Synthesized Population Databases: A Geospatial Database of US Poultry Farms

Mark C. Bruhn, Breda Munoz, James Cajka, Gary Smith, Ross J. Curry, Diane K. Wagener, William D. Wheaton

January 2012

RTI Press

About the Authors

Mark C. Bruhn, BS, is a research GIS analyst in RTI International's Geospatial Science and Technology Program.

Breda Munoz, PhD, is an environmental statistician at RTI International.

James Cajka, MA, is a senior GIS analyst in RTI's Geospatial Science and Technology Program.

Gary Smith, MA (Oxon), MA (Cantab), DPhil, is Professor of Population Biology and Epidemiology at the University of Pennsylvania School of Veterinary Medicine.

Ross J. Curry, MA, is a research GIS analyst in RTI's Geospatial Science and Technology Program.

Diane K. Wagener, PhD, is a senior epidemiologist in RTI's Statistics and Epidemiology program and the Principal Investigator of the Models of Infectious Disease Agents (MIDAS) Informational Technology Resource at RTI International.

William D. Wheaton, MA, is a senior research geographer and director of RTI's Geospatial Science and Technology program.

Suggested Citation

Bruhn, M. C., Munoz, B., Cajka, J., Smith, G., Curry, R. J., Wagener, D. K., & Wheaton, W. D. (2012). *Synthesized population databases: A geospatial database of US poultry farms*. RTI Press publication No. MR-0023-1201. Research Triangle Park, NC: RTI Press. Retrieved from http://www.rti.org/rtipress.

doi:10.3768/rtipress.2012.mr.0023.1201

www.rti.org/rtipress

Synthesized Population Databases: A Geospatial Database of US Poultry Farms

Mark C. Bruhn, Breda Munoz, James Cajka, Gary Smith, Ross J. Curry, Diane K. Wagener, William D. Wheaton

Abstract

The pervasive and potentially severe economic, social, and public health consequences of infectious disease in farmed animals require that plans be in place for a rapid response. Increasingly, agent-based models are being used to analyze the spread of animal-borne infectious disease outbreaks and derive policy alternatives to control future outbreaks. Although the locations, types, and sizes of animal farms are essential model inputs, no public domain nationwide geospatial database of actual farm locations and characteristics currently exists in the United States. This report describes a novel method to develop a synthetic dataset that replicates the spatial distribution of poultry farms, as well as the type and number of birds raised on them. It combines county-aggregated poultry farm counts, land use/land cover, transportation, business, and topographic data to generate locations in the conterminous United States where poultry farms are likely to be found. Simulation approaches used to evaluate the accuracy of this method when compared to that of a random placement alternative found this method to be superior. The results suggest the viability of adapting this method to simulate other livestock farms of interest to infectious disease researchers.

Contents

RTI
INTERNATIONAL

Introduction

Because of the potential spread to humans, infections and diseases in poultry and other livestock populations are of keen public health interest. Avian influenza (H5N1), for example, is of particular concern because of its high case mortality (14–33 percent) among previously healthy children and adults (Li, Choi, Sly, & Pak, 2008; World Health Organization, 2005). Of equal concern is the potentially severe economic impact to the livestock and poultry industries, such as that caused by outbreaks of foot-and-mouth disease in the United Kingdom (UK). These outbreaks also threaten the food supply and thus are a concern for everyone (Smith & Kelly, 2008).

Infectious disease researchers are increasingly using mathematical models to simulate the transmission dynamics of infections of veterinary interest to inform policy decisions affecting specific control and containment strategies (e.g., Bates, Carpenter, & Thurmond, 2003; Bates, Thurmond, & Carpenter, 2003a, 2003b; Kobayashi, Carpenter, Dickey, & Howitt, 2007a, 2007b). In 2001, for example, British policymakers relied on mathematical models to evaluate the effectiveness of different strategies to control and contain a large outbreak of foot-and-mouth disease in the UK (Keeling, Woodhouse, May, Davies, & Grenfell, 2003; Keeling et al., 2001; Morris, Wilesmith, Stern, Sanson, & Stevenson, 2001). Containment spared livestock that might have been infected had the outbreak expanded. It also resulted in significant economic savings for animal farmers outside of the containment areas, as well as for consumers of UK animal products around the world.

These models of foot-and-mouth disease in the UK relied on the availability of data containing the locations of UK animal farms with associated animal types and numbers on each farm. The attributed farm locations were needed because many infectious disease mitigation strategies, such as culling or ring vaccination, explicitly rely on spatial relationships between susceptible animal populations and pathogen reservoirs.

In the absence of more detailed information, some researchers have examined the usefulness of models based on aggregate census records and maps consisting of farm locations uniformly distributed at defined scales (e.g., townships, ZIP Codes, or counties). The usefulness of the models turned out to depend on the end point of interest (e.g., epidemic impact, radius of the culling zone, duration of the epidemic) and the spatial scale of the map (Le Menach et al., 2005; Rorres, Pelletier, & Smith, 2011; Tildesley et al., 2010). It was clear, however, that all of the models were impaired to a greater or lesser extent by the lack of good information on the spatial clustering of farms at a fine scale.

To accurately and effectively simulate avian disease transmission and the requisite control strategies, US researchers performing simulations need good farm locations, particularly with respect to how they cluster spatially (Rorres, Pelletier, Keeling, & Smith, 2010; Rorres, Pelletier, Bruhn, & Smith, 2011). However, unlike their counterparts in the UK and many other European countries, most US researchers lack access to explicit geospatial farm data. For instance, only a handful of US states collect spatial data on poultry, swine, and cattle operations. Additionally, even when spatial data exist, they are rarely in the public domain. The National Animal Identification System (NAIS) collects detailed poultry (and other) data, for example, but access is restricted to state and federal animal health officials, and only in the event of a serious infectious disease epidemic. These data are not available for research purposes (Baca, 2004).

Furthermore, the record is and will remain incomplete because the NAIS, as originally conceived, was abandoned in February 2010. The US Department of Agriculture (USDA) announced that it would revise the prior animal identification policy and offer a new approach to achieving animal disease traceability (USDA, 2010). By abandoning a policy that might have led to an individual animal and farm identification database and moving to a state-by-state-determined policy intended to track only animals that cross state lines, there is no longer any realistic prospect of access to an accurate map of existing individual farm locations in the United States.

To circumvent issues of access to actual farm data, researchers have attempted to use public information to simulate farm and animal populations. For example, USDA researchers simulated beef and dairy farm locations and animal populations in North Carolina by randomly locating the farms within restrained areas determined by several geographic factors, such as roads and water bodies (Geter, 2006; Miller et al., 2007). Although promising, this approach resulted in local farm densities that were too high as well as an unreasonably large spatial distribution. Additionally, Engle (2006) used high-resolution aerial imagery to create poultry house identification algorithms. This method, however, created a large number of false-positive identifications, required the type of detailed data that are not yet widely available in the United States, and has only been attempted in a few small areas.

To fill this gap in the science, our study aimed to create a nationwide spatial dataset of synthetic poultry farms from publicly available data that infectious disease researchers could use as input to their infectious disease models in the United States until more accurate data are available. The final database is termed "synthetic" because, while it is realistic (i.e., synthetic farms will be located in locations suitable for commercial poultry production and have attribute values similar to actual farms raising poultry) and it is derived from actual poultry farm data, it does not contain names, addresses, or any other identifying or sensitive characteristics of actual poultry farms.

