Farm Animal Demographics Simulator Aids in Disease Modeling

Written by Ken Geter, CEAH GIS

A significant limitation in developing animal disease simulation models is the lack of data on locations of individual farms and on the numbers and types of animals on each farm. The National Agricultural Statistics Service (NASS) Census of Agriculture provides data on numbers of farms and animals by county, as well as numbers of animals by Zip Code. However, because of privacy concerns, information on individual farms is not publicly available.

The farm animal demographics simulator creates agriculturally and geographically realistic data sets based on the NASS data for disease simulation models. The simulator uses NASS data to estimate the number and size of animal facilities within a geographic region (e.g., counties or States). Over the region of interest, a geographic constraint is applied so that facilities are not located in urban areas, lakes, parks, natural areas, or other locations where they are unlikely to occur. Facilities are located randomly, unless a weight matrix is used to concentrate certain types of farms closer to urban areas, live animal markets, and feed suppliers. Numbers of animals are randomly assigned to the simulated facilities in accordance with the NASS data.

Because States are not required to report their animal numbers to NASS, there is a fair amount of missing data in the publicly available Census of Agriculture. Part of the function of the farm animal demographics simulator is to estimate the missing county-level data.

Another use of this type of simulated farm data would be for emergency management training scenarios. In these tabletop exercises, there is a need for agriculturally realistic data that do not violate privacy concerns.

Methods

The farm animal demographics simulator is currently a system of manual processes using ESRI ArcGIS software, in conjunction with a privately developed software extension, and Python programming algorithms. In the near future, the simulator will be developed into an application to provide data for inputs into the Avian Influenza Modeling Project at CEAH.

Geographic Masking

The first step in simulating farm locations is to create a geographic mask of probable locations. Areas such as lakes, rivers, parks, federal lands, or urban areas are masked out as possible farm locations. The mask layer also typically constrains farm locations to be within a given distance of roads. In the example map of Minnesota beef cattle, farms are within 300 meters of roads. The mask layer is generated in ESRI ArcMap, using the geoprocessing Model Builder tool.

This step is generally the most time-consuming part of the process requiring geospatial technical expertise to accomplish. The mask generation is computationally intensive and can take several days to run and debug, even on a high-end dual-processor workstation computer. States with dense road networks, e.g., eastern seaboard States, take longest since the masking procedure places a distance buffer on all the roads within a State. States with dense road networks can be subdivided into regions to optimize processing.

Generation of Simulated Farm Locations

The simulated farm locations are generated within the constrained area of the mask layer using the "Generate Random Points" tool in the "Hawth's Tools" ArcMap extension (see http://www.spatialecology.com/htools/tooldesc.php). This tool places points at random anywhere within a particular polygon layer, in this case within the mask layer.

Assigning Farm Sizes and Animal Populations

NASS usually categorizes farms according to size. The size categories correspond to the number of animals on a farm. Total numbers of farms per category and animals per category are reported in the Census of Agriculture. The Python programming language (http://www.python.org) is used to first randomly assign a size category attribute to each of the simulated farms generated in the step above. Then the algorithm randomly attributes a number of animals to each farm within the size category range. The process takes into account the total number of farms and the total of animals in each size category for a county, so that the sum of the assigned farm sizes and animal populations equals those totals.

Outputs

The main outputs of the farm animal demographics simulator are maps and data sets for input into animal disease modeling projects such as the Avian Influenza Modeling Project at CEAH. An example map of beef cattle in Minnesota is shown. After the mask layer is generated, multiple randomizations of farm locations and farm sizes can be performed, as these processes are the least time- and computer-intensive to run. Therefore the modelers can generate any number of data sets and maps for input into their analyses. Data sets and maps generated from the simulator could also be used in emergency management tabletop exercises.

Missing Data

In order to protect the privacy of farms in counties with small numbers of farms, States frequently do not report numbers of farms or animals for those counties. This presents a problem for the simulator, since the Python algorithm requires county totals to accurately assign the farm size and number of animal attributes to each farm. There are several ways to estimate the missing value counties including distributing the difference between the known values and the State totals or by interpolating values from the neighboring

counties. These estimations usually require decisions to be made by an analyst, and generally cannot be made automatically by computational processes.

Mask Layer Generation

This part of the simulator process is the most computationally intensive. It usually requires several iterations to create the mask layer for a given State, depending on the density of roads and water features. Optimization of this process requires further examination: certain input parameters of the process may need to be changed in order to find the best balance between processing time and simulation accuracy. Again, for a given State, these decisions probably need to be made by an experienced analyst. However, once a mask layer is generated, it need not be changed for further randomizations.

Validation of the Simulated Data

Ideally, feedback on the distribution of the simulated farm placements and animal demographics is necessary. Such feedback may be obtained through cooperation with State and local officials, or possibly NASS staff. It would be most useful in validating methodologies to estimate data in the counties with missing values.

Application Development

The farm animal demographics simulator is currently in the process of being contracted for development as a software application. It will be used as a component of the Avian Influenza Modeling Project to provide analysts with data needed to run simulation models. In its final form, the application may be an ArcGIS toolbar extension, a stand-alone desktop application, a Web application, or a combination of these options. The challenges to be addressed include those listed above, i.e., estimating missing data and optimization of mask layer generation. The application will undoubtedly require some human-entered parameters and interaction in order to deal with these challenges, but in the end will be much more efficient than the current manual processes.

For more information on the farm animal demographics simulator, please contact Ken Geter at (970)494-7253 or e-mail to kenneth.d.geter@aphis.usda.gov.

Figure 1. Example of Farm Animal Demographics Simulator for beef cattle in Minnesota.