census Bureau, 2005b). Data variables on population and housing are available in these files.

### Public Use Microdata Sample

The Public Use Microdata Sample (PUMS) data contain records representing a 5 percent sample of the occupied and vacant housing units in the United States and the individuals in the occupied units (U.S. Census Bureau, 2005c). These data are actual responses to census long-form questionnaires and, therefore, retain family structure information, including household data, such as number of individuals in the household, number of bedrooms, age of building, access to telephone service, type of heating, mortgage data, and many other variables. The data that PUMS provides on individuals includes an individual's age, sex, ethnicity, language spoken, school enrollment, occupation, travel time to work, military service, and many other variables. Additionally, the PUMS dataset maintains linkages between individuals and households, which allow the household population structure to be brought forward through further analyses.

### Synthetic Household Generation Methods

The generation of synthesized households and their resident agents included the following basic activities:

- Generating household locations
- Generating microdata records for all households.

We also compared the synthetic population to census counts as a quality control measure.

### **Generating Household Locations**

Each household in the database is represented as a GIS point feature, which has a unique  $x,y$  location and containing descriptive tabular attributes. Because no national database of household locations is available, we generated point features to represent the location of each household. We determined the number of points to be generated within each census block group by the count of households in the block group according to SF3 data. A dasymetric redistribution approach was used to locate each household within a census block group to conform to the population distribution at a 90-meter grid cell-resolution provided by LandScan population distribution data. Figure 3 illustrates the dasymetric method.

### **Generating Microdata Records for All Households**

Because the microdata are available for only 5 percent of U.S. households, we needed to devise a process to replicate the PUMS household data records to generate microdata records for 100 percent of the households represented by the previously described point features. We did this by using a statistical technique known as iterative proportional fitting (IPF) (Beckman et al., 1996).

To carry out the IPF procedures, we used an existing computer program, the Population Generator, which was developed at Los Alamos National Laboratory as part of the TranSims (TranSims Open Source, 2008) transportation simulation modeling package. The Population Generator implements a procedure developed by Beckman and colleagues (Beckman et al., 1996) to generate synthetic populations from the three previously described U.S. Census datasets (i.e., TIGER, SF3, and PUMS) and the cross-walk table that associates PUMS with census block group polygons. Beckman and colleagues (Beckman et al., 1996) fully describe the details of the IPF procedure used in the TranSims Population Generator.

The IPF procedure selects records from the 5 percent sample of households contained in the PUMS data to represent households in a particular census block group such that when the assignment process is complete, 100 percent of the known households in a block group are represented. Because TranSims uses

specific criteria to assign households to block group areas, we expect the synthetic population to have the best fit for the following population characteristics:

- Number of people in the household under the age of 18
- Household income
- Household size
- Household population
- Vehicles available.

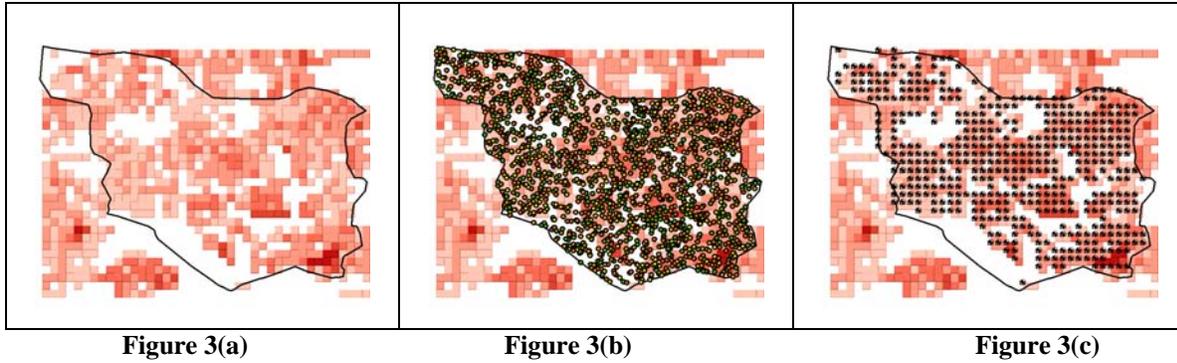
Different synthesized populations can be constructed by varying the matching variables. For example, if “language spoken” were a key characteristic in an agent-based model, then a synthesized dataset could be built to include it as one of the matching variables.