The resulting synthesized data contain the number of farm operations by poultry type and size that match aggregated counts by county. The locations of the synthesized farms are generated so that they are consistent with poultry farm siting factors—such as physical and regulatory constraints, existence of poultry industry support infrastructure, and likelihood of acceptance by neighbors—observed from real poultry farm maps from two states, Delaware and Pennsylvania.

In this report, we describe the construction of the synthesized poultry farm database, evaluate its accuracy, and discuss the implications for future research.

Methods

We used a variety of data processing steps to create the synthesized poultry-farm database, including obtaining the number of poultry farms by type and size within each county; identifying common factors that influence the siting of poultry farms; developing geographic layers using suitability scores associated with each siting factor; merging site suitability layers into a single combined suitability surface; and placing poultry operations within each county based on that surface. We expect some variability in how each type of poultry farm (i.e., broilers, layers, pullets, turkeys, and other poultry) is influenced by each siting factor, but in this exercise we assumed that the differences are negligible, as noted in the discussion.

Obtain Data on Poultry Farm Counts, Types, and Sizes of Operations, Aggregated by County

The best source of data on poultry farm counts, types, and sizes of operations, aggregated by county, is the 2002 US Census of Agriculture published by the USDA's National Agriculture Statistics Service (USDA, 2002). This census is conducted every 5 years to obtain data on farming and ranching operations and includes agricultural operation (i.e., farm) characteristics, such as the types and amount of produce grown and animals raised. Participation is required by law.

We extracted and downloaded county-level data from the Census of Agriculture's Web site (USDA, 2002). We selected the county level because it is the smallest geographical division with complete summary statistics of the relevant variables—including counts for each poultry type, number of animals, and for broilers, the number sold (because the actual number of broilers was not available).

Additional processing and summarization were required to address double counting of farms at the county level that have a combination of types (e.g., a farm may raise both broiler and layer chickens). To reconcile these data, we developed an algorithm that assigned a unique farm type to synthetic farm locations based on type categories and farm size (see the appendix).

Identify the Factors That Affect the Siting of Actual Poultry Farms

After analyzing available poultry farm data and conducting a literature review, we determined that the following four siting factors are associated with the locations of poultry farms (Berry, 2007; Goan, 2005; Sambidi, 2003; Sambidi, Harrison, & Farr, 2004; USDA, 2002):

- *Zoning and local regulations*—can influence the likelihood of finding a poultry farm at a given location in several ways:
  - Municipal versus agricultural land use—areas zoned for agricultural use have a higher likelihood of allowing commercial poultry production, whereas municipalities are more likely to have specific regulations prohibiting it.
  - Constraints on where businesses can operate— typically, commercial poultry operations are not allowed to operate on public lands.
  - Separation of potentially conflicting land uses—commercial poultry operations are not usually found in areas of high population whose occupants might object to being near a commercial poultry operation (e.g., residential areas, schools, churches), often because of odor concerns.

- *Available affordable land*—other existing or competing land uses, such as nonagricultural businesses, transportation and residential areas, can exclude or deter the establishment of poultry operations.

- *Poultry industry infrastructure*—poultry farms are more likely to be located within close proximity of supporting infrastructure and businesses, such as feed mills, hatcheries, and processing plants.

- *Physiogeographic constraints on poultry house siting*—poultry houses are not found in water bodies (e.g., rivers, lakes, and ponds) and are unlikely to be located on steep slopes.

Using these four siting factors, we specified rules to identify areas where poultry houses (which are commonly about 40 feet wide by 500 feet long for broilers) could not exist (e.g., water bodies), were unlikely to exist (e.g., in urban areas or on steep slopes), or were relatively likely to exist (e.g., agricultural areas) (Table 1). Because the poultry often reside on only a small portion of the farm on which they are located, we considered the siting of commercial poultry houses rather than the location of the entire farm.

Table 1. General factors influencing the siting of poultry houses, placement rules, and GIS data used to implement the placement rules

Siting Factor	Placement Rules	GIS Data Layers
Municipal vs. agricultural use	Assign agricultural land uses an optimum score and non-agricultural land uses scores 20% less than optimum.	Agricultural/non-agricultural land use/ land cover
	Exclude city downtown areas and assign low placement scores to the rest of the municipalities/urbanized areas.	Municipalities and urbanized areas
	Exclude areas of high-intensity development and many areas of medium-intensity development. Small, isolated areas of medium-intensity development surrounded by agricultural land had a positive correlation with larger farm operations in our sample data and were not excluded.	Developed land use/land cover
Constraints on where businesses can operate	Exclude most federal lands. Assign a low placement score to federal lands that sample farm data fell within (e.g., some national forests) or ones that covered an entire county that the US Census of Agriculture indicated had poultry farms.	Federal lands
	Exclude state and local public lands.	State and local parks
	Place at least 50 m from streams and other surface water bodies (possible human drinking water sources).	Surface waters

(continued)

Table 1. General factors influencing the siting of poultry houses, placement rules, and GIS data used to implement the placement rules *(continued)*

Siting Factor	Placement Rules	GIS Data Layers
Separation of potentially conflicting land uses	Do not place on or very near sensitive areas.	Sensitive areas (e.g., churches, schools)
Separation of potentially conflicting land uses + available affordable land	Do not place on or very near populated or developed areas (includes suburban areas not well identified by the National Land Cover Dataset).	Residential and developed areas
Available affordable land	Within agricultural land, most sample poultry farms were found preferentially in land classified as cropland by the National Land Cover Dataset, so croplands were considered a complimentary (and optimum) land use instead of a competing land use. Croplands were assigned an optimum score as compared with other land uses.	Agricultural/non-agricultural land use/ land cover
	Sample poultry farms were found on pasture to a much lesser degree, especially the middle or "core" pasture areas. We believe these areas were preferentially used by non-poultry animal operations (e.g., cattle, sheep) for grazing. Peripheral pasture areas were assigned low scores and core pasture areas were assigned very low scores.	Agricultural/non-agricultural land use/ land cover
	Exclude areas where other businesses exist that are likely incompatible with commercial poultry production.	Non-agriculture businesses
	Exclude areas where other land uses exist that are likely incompatible with commercial poultry production.	Transportation (e.g., airports, railroads)
Poultry industry infrastructure	Place near roads accessible by large trucks. This rule addresses poultry operations' accessibility to supplies (feed and chicks) and ability to export products (eggs, grown broilers).	Transportation (roads)
	Place preferentially closer to (within 50 km of) known support services (e.g., hatcheries, feed mills).	Poultry support businesses
	Place within 30 km of poultry consumers (i.e., municipalities).	Municipalities and urbanized areas
Physiogeographic constraints	Preferentially place on gentle terrain (<10% slope).	Slope
	Exclude land cover types where construction of buildings used on animal operations is not feasible (e.g., water, forests, wetlands).	Incompatible land cover

GIS = geographical information system.

Having identified the broad factors that influence the siting of poultry houses, we analyzed actual farm locations to quantify the relevance of each factor to the likelihood of finding a farm at a given site. We used three sources of actual farm locations: the Dun & Bradstreet (D&B) business database (Dun & Bradstreet, 2007), a restricted access database of poultry farms developed by University of Pennsylvania researchers, and the Delaware Environmental Navigator (DEN) (State of Delaware, 2006).