### **ASSIGNING AGENTS TO SCHOOLS**

Infectious disease transmission is known to occur at a higher rate in schools because of the relatively close and sustained contact between students (Longini and Halloran, 2005). Therefore, the synthesized agent database must assign school-age individuals to schools to enable the explicit modeling of social contacts between agents in school settings.

We used different methods to assign students to public and private schools to reflect the different geographical processes inherent in enrollment decisions for public versus private schools. We determined how many students should be assigned to public or private schools in each area based on codes found in the PUMS data.

The public schools assignment method is based on the assumption that public school students are enrolled at the closest school having adequate capacity. This assumption is necessary because no national data source of school catchment areas exists. Even within school districts, there may be magnet schools, which draw students from throughout a county or city, or there may be busing or other school assignment methods that are not based on proximity. However, in the absence of reliable data on these factors, we believe the minimum distance assumption for school assignments is valid and represents an appropriate simplification that allows students to be assigned to schools in a nationally consistent manner.



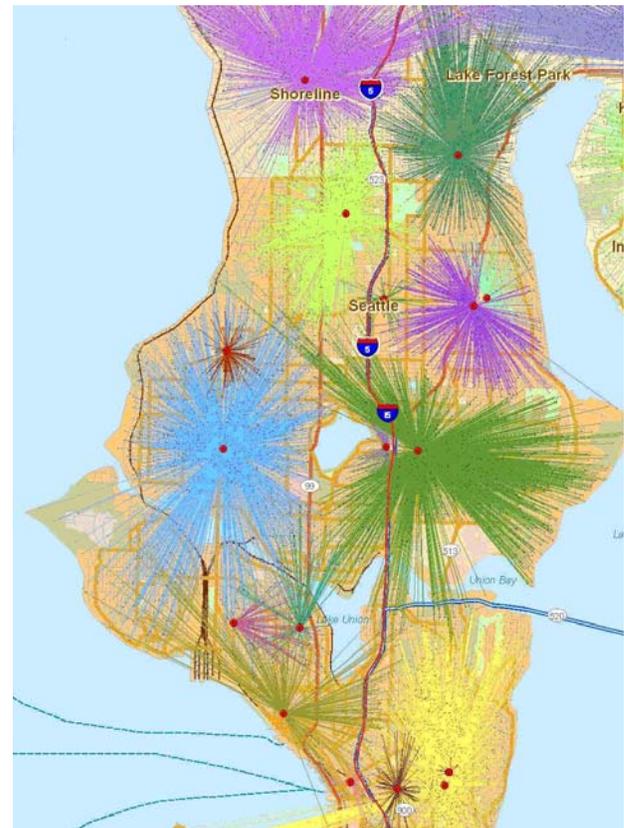
**Figure 3. Positioning of Households to Match LandScan Population Data; (a) LandScan 90-meter population, (b) random location of households, (c) dasymetrically relocated households based on LandScan distribution**

Figure 4 illustrates the spatial distribution of high schools in a portion of Kings County, Washington, and the allocation of high school students from the synthesized agent database to those schools. Each dot on the map represents the location of a high school. For each student assigned to a high school, a line is drawn between the location of the student's household and the school. The figure illustrates the results after all high school grade assignments have been run. Because the process is run grade by grade, in some cases students seem to be drawn from other schools' catchments. In effect, each grade has its own catchment, which results in assignment overlapping. If a given school's grade-by-grade capacity has not been met and unassigned students are available at greater distances, the algorithm will assign those students to that school even if they are closer to a different school, providing that the closer school's capacity has already been met.

After public-school students have been assigned, we implemented the private-school assignment method on school-age individuals not assigned to public schools and not reserved as homeschooled. Catchment areas for private schools are broader than for public schools; therefore, the private-school method allows students to be assigned to a private school that is not necessarily in the same county. Because private schools draw students from across county borders, we cannot process these assignments one county at a time, as we did with public-school assignments. Instead, we did private-school assignments one state at a time. This allowed the schools to draw from a more natural distribution of potential students.

Individual choice is a major factor affecting whether students attend private schools and which specific private schools they might attend. These variables are difficult to account for in assigning a synthetic

population to private schools. We assumed that although distance is important in private school assignments, it is less critical than in public school assignments. Therefore, distance was still the key criteria in assigning students to private and parochial schools but, unlike the public school-allocation method, private school students were not constrained to the nearest private school with capacity. Instead, a concentric-ring approach was used to assign students to private schools.



**Figure 4. Distribution of High Schools and Linked Students**

## **ASSIGNING AGENTS TO WORKPLACES**

To identify individuals and locations where infectious disease may be transmitted, we needed to be able to assign the non-school-age population to workplaces.

We developed a two-stage method to assign synthesized agents to workplaces. The first stage assigned workers to a census tract. The second stage created individual workplace locations within each census tract and assigned workers to specific workplaces.

The Census Bureau's Special Tabulation 64, STP64 (U.S. Census Bureau, 2004), data contain records for each combination of census tract of residence and census tract of work for which individuals from one census tract work in the same or another census tract. We assigned a set of synthesized agents from each census tract of residence to a census tract of work as specified by the STP64 data. For example, if the STP64 data indicated that 50 individuals living in tract A work in tract B, then 50 agents living in tract A were coded with a work identifier of tract B.

After assigning agents to a census tract of work, we ran a process to generate specific workplaces that meet the workplace counts and sizes in the InfoUSA data. For each census tract, we generated a record for each business located within the tract.

This analysis resulted in a workplace identifier for each individual assigned to a place of work. The resulting data table contains a record representing each business workplace—along with the number of employees who work there—found in the census tract. Each workplace is placed at the center of the census tract for this analysis.

After determining which workers work in each census tract and generating records for each workplace in each tract, we assigned the synthesized agents, working in a given census tract, to specific workplaces. We used the specified capacity of each workplace site to determine how many workers should be assigned to that site.

At the conclusion of the workplace assignment process, we assigned agents to individual workplaces to meet two main criteria: (1) each agent works in the correct tract, as specified in the STP64 data, and (2) each agent is assigned to a workplace such that the distribution and capacity of each workplace within each tract is honored. Agents who work in the same workplace have the same workplace identifier;

therefore, developers of agent-based models know explicitly which workers may come into contact with each other based on their workplace assignments. This method does not account for workers working different shifts or for telecommuters.

## **INTERNATIONAL SYNTHETIC POPULATIONS**

In response to the recent interest in H1N1 influenza outbreaks in Mexico, we developed a process for generating synthetic populations for other countries and used Mexico as a test case.

The data sources needed to generate a synthetic population database in foreign countries include public use microdata, census data, LandScan population distributions, and geospatial administrative boundary data.

### **Steps in Developing the Mexican Synthetic Population database**

The International Public Use Microdata Organization (IPUMS) ([www.ipums.org](http://www.ipums.org)) provides public use microdata for a number of countries. The public use microdata for Mexico from the 2005 census were available from [ipums.org](http://ipums.org). This data counted people by age. Census counts of households and individuals at the Mexican State geographic level for the 2005 census were also available from various sources on the Internet, including from the Mexican government. LandScan 2005 data contains a gridded population dataset for the world at a 1KM grid cell resolution.

We developed a weighting and matching process to select records from the public use microdata set such that, for each Mexican state, the total number of households and the total number of individuals in each age group matched the counts given for the Mexican census.