Although the D&B business database contains poultry farm locations from around the country, it contains only about 30 percent of the poultry farms reported by the 2002 US Census of Agriculture. We believe that smaller farms in particular are underrepresented. Also, the locations from this data source were mostly georeferenced by address matching; consequently, their locations were not always as precise as the locations from the other two data sources. For example, we examined a small sample of the D&B poultry businesses in Pennsylvania and Delaware and determined that most D&B locations were within 500 m of the actual poultry houses, but many were not within the boundaries of the farm.

Nonetheless, the D&B database allowed us to test and refine some of our potential siting factors that involved larger areas or distances. For example, we

identified some rare cases where a true poultry farm was located on public lands, such as national forests, and on Indian reservations, and a small percentage (less than 4 percent) of true farms that are located within municipal boundaries. These findings were contrary to expectations and allowed us to adjust our siting factor assumptions. We assigned a small, non-zero suitability score for municipal areas and public lands, allowing a small number of synthesized farms to be placed in municipal areas and within certain public lands if other suitability conditions were also favorable. We created the small non-zero suitability score for only the public lands that contained a D&B poultry farm or for public lands that covered an entire county in which the Census of Agriculture indicated poultry farms existed.

We used poultry farms maps from Lancaster County, Pennsylvania, and the state of Delaware (referred to as "truth" datasets because they represent the known locations of some poultry farms) to estimate suitability scores for siting factors, including land cover areas and distances to other features, such as roads and support infrastructure. While Lancaster County and the state of Delaware have a high density of poultry farms, the densities of poultry farms vary within the area. For example, Lancaster County and Sussex County (Delaware) have very high farm densities (1.00 and 0.59 poultry farms per square mile, respectively), whereas Kent and New Castle counties (Delaware) have lower farm densities (0.21 and 0.04 poultry farms per square mile, respectively). We used these truth data to represent the location patterns and siting factors for the conterminous United States.

Finally, we obtained locations of poultry farms for the state of Delaware from the DEN (State of Delaware, 2006). The DEN provided actual poultry farm locations, but did not provide data on the type of poultry or count of birds at each farm. We examined the locations of the Delaware farms in the DEN and identified some spatial inaccuracies. We then edited and adjusted these locations using aerial imagery so each farm point was located on a poultry house, as seen in the aerial imagery. Based on the types of farms from the D&B database for Delaware and professional knowledge of the poultry operations

in the state, we concluded that the vast majority of Delaware poultry farms are broiler farms.

We used the DEN data in conjunction with the truth datasets to estimate suitability scores for distances from poultry infrastructure. A suitability score is a value between zero and 1,000 that reflects how suitable a given location is for a poultry farm. A value of zero denotes a location in which the likelihood that a farm would be placed there is nil, whereas 1,000 denotes an area that is very suitable for placement of a farm.

For instance, the likelihood that a farm would be placed inside core developed areas (high- or medium-intensity development) is extremely low, so the suitability score is set to zero; whereas the likelihood that a farm would be placed in areas other than core developed areas is very high, so the suitability score is set to 1,000. Table 2 presents the suitability scores for various siting factors (presented as data layers in Table 2). In many cases, the suitability score varies with distance from features or areas, such as a town or road.

We further categorized suitability scores to represent the different levels of suitability for poultry house placement. For distance from roads, for example, we considered areas within 50 to 300 m of a road to be optimum and assigned them the highest score (1,000) for placing a poultry house, whereas the areas very close to (<50 m) or far from (>1 km) the roads had the lowest scores for placing a poultry house (5 to 10 percent of the optimum).

To determine scores for proximity data layers, we used the actual farm location data (D&B, DEN, Lancaster County) and compared those locations to zones of increasing distance from features (i.e., buffer zones). This enabled us to calculate suitability scores for areas based on their distances from various geographic features. For instance, we generated multiple buffer zones (or rings) at various distances from some roads and poultry support service businesses, including hatcheries. We based the scores assigned to each buffer zone on the distribution of actual poultry farms that fell within them.

Table 2. GIS data layer classes and their associated suitability scores for the likelihood of placing a farm there

GIS Data Layer	Data Source	Data Classes	Suitability Scores
Agricultural/non-agricultural land use/land cover	National Land Cover Dataset (EPA, 2001)	Non-agricultural land use (land use) codes 21–23, 51–52, 72–74	800
		Inside crop land (land use code 82)	1,000
		Inside peripheral pasture/hay (land use code 81)	100
		Inside core[a] pasture/hay (land use code 81)	10
Municipalities and urbanized areas	Esri ArcGIS version 9.2 data disks (Esri, 2006b; US Census Bureau, 2000)	Inner city ("downtown")	10
		Outer city	40
		Near city	950[b]
		Not near city	10
Developed land use/land cover	National Land Cover Dataset (EPA, 2001)	Inside NLCD land use code 24 or core developed area[c]	0
		Outside those areas	1,000
Federal lands	Esri ArcGIS version 9.2 data disks (Esri, 2006b)	Inside	0
		Inside special[d]	250
		Outside	1,000
State and local parks	Esri ArcGIS version 9.2 data disks (Esri, 2006b)	Inside	0
		Outside	1,000
Surface waters	National Hydrography Dataset (medium resolution) (USGS, 2010)	Inside or within 50 m	0
		Outside of 50 m	1,000
Sensitive areas (e.g., churches, schools)	Esri ArcGIS version 9.2 data disks (Esri, 2006b)	Inside and within 50 m	0
		Outside	950
Residential and developed areas	Esri ArcGIS Business Analyst street map (2006 TeleAtlas) (Esri, 2006a)	Inside	0
		Outside	1,000
Non-agriculture businesses	Esri ArcGIS Business Analyst (Esri, 2006a)	Within 50 m	0
		Outside 50 m	1,000
Transportation (roads)	Esri ArcGIS Business Analyst street map (2006 TeleAtlas) (Esri, 2006a)	<50 m	108
		50–300 m	1,000
		300–500 m	230
		500–1,000 m	150
		1,000+ m	50
Transportation (airports)	Esri ArcGIS version 9.2 data disks (Esri, 2006b)	Inside	0
		Outside	1,000
Transportation (railroads)	Esri ArcGIS version 9.2 data disks (Esri, 2006b)	Within 15 m	0
		Outside 15 m	1,000

(continued)

Table 2. GIS data layer classes and their associated suitability scores for the likelihood of placing a farm there *continued*

GIS Data Layer	Data Source	Data Classes	Suitability Scores
Poultry support businesses—all but hatcheries	Esri ArcGIS Business Analyst (Esri, 2006a)	<5 km	1,000
		5–10 km	864
		10–15 km	632
		15–20 km	428
		20–25 km	322
		25–30 km	196
		30+ km	62
Poultry support businesses—hatcheries	Esri ArcGIS Business Analyst (Esri, 2006a)	<10 km	1,000
		10–20 km	347
		20–30 km	172
		30–40 km	92
		40–50 km	71
		50+ km	30
Slope	National Elevation Dataset (USGS, 2006)	<10% slope	1,000
		10–20% slope	286
		20–35% slope	143
		>35% slope	10
Incompatible land cover	2001 National Land Cover Dataset (EPA, 2001) & National Wetlands Inventory (US Fish and Wildlife Service, 2010)	NLCD land use codes 11–12, 31–32, 41–43, 90–99 plus wetlands from NWI	0
		Other land uses	1,000

[a] The core pasture was created by shrinking the pasture/hay NLCD zones by two 30 m cells within ArcGIS. The peripheral pasture/hay areas are composed of all of the non-core pasture/hay cells.