The matching process resulted in (1) a text file containing a record for each household in Mexico coded by State and number of individuals in the household and (2) a file containing one of over 103 million records for each individual in Mexico. Each file contains the age and sex of an individual along with a key to link that individual back to a particular household.

To place the households on the landscape, we developed a dasymetric process for locating them within the LandScan grid cells such that the count of individuals from the synthetic data within the grid

cell matched the LandScan population totals for the grid cell.

After calculating the grid cell location for each household, we used a randomized positioning process to place the household randomly within the grid cell. We then linked the household data records to the household geographic points to complete the dataset.

The resulting synthetic dataset for Mexico has been prepared to construct agent-based models for “what-if” analysis of the potential effects of mitigation strategies on swine flu in Mexico. Additional efforts for modeling epidemics in Mexico include constructing the daily movement patterns for the agents. This effort will combine the methods developed for large urban areas (Germann et al., 2006; Halloran et al., 2008) with methods developed for agent-based models of rural populations in Southeast Asia (Longini et al., 2005).

The same process used to develop the Mexican database can be applied to any country for which a public-use microdata sample is available. As of this writing, the ipums.org datacenter has microdata for 43 countries.

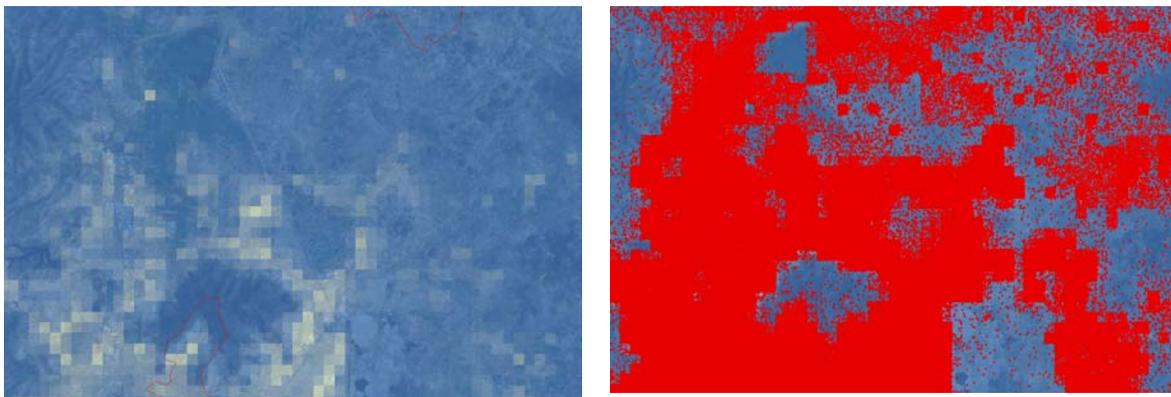
### **Building Agent-Based Simulations from Synthetic Populations**

Large-scale agent-based simulations have been used to support planning for responses to potential epidemics such as avian flu (Cooley et al., 2008). In these models, each individual in an urban area is represented as an agent. The location and attributes of each agent are determined by the synthetic population model. A simulation then models the behavior of these agents over a period of days and months and

captures critical interactions between agents (e.g., the transmission of an infectious disease). These simulations also model the ways in which the agents react to various interventions, including social distancing, travel restrictions, medical treatment, and vaccination.

For studies of human-transmitted diseases, the agent-based models are built using normal-behavior and sick-behavior models for the populations. These behavior models for the agents represent the interaction patterns of the people in the area being modeled and are typically state transition models. Different models are available for different demographic groups. For example, if the synthetic population model specifies the age of each agent, then an age-appropriate behavior model will be used. Age-appropriate normal behavior models for the United States will have infants staying at home, school-age children attending school, working-age adults commuting to work, and the elderly staying at home or in retirement communities or assisted living facilities. The transition rates for state changes in the model are parameters. Some of these parameters are calculated using census data or other sources. For example, agent commute times may be calculated from census data about the location and number of workers at workplaces in the region. Similarly, the time between becoming infected and displaying symptoms may be estimated from medical evidence for appropriate diseases.

Some parameters are estimated by a calibration process whereby selected model outputs are matched against historical data. For example, infection rates within and between age groups may be calibrated based on epidemiological evidence identifying where infections occur (e.g., household, workplace, school).



**Figure 5. Gridded Distribution of Population (left) and Associated Household Locations (right)**

## **Training Applications for Agent-Based Simulations**

The primary training application of agent-based simulations is to help the trainee appreciate the consequences of various interventions on large host-nation populations. Agent-based simulations of epidemics have been used to study the appropriate use of limited stockpiles of vaccines and antiviral medicines in combination with social-distancing interventions such as closing schools and religious institutions, quarantining households and hospitals, and restricting travel (Longini et al. 2005, Ferguson et al., 2006). These social-distancing interventions are relevant to Security, Stability, Transition, and Recovery (SSTR) missions. Agent-based simulations have been used in training to clarify roles of healthcare workers and to identify critical healthcare infrastructure as well as to assess requirements for stockpiles of healthcare supplies (Longini et al. 2005, Ferguson et al. 2006). An important aspect of these simulations is the definition of triggering events or thresholds for interventions and the surveillance methods used to detect when the triggering events have occurred. For example, triggering events for interventions to prevent influenza epidemics that are based on healthcare institution reporting can be confused by seasonal influenza case reporting (Bobashev et al., 2008).

## **Future Applications of Agent-Based Simulations Using Synthetic Populations**

Existing synthetic population models of developed countries use sophisticated census and survey data to ensure accurate modeling. The lack of these data in developing countries presents a challenge that we are currently addressing by exploring alternatives to census data as a means of collecting geo-referenced data.

Epidemiological simulations sometimes require information about populations other than humans. For example, various strains of influenza can be carried by birds (e.g., avian flu) or pigs (e.g., swine flu). [BW: this last sentence needs to be checked...Julie—I'll confirm.] In these cases, synthetic animal populations should be developed to support simulation of transmission of disease to humans from other hosts.

Agent-based models hold promise for supporting “what-if” analyses of the political, economic, and social consequences of many interventions commonly employed in SSTR missions. These include social-distancing interventions such as

curfews and road blocks and the closing of schools, workplaces, and religious institutions. The long-term economic side effects of workplace closures and absenteeism due to illness have been analyzed using epidemiological models (Bobashev et al, 2008).

Analysis of social consequences of interventions will depend upon access to data about the existing social networks that can be integrated into the synthetic population databases. Also, relevant historical studies and supporting demographic data are needed to calibrate the models and validate their potential predictive accuracy.

## **CONCLUSIONS**

Synthetic population models that accurately reflect the demographics of the population of interest are essential for building agent-based simulations for predicting how populations will react to interventions. Similar to the virtual terrain for land-battle simulations, synthetic population models serve as the “human terrain” for models of large populations.

The design of agent-based models for training or mission rehearsal requires tradeoffs that balance the availability and quality of the demographic data with the range of interventions that are possible in the training. The agent-based models must incorporate behaviors that distinguish the interventions. However, the state transition rates for these behaviors must be calibrated with demographic evidence to ensure a valid training experience.

This paper describes statistical techniques for ensuring that a synthetic population model accurately reflects the demographic data provided from many sources. These techniques include geographically locating households, populating the households with appropriate numbers of agents, and calculating the commuting schedules, workplace interactions, and school interactions of the populations. These techniques have been used for both large urban populations and rural populations.

## **ACKNOWLEDGEMENTS**

The project described was supported by grant number U01GM070698 (Models of Infectious Disease Agent Study—MIDAS) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute of General Medical Sciences or the National Institutes of Health.