[b] Processing error: The value 950 was used accidentally; the value should have been 1,000. We believe the effect to the end result to be minimal.

[c] Core developed areas consist of high- or medium-intensity developed (codes: 23, 24) land use 30 m cells completely surrounded by other high- or medium-intensity developed 30 m cells.

[d] Our first attempt at the surfaces excluded all public lands and Indian lands until we concluded that not all of them are exclusionary of commercial farms with poultry; some covered entire counties that the US Census of Agriculture stated had poultry farms or for which our national business data had poultry business points that fell well within our Public Lands GIS data layers.

Develop Geographic Layers Using Suitability Scores Associated with the Siting Factors

This data processing step involved developing geographic information system (GIS) data layers that contain geospatially specific farm suitability scores, using the site-location factors and suitability scores specified in Tables 1 and 2. We processed the data layers for each siting factor independently, but we used a consistent map projection and coordinate system so we could easily combine the datasets. We disaggregated all nationwide layers into separate files for individual states, which included an extra 1 km zone around each state to capture features from neighboring states that might influence the final surface in each state.

To create some of the data layers, we needed to reformat or process the source data. For example, we used GIS techniques to compute slope values from the elevation data (National Elevation Dataset [US Geological Survey, 2006]). We also developed a data layer consisting of a 75 m buffer zone around short local road segments (i.e., those less than 0.5 km in length) in a detailed road dataset to approximate exurban developed or residential areas.

We converted each data layer to a uniform raster format—a geospatial layer made up of a series of regularly spaced grid cells all having the same size, such as those seen on typical street-routing Web sites—with grid cells 30 m on a side. We assigned each grid cell the farm suitability score of the category or buffer zone that its center fell within.

Merge Site Suitability Layers into a Single Combined Suitability Surface

We created the raster grid cell layers so that each individual data layer could be merged with other data layers into a single layer (i.e., we combined the individual layer raster grid scores to create an overall suitability score for each 30 m grid cell). The overall suitability score was the product of all the input layer scores calculated separately for each cell and rescaled back to the zero to 1,000 range. Consequently, a cell with a zero suitability score for any of the data layers would have an overall score of zero, enabling us to filter out these cells for farm placement. The resulting set of cell-specific scores is a suitability surface that combines all of the individual data layers. Note that this process uses the same score for all animal types.

Figure 1 is a simplified example of the likelihood combination process that combines three data layers into a single resulting likelihood surface. In future versions of the database, separate likelihood surfaces for each animal type and size class may be necessary depending on the use of the data.

Place Poultry Operations Based on the Combined Suitability Surface

To locate poultry operations for each county, we developed a process consistent with the expectations of the poultry suitability surface. Figure 2 is an example of an input suitability surface, with colors indicating ranges of suitability (left side) and the resulting placement of synthetic farms (right side) for Lancaster County.

To place poultry operations using the constraints of the suitability surface, we first randomly generated 10 times the number of desired poultry locations in the county and then screened out candidate locations based on the suitability score at that location using a random number generator. We assigned random numbers from a uniform distribution between zero and 1,000 to each of the potential synthetic farm locations. We then compared the assigned random number for each potential farm location with the farm suitability score at that location. If the assigned random number was less than the suitability score, we retained the location and considered it still viable to become a synthetic poultry farm location. If the converse was true, we removed the potential farm location from consideration. The intended net effect was that potential locations in more suitable areas were preferentially kept as synthetic farm locations.

In some cases, this process resulted in fewer potential farms being left for consideration in a county than the total number of farms reported in the Census of Agriculture county poultry farm counts. This could happen in counties where the majority of the county had very low farm-siting scores. In these cases, we created a new set of potential farm locations, generating 10 times the previous number of candidates (i.e., 100 times the desired number) and repeated the process for the county, discarding the first attempt.

Figure 1. Example of multiplication of individual suitability layers to generate a combined likelihood surface
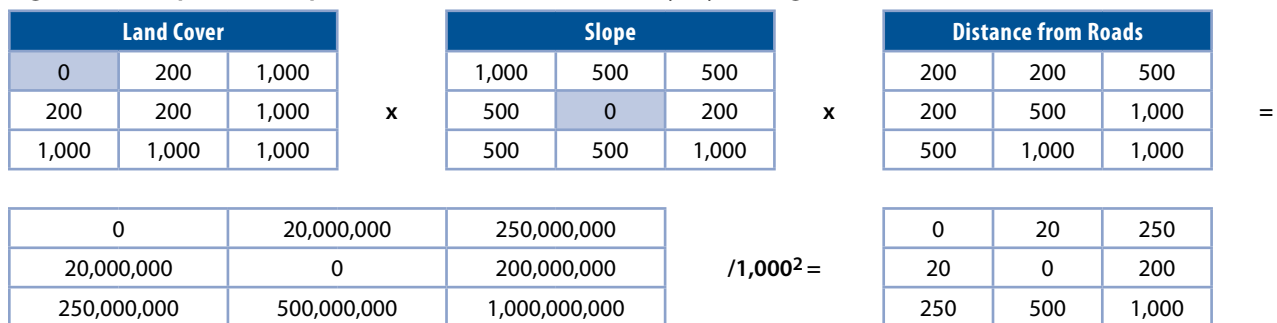
Land Cover				Slope				Distance from Roads			
0	200	1,000		1,000	500	500		200	200	500	
200	200	1,000	x	500	0	200	x	200	500	1,000	=
1,000	1,000	1,000		500	500	1,000		500	1,000	1,000	

0	20,000,000	250,000,000		0	20	250
20,000,000	0	200,000,000	$/1{,}000^2 =$	20	0	200
250,000,000	500,000,000	1,000,000,000		250	500	1,000