## REFERENCES

- Beckman RJ, Baggerly KA, McKay MD. Creating synthetic baseline populations. *Annals of Transportation Research*. 1996;30(6):415-29.
- Bobashev G, Goedecke M, Morris R et al. Distribution strategies for the limited amounts of vaccines in the context of pandemic influenza. *Presented to the World Health Organization*, October 21, 2008.
- Cooley P, Ganapathi L, Ghneim G, et al. Using influenza-like illness data to reconstruct an influenza outbreak. *Mathematical and Computer Modelling*. 2008;48(5,6):929–939.
- Epstein JM, Axtell R. *Growing artificial societies: social sciences from the bottom up*. Cambridge (MA): MIT Press; 1996.
- Eubank S. Network-based models of infectious disease spread. *Jap J Infect Dis*. 2005;58(6): S9-13.
- Ferguson N, Cummings DAT, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*. 2005;437:209–214.
- Ferguson N, Cummings DAT, Fraser C, et al. Strategies for mitigating an influenza pandemic. *Nature*. 2006;442:448–452.
- Germann TC, Kadau K, Longini IM Jr., et al. Mitigation strategies for pandemic influenza in the United States. *PNAS*. 2006;103(15):5935–5940.
- Halloran ME, Ferguson NM, Eubank S, et al. Modeling targeted layered containment of an influenza pandemic in the USA. *PNAS*. 2008;105(12):4639–4644.
- Hobeika, 2009. May, 2009. “TRANSIMS - the next generation planning/simulation model.” *Transactions of the Wessex Institute*. 10.2495/UT040151.
- Lin F, Lin S. Enhancing the supply chain performance by integrating simulated and physical agents into organizational information systems. *Journal of Artificial Societies and Social Simulation*. 2006;9(4):1.
- Longini I Jr, Nizam A, Xu S, et al. Containing pandemic influenza at the source. *Science*. 2005;309:1083–1087.
- Longini IM, Halloran ME, Nizam A, Yang Y, Xu S, Burke D, et al. Containing a large bioterrorist smallpox attack: A computer simulation approach. *Int J Infect Dis*. 2007 Mar;11(2):98-108.
- Longini IM, Halloran ME. Strategy for distribution of influenza vaccine in high risk groups and children. *Am J Epidemiol*. 2005;161:303-6.
- Orcutt GH, Mertz J, Quinke H, editors. *Microanalytic simulation models to support social and financial policy*. Amsterdam: North-Holland. 1986.
- Sallach D, Macal C. The simulation of social agents: an introduction. *Special Issue of Social Science Computer Review*. 2001;19(3):245-8.
- TranSims Open Source [homepage on the Internet]. Los Alamos (NM): Los Alamos National Laboratory; 2008. *Transportation Analysis Simulation System (TRANSIMS)*. Originally available from Los Alamos National Laboratory; now available from: <http://transims-opensource.org>
- U.S. Census Bureau: Department of Commerce, Economics and Statistics Administration. *Census 2000 special tabulation: census tract of work by census tract of residence (STP 64)* [Internet]. Washington, DC: US Census Bureau; 2004 April. [cited 2008 March 6]. Available from: <http://www.census.gov/mp/www/spectab/stp64-webpage.html>
- U.S. Census Bureau; Department of Commerce, Economics and Statistics Administration. 2000. *Topologically Integrated Geographic Encoding and Referencing (TIGER) system*. Washington, DC: US Census Bureau; 2005a.
- U.S. Census Bureau; Department of Commerce, Economics and Statistics Administration. 2000. *Census of population and housing, summary file 3*. Washington, DC: US Census Bureau; 2005b March.
- U.S. Census Bureau; Department of Commerce, Economics and Statistics Administration. 2000. *Census of population and housing, public use microdata sample 2000*. Washington, DC: US Census Bureau; 2005c Dec.
- Wheaton, W.D., Cajka, J.C., Chasteen, B.M., Wagener, D.K., Cooley, P.C., Ganapathi, L., Roberts, D.J., and Allpress, J.L. (2009). *Synthesized Population Databases: A US Geospatial Database for Agent-Based Models*. RTI Press publication No. MR-0010-0905. Research Triangle Park, NC: RTI International.