Figure 2. Generating synthetic farm locations for Lancaster County, Pennsylvania, using the suitability surface
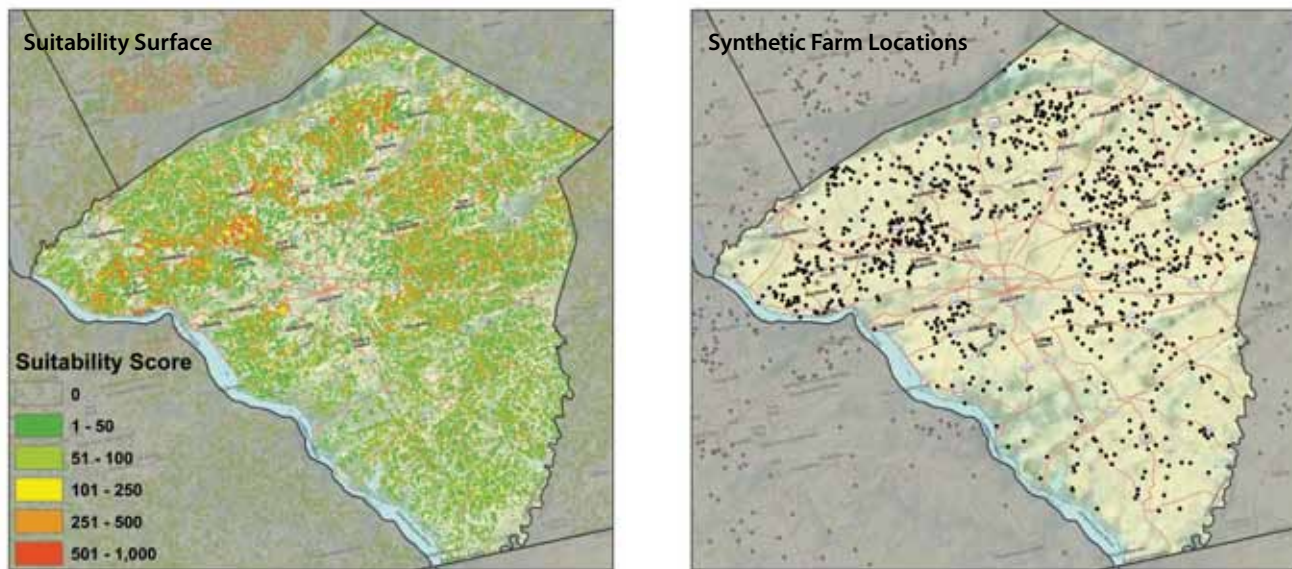
More commonly, too many potential synthetic poultry farm locations passed the screening process. To obtain the desired number of farm locations, we randomly selected locations from the remaining candidates.

Once we reached the desired number of synthetic poultry farm locations in a county, we randomly assigned a bird type and bird count range (i.e., farm size) attribute to each synthetic poultry farm location according to the distribution of poultry farm type and size combinations found in the Census of Agriculture for the county. Finally, we assigned a bird count generated at random from the location's bird count range.

Figure 2 shows the results for the pilot area, Lancaster County. Note the large variability and coverage of the suitability scores across the county and that the high-suitability areas are clustered (left-hand graphic). According to a visual analysis of the pilot results by University of Pennsylvania researchers, the distribution of the synthetic poultry farms (right-hand graphic) reflects the distribution of actual poultry farms much more accurately than farms placed randomly within the county's borders using the same census counts (S. Dunipace, personal communication, 2006). We considered this to be sufficient evidence to proceed with generating synthetic poultry farms for the rest of the conterminous United States.

Accuracy Assessment Method

We used agreement matrices (also known as error, confusion, or correlation matrices), a standard method of evaluating map accuracy (Jensen, 1996; Munoz, Lesser, Dorney, & Savage, 2009; R Development Core Team, 2010; Stehman & Czapewiski, 2003), and statistics derived from these matrices to evaluate the accuracy of our synthesized data. When using agreement matrices to evaluate classification accuracy, one dataset often serves as a reference dataset (i.e., "truth data") and the other serves as a predictive model of the truth.

In this case, we compared locations from the synthesized dataset that could support a farm (i.e., had a non-zero suitability score) and locations that could not support a farm (i.e., had a zero suitability score) with actual land use data from aerial imagery and supplemental data sources.[1] The actual land use data served as the reference dataset that provided the classifications.

We also recalculated the suitability surface following the more common approach, using a uniform distribution to generate a dataset of randomly placed farms (from which we derived random placement classifications) and compared

---

1   Data sources include 1 m resolution 2006 National Agricultural Imagery program (USDA, 2006), USGS topological maps, Street Map from ESRI's ArcGIS Online map service (ESRI, 2010), and 6 in to 1 ft resolution multiyear aerial imagery from North Carolina county sources from the NC OneMap map service (NC OneMap, 2010).

that to both the reference classifications and the synthetic classifications.

We selected North Carolina for this comparison because in North Carolina (1) commercial poultry operations range from very large factory operations to small family-owned farms, (2) the state has a diversity of terrains (e.g., coastal plain, rolling farmlands, foothills, mountains), many of which are similar to other US poultry-producing areas, and (3) we had access to high-resolution (1 m) and very high-resolution (6 in to 1 ft) true-color aerial imagery for most of the state, as well as online US Geological Service (USGS) 1:24,000 digital topographic maps (known as digital raster graphics, or DRGs).

The first step involved the selection of a reasonable number of locations that could be compared across all three surfaces (i.e., synthesized, random, and actual use). The 2002 Census of Agriculture for North Carolina indicated that there were 6,251 poultry farms in the state. Consequently, we randomly selected 6,251 locations from the synthesized surface that had a non-zero suitability score of supporting a poultry farm (i.e., these were classified as "suitable" for the synthesized classification). We repeated the process of selecting and classifying locations using the random placement surface (i.e., these new locations were classified as "suitable" for the random placement classification). We then generated an equal number of "unsuitable" locations from the synthesized surface and from the random placement surface by randomly selecting from among the locations with zero suitability scores. We did this because we are as interested in avoiding incorrect placement as placing farms correctly. From the resulting 25,004 locations (i.e., $4 \times 6,251$), we randomly selected a 10 percent sample of 2,499 for the calculation of the agreement matrices.

The second step involved determining the reference classifications at each of the 2,499 sampled locations from each of the three datasets. To create our reference classifications (actual land use), we developed a simplified set of "suitable" and "not suitable" land use categories (Table 3) that reflects the land uses found in North Carolina, with a particular focus on the land uses associated with commercial poultry production. To some extent, the land use interpretations were limited by what is discernable from high-resolution aerial imagery and topological maps of the area. For example, because small poultry operations are unlikely to have large commercial poultry houses, we may have misidentified them as a non-animal farm or even rural housing. Consequently, small poultry operations may be underrepresented.

Table 3. Land use categories classified as suitable and not suitable for commercial poultry production (sorted by relevance)

Suitable for Commercial Poultry Production
1. Commercial poultry house
2. Commercial poultry operation (poultry houses visible on the property)
3. Livestock farm (e.g., swine, cattle)
4. Other farm (with compatible crops or pastures present; within 3 km of commercial poultry houses)
5. Other farm (with compatible crops or pastures present; no commercial poultry houses visible within 3 km)

Not Suitable for Commercial Poultry Production
6. Rural housing (often larger lots, often have gardens or some small agricultural use associated, not in or near population centers; included some vacation housing as well)
7. Rural area (non-agricultural rural areas not in another category; may include wetlands, barren land, scrub land)
8. Forest/patches of trees (not individual tree)
9. Industrial areas
10. Non-rural housing (organized subdivision or grid structure present)
11. Commercial areas/offices/research parks
12. Water bodies
13. Transportation (most often roads)
14. Other (not described in any other category; also includes areas of farms deemed unsuitable because of steep slope)

To evaluate the overall similarity of the approaches, we calculated three agreement matrices: a matrix comparing the agreement of the synthetic farm locations with the truth (or reference) data; a matrix comparing the randomly placed farm locations with the reference data; and a matrix comparing the synthetic farm locations with the random farm locations. The accuracy statistics calculated from these matrices (Figure 3) include the Total Accuracy (i.e., the proportion of the predictions that were correct), the User's Accuracy (i.e., the proportion of correct predictions given a particular prediction value [Yes/No]), and the Producer's Accuracy (i.e., the proportion of correct predictions given a specific truth value [Yes/No]). Note that the Producer's Accuracy reflects the percentage of poultry locations in the reference dataset (truth data) correctly classified by the synthetic or random surfaces, and that the User's Accuracy reflects the percentage of synthetic or random placement locations that matched the truth data.

Figure 3. Accuracy statistics derived from agreement matrices

		Predicted		Total
		No	Yes	
Truth	No	a	b	a + b
	Yes	c	d	c + d
Total (T)		a + c	b + d	T = a + b + c + d

Accuracy Statistics

Total Accuracy (TA) = (a + d)/T
User's Accuracy (UA):
  Given Prediction = Yes: UA(Yes) = d/(b + d)
  Given Prediction = No: UA(No) = a/(a + c)
Producer's Accuracy (PA):
  Given Truth = Yes: PA(Yes) = d/(c + d) = Sensitivity
  Given Truth = No: PA(No) = a/(a + b) = Specificity
False Positive (FP) = b/(a + b) = Errors of Commission
False Negative (FN) = c/(c + d) = Errors of Omission

We also calculated the proportions of false-positive and false-negative cases, also known as commission and omission errors, respectively. The false-positive statistic is the proportion of locations unsuitable for poultry production that were, in fact, classified as suitable by the synthetic or the random surfaces. The false-negative statistic denotes the proportion of

suitable poultry locations that were, in fact, classified as unsuitable by the synthetic or random surfaces. It is preferable for the false-negative and false-positive statistics to be very low, and for the Producer's, User's, and Total Accuracy statistics to be very high.

We used a bootstrap approach to estimate 95 percent bootstrap confidence intervals (CIs) and mean values for each of the agreement statistics, using the open-source statistical software package R to implement appropriate code (R Development Core Team, 2010). Using sampling with replacement, 1,000 samples of size 500 were drawn at random from the set of 2,499 selected locations. For each of the 1,000 samples, we calculated the agreement statistics (Total Accuracy, User's Accuracy, and Producer's Accuracy, as well as the commission and omission errors), resulting in a sampling distribution of 1,000 values for each statistic. The mean of each statistic is the mean of the resulting distribution, and the upper and lower 95 percent CIs correspond to the 2.5 and 97.5 percentiles of the sampling distribution for each statistic (e.g., Total Accuracy) values.

We used these bootstrap CIs to test the null hypothesis of no difference between accuracy statistics calculated from the two methods. In particular, if the mean of a given accuracy statistic for the synthetic poultry surface did not fall within the 95 percent bootstrap CI of the corresponding accuracy statistic for the randomly placed surface, and vice versa, then we could say that the two given accuracy statistics are significantly different with 95 percent confidence.

Results

Table 4 presents the distribution of classifications for the sampled locations from all three datasets (i.e., the synthetic dataset, the random placement dataset, and the reference dataset, or actual land use) by land use category. For the synthetic surface, a larger percentage of locations with non-zero scores (73.5 percent) were classified in the suitable categories compared with the percentage of the zero score locations (26.5 percent) classified in the same categories. The opposite was observed for locations classified in the suitable categories by the random placement surface. When looking at the not suitable categories, both surfaces

show a larger percentage of the zero score locations classified in these categories. Overall, the estimates presented in Table 4 indicate that the synthetic surface was more successful in replicating the actual land use distribution.

Table 5 displays the accuracy statistics for the synthetic and random placement surfaces. A 95 percent bootstrap CI for the Total Accuracy is 75

to 82 percent, which suggests good prediction capabilities of this surface. The low percentage of false-negative and false-positive cases suggests a low rate of misclassification. The high rates for User's Accuracy No (95% CI: 84.3, 91.3) and Producer's Accuracy Yes (95% CI: 66.4, 80.5) reflect the rate for which the synthetic surface correctly predicts that a site is not a poultry farm.

Table 4. Classification results of sampled locations in North Carolina, comparing land use categories by suitability scores for synthetic and random placement surfaces

Land Use Categories		Synthetic Poultry Surface		Random Placement Surface	
Description	Reference n	Non-zero score areas n (%a)	Zero score areas n (%a)	Non-zero score areas n (%a)	Zero score areas n (%a)
Suitable for Commercial Poultry Production[b]					
1. Poultry house	7	6 (85.7)	1 (14.3)	1 (14.3)	6 (85.7)
2. Poultry farm/ operation	158	118 (74.7)	40 (25.3)	37 (23.4)	121 (76.6)
3. Livestock farm	94	58 (61.7)	36 (38.3)	29 (30.9)	65 (69.1)
4. Other farm (near poultry)	321	250 (77.9)	71 (22.1)	58 (18.1)	263 (81.9)
5. Other farm	163	114 (69.9)	49 (30.1)	36 (22.1)	127 (77.9)
Suitable Subtotals	**743**	**546 (73.5)**	**197 (26.5)**	**161 (21.7)**	**582 (78.3)**
Not Suitable for Commercial Poultry Production[c]					
6. Rural housing	106	33 (31.1)	73 (68.9)	20 (18.9)	86 (81.1)
7. Rural area	102	47 (46.1)	55 (53.9)	18 (17.6)	84 (82.4)
8. Forest/trees	1,239	171 (13.8)	1,068 (86.2)	349 (28.2)	890 (71.8)
9. Industrial areas	29	14 (48.3)	15 (51.7)	6 (20.7)	23 (79.3)
10. Non-rural housing	87	11 (12.6)	76 (87.4)	29 (33.3)	58 (66.7)
11. Commercial/office	23	5 (21.7)	18 (78.3)	5 (21.7)	18 (78.3)
12. Water	44	3 (6.8)	41 (93.2)	12 (27.3)	32 (72.7)
13. Transportation	64	21 (32.8)	43 (67.2)	22 (34.4)	42 (65.6)
14. Other	62	26 (41.9)	36 (58.1)	13 (21.0)	49 (79.0)
Non-Suitable Subtotals	**1,756**	**331 (18.8)**	**1,425 (81.2)**	**474 (27.0)**	**1,282 (73.0)**
Totals	**2,499**	**877**	**1,622**	**635**	**1,864**

[a] Percentage of land use category; e.g., 85.7% of the 7 poultry house locations were classified as suitable in the synthesized dataset, but only 14.3% were classified as suitable in the random dataset.

[b] A successful predictive model for these locations would classify these locations as non-zero suitability areas; that is, ideally the percentage of locations that were non-zero on the surface should be close to 100%, and the percentage that were zero should be close to 0% for each category.

[c] A successful predictive model for these locations would classify these locations as zero suitability areas; that is, ideally the percentage of locations that were zero on the surface should be close to 100%, and the percentage that were non-zero should be close to 0% for each category.

Table 5. Accuracy statistics: mean and bootstrap 95 percent confidence interval (CI) by surface

Statistic	Synthetic Mean (95%CI)	Random Placement Mean (95%CI)
Total Accuracy Rate	78.8% (74.8, 82.2)	57.8% (53.6, 62.0)
User's Accuracy (Yes)	62.0% (54.5, 69.5)	25.2% (17.9, 33.3)
User's Accuracy (No)	87.9% (84.3, 91.3)	68.9% (64.1, 73.3)
Producer's Accuracy (Yes)	73.4% (66.4, 80.5)	21.6% (15.4, 28.6)
Producer's Accuracy (No)	81.0% (76.7, 85.0)	73.0% (68.7, 77.6)
False Positives	19.0% (15.0, 23.3)	27.0% (22.4, 31.3)
False Negatives	26.6% (19.5, 33.6)	78.4% (71.4, 84.6)

With respect to the random placement surface, a bootstrap CI for the Total Accuracy is from 54 to 62 percent. The high rate of false negative values (71 to 84 percent) indicates a high rate of misclassification by errors of omission, and a moderate to high tendency to classify unsuitable locations as suitable. The lower rate of false positive values (22 to 31 percent) indicates that there is a low tendency to classify unsuitable locations as suitable for poultry farms. The low rates of User's Accuracy Yes (18 to 33 percent) and Producer's Accuracy Yes (15 to 29 percent) suggest a poor ability to find suitable poultry locations where they exist, whereas higher rates for User's Accuracy No (64 to 73 percent) and Producer's Accuracy No (69 to 78 percent) suggest better predictive power for areas not suitable for poultry production.

All of the accuracy statistics for the two surfaces are significantly different (at the 95 percent CI) from one another.

Table 6 displays the accuracy statistics for the synthetic surface compared with the random placement surface. The Total Accuracy ranges from 48 to 57 percent, which suggests weak agreement between the two surfaces. The low rates of User's Accuracy Yes (18 to 34 percent) and Producer's Accuracy Yes (13 to 25 percent) suggest very poor agreement between the two methods on where suitable poultry sites are located. The higher rates for User's Accuracy No (56 to 67 percent) and Producer's Accuracy No (66 to 76 percent) suggest a better agreement on where suitable poultry locations are not located.

Table 6. Accuracy statistics: mean and bootstrap 95 percent confidence interval (CI) for synthetic vs. random placement surfaces

Statistic	Mean (95% CI)
Total Accuracy Rate	52.5% (48.2, 56.8)
User's Accuracy (Yes)	25.6% (18.0, 33.6)
User's Accuracy (No)	61.6% (56.4, 66.7)
Producer's Accuracy (Yes)	18.5% (13.1, 24.5)
Producer's Accuracy (No)	70.9% (66.2, 75.6)
False Positives	29.1% (24.4, 33.8)
False Negatives	81.5% (75.5, 86.9)

The lack of strong agreement between the two surfaces and the superior poultry farm prediction capability of the synthetic surface suggests that this method can be used to create more reliable inputs to agent-based models than those from random placement methods.

Discussion

By using US Census of Agriculture farm counts and publicly available GIS data, this study demonstrates that it is possible to generate a nationwide poultry farm dataset with locations that are significantly more realistic than locations created using a common random placement method alternative. Our method, and the maps it provides, has already proved useful in obtaining preliminary estimates of the parameters that govern the transmission of highly pathogenic avian influenza in the USA (Rorres, Pelletier, Bruhn, & Smith, 2011).

The success rate of the synthetic poultry surface in correctly identifying commercial poultry production locations (62.3 percent) is more than twice that of the random placement surface (25.4 percent). The accuracy statistics (i.e., Total Accuracy, User's Accuracy, and Producer's Accuracy) for the synthetic poultry surface are all significantly higher (95 percent confidence level) than the corresponding random placement surface's accuracy statistics. Most importantly, the Total Accuracy statistic is quite high for the synthetic poultry surface, with a mean of 79 percent, as compared with a mean of 58 percent for the random placement surface. Also, the rates of misclassifications (i.e., false negative and false positive cases) are lower (i.e., better) for the synthetic surface.

These synthetic data can be used in modeling and simulations—for example, infectious disease spread modeling of avian influenza—and when regional estimates of actual farm distributions are needed, for example, these data can be used in the determination of the approximate number of poultry farms and birds by watershed. However, we do not recommend using the data when precise locations and/or reliable associated poultry counts for specific farms are needed, such as might be the case in site pollution estimation scenarios.

Overall, the results suggest the viability of adapting this method to simulate other livestock farms of interest to infectious disease researchers, such as swine and cattle. To improve on the realism of the locations and estimated bird counts, we are investigating five research avenues.

First, we are considering incorporating actual farm locations when available after the locations have been confirmed using aerial imagery and ancillary datasets, such as the datasets used to classify the locations in our accuracy assessment plus any local data sources if they are available. The associated farm capacity (number of animals) might be estimated if the imagery supports the estimation of animal house sizes. Farm type is often included with the actual farm location data, or might be estimated based on the animal building size, number, configuration, and contextual information. We will use these actual farm locations to perform additional accuracy assessments as well.

Second, to reduce prediction errors related to precision or classification errors present in the data sources used to create the synthetic poultry surface, we are considering substituting or adding newer datasets, such as LandScan 90 m population counts (Oak Ridge national Laboratory, 2010), better approximations for rural housing and industrial areas, more accurate forest data, soils, and detailed croplands.

Third, if further investigation determines that the different combinations have significantly different placement profiles, we can create separate placement surfaces for the different animal farm type and size combinations. For example, large broiler farms (>10,000 birds) would be expected to have different spatial constraints than small layer farms (<100 birds) because of their different infrastructure, food, and waste-handling requirements.

Fourth, we plan to examine the suitability factors more closely to determine whether a similar result might be obtained with fewer data inputs, and in that case develop a simpler process with fewer data inputs and less processing to generate more current versions of the synthetic population. An examination using fewer data points might provide valuable insights that could be used if similar data were to be generated for other countries that may not be as data rich or may have different local suitability factors.

Lastly, we looked at the county-summarized estimated bird counts and noted that the average difference from the Census of Agriculture county counts was -4 percent for layers, 24 percent for broilers, 48 percent for pullets, and 73 percent for turkeys. The largest average percentage difference was noted for counties with the smallest number of birds (e.g., turkey farms with 1,000 or more turkeys had an average difference of only 13 percent). The differences are due to the need to assume a sampling distribution within farm size categories to assign synthetic farm sizes. We assumed a uniform distribution. In future research, we will explore other distributions and enforce tolerances on the acceptable difference between the estimated bird counts and the reported bird counts at the county level.

References

Baca, R. (2004). National Animal Identification System. *NAHSS Outlook, November/December*. Retrieved from Animal and Plant Health Inspection Service website: http://nsu.aphis.usda.gov/outlook/issue3/national_animal_identification_system.pdf.

Bates, T. W., Carpenter, T. E., Thurmond, M. C. (2003). Benefit-cost analysis of vaccination and preemptive slaughter as a means of eradicating foot-and-mouth disease. *American Journal of Veterinary Research, 64(*7), 805–812.

Bates, T. W., Thurmond, M. C., & Carpenter, T. E. (2003a). Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. A*merican Journal of Veterinary Research, 64(*2), 195–204.

Bates, T. W., Thurmond, M. C., Carpenter, T. E. (2003b). Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease. A*merican Journal of Veterinary Research, 64(*2), 205–210.

Berry, J. (2007, May). *Factors involved in site selection for new and modified poultry facilities*. Oklahoma Cooperative Extension Fact Sheet F-8213. Retrieved from: http://pods.dasnr.okstate.edu/docushare/dsweb/Get/Document-2100/ANSI-8213web.pdf.

Dun & Bradstreet, Inc. (2007). Dun & Bradstreet business database. Retrieved from http://www.dnb.com/us/.

Engle, J. (2006, October). *Poultry barn extraction from DigitalGlobe and PAMAP imagery using segmentation-based object-oriented classification*. Presentation at the meeting of the Pennsylvania Geological Survey.

Esri. (2010). Esri ArcGIS Online map services [Internet homepage]. Retrieved from http://www.arcgis.com

Esri (2006a). Esri Business Analyst Desktop [Internet homepage]. Retrieved from http://www.esri.com/software/arcgis/extensions/businessanalyst/index.html

Esri (2006b). Data and maps for ArcGIS. Supplied with ArcGIS software update to version 9.2. Retrieved from http://www.esri.com/data/data-maps/index.html

Geter, K. (2006). Farm animal demographics simulator aids in disease modeling. *NAHSS Outlook*, Quarter Two. Retrieved from http://nsu.aphis.usda.gov/outlook/issue10/outlook_apr06_fads.pdf.

Goan, C. (2005). *Site selection factors for new poultry facilities*. The University of Tennessee Agricultural Extension Service, SP 592. Retrieved from https://utextension.tennessee.edu/publications/Documents/SP592.pdf.

Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective*. 2nd ed. (pp. 247–251). Upper Saddle River, New Jersey: Prentice Hall.

Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G., & Grenfell, B. T. (2003). Modeling vaccination strategies against foot-and-mouth disease. *Nature, 421*, 136–142.

Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., & Grenfell, B. T. (2001). Dynamics of the 2001 UK foot-and-mouth epidemic: Stochastic dispersal in heterogeneous landscape. *Science, 294*, 813–817.

Kobayashi, M., Carpenter, T. E., Dickey, B. F., & Howitt, R. E. (2007a). A dynamic, optimal disease control model for foot-and-mouth disease: I. Model description. *Preventive Veterinary Medicine, 79*(2/4), 257–273.

Kobayashi, M., Carpenter, T. E., Dickey, B. F., & Howitt, R. E. (2007b). A dynamic, optimal disease control model for foot-and-mouth-disease: II. Model results and policy implications. *Preventive Veterinary Medicine, 79*(2/4), 274–286.

Le Menach, A., Legrand, J., Grais, R. F., Viboud, C., Valleron, A-J, & Flahault, A. (2005). Modeling spatial and temporal transmission of foot-and-mouth disease in France: Identification of high risk areas. *Veterinary Research, 36*, 699–712.

Li, F. C.K., Choi, B. C. K., Sly, T., & Pak , A. W. P. (2008). Finding the real case-fatality rate of H5N1 avian influenza. *Journal of Epidemiology and Community Health, 62*, 555–559.

Miller, R., Geter, K., Corso, B., Forde-Folle, K., Neufeld, D., Freier, J., & Russell, J. (2007, August). *Farm location and animal population simulator: A system for estimating farm and animal populations.* Poster presented at the GISVet 2007 Conference, Copenhagen, Denmark.

Morris, R. S., Wilesmith, J. W., Stern, M. W., Sanson, R. L., & Stevenson, M. A. (2001). Predictive spatial modeling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain. *Veterinary Record, 149*, 137–144.

Munoz, B., Lesser, V. M., Dorney. J. R. & Savage R. (2009). A proposed methodology to determine accuracy of location and extent of geographically isolated wetlands. *Environmental Monitoring and Assessment, 150*(1–4), 53–64. doi: 10.1007/s10661-008-0672-0

NC OneMap. (2010). 2010 Orthoimagery Project. [Internet page of map service using ArcGIS software.] Retrieved from http://www.nconemap.com/NCOrthos/tabid/425/Default.aspx

Oak Ridge National Laboratory, Geographic Information Science and Technology. (2010). LandScan [dataset]. Retrieved from http://www.ornl.gov/sci/landscan/.

R Development Core Team. (2010). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org

Rorres, C., Pelletier, S. T. K., Keeling, M. J., & Smith, G. (2010). Estimating the kernel parameters of premises-based stochastic models of farmed animal infectious disease epidemics using limited, incomplete, or ongoing data. *Theoretical Population Biology, 78*, 46–53.

Rorres, C., Pelletier, S. T. K., Smith, G. (2011). Stochastic modeling of animal epidemics using data collected over three different spatial scales. *Epidemics, 3*, 61–70.

Rorres, C., Pelletier S. T. K., Bruhn, M. C., & Smith G. (2011). Ongoing estimation of the epidemic parameters of a stochastic, spatial, discrete-time model for a 1983–84 avian influenza epidemic. *Avian Diseases, 55*, 35-42.

Sambidi, P. R. (2003, May). *Factors affecting plant location decisions of US broiler executives (Master's thesis). Louisiana State University*. Retrieved from http://etd.lsu.edu/docs/available/etd-0409103-091155/unrestricted/Sambidi_thesis.pdf.

Sambidi, P. R., Harrison, R. W., & Farr, A. J. (2004). *A conjoint analysis of site selection for the US broiler industry: Implications for Louisiana*. Louisiana State University Agricultural Center Bulletin Number 882. August. Retrieved from http://www.lsuagcenter.com/NR/rdonlyres/B478CEEC-3822-46DF-B044-4D943862AE53/4346/B884.pdf.

Smith, G., & Kelly, A. M. (2008). *Food security in a global economy*. Philadelphia, PA: University of Pennsylvania Press.

State of Delaware, Department of Natural Resources and Environmental Control. (2006). Delaware Environmental Navigator. Retrieved from http://www.nav.dnrec.delaware.gov/DEN3/Search/FacilitySearch.aspx

Stehman, S. V., & Czapewiski, R. L. (2003). Introduction to special issue on map accuracy. *Environmental and Ecological Statistics, 10*(3), 301–308.

Tildesley, M. J., House, T. A., Bruhn, M. C., Curry, R. J., O'Neil, M., Allpress, J. L. E., Smith, G., & Keeling, M. J. (2010). The impact of spatial clustering on disease transmission and optimal control. *Proceedings of the National Academy of Sciences, USA, 107*, 1041–1046.

US Census Bureau. (2000). Urbanized areas [Maps and data files]. Retrieved from http://www.census.gov/geo/www/ua/uaucbndy.html

US Department of Agriculture, National Agricultural Imagery Program (NAIP). (2006). [Internet homepage.] Retrieved from http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=nai

US Department of Agriculture, US Census of Agriculture. (2002). *2002 Census publications*. (All counties by state by table, Table 13: Poultry—inventory and sales: 2002 and 1997). Retrieved from http://www.agcensus.usda.gov/Publications/2002/index.asp

US Department of Agriculture, Animal and Plant Health Inspection Service. (2010). *Questions and answers: New animal disease traceability framework*. APHIS Veterinary Services Fact Sheet. Retrieved from http://www.aphis.usda.gov/publications/animal_health/content/printable_version/faq_traceability.pdf.

US Environmental Protection Agency, Multi-Resolution Land Characteristics Consortium. (2001). 2001 National Land Cover Data (NLCD 2001). Retrieved from http://www.epa.gov/mrlc/nlcd-2001.html

US Fish & Wildlife Service. (2010). National Wetlands Inventory. Retrieved from http://www.fws.gov/wetlands/.

US Geological Survey. National Elevation Dataset (NED). (2006). Retrieved from http://ned.usgs.gov/.

US Geological Survey. (2010). National Hydrography Dataset. Retrieved from http://nhd.usgs.gov/.

World Health Organization. (2005, November 4). Avian influenza frequently asked questions. *Weekly Epidemiological Record, 80*(44), 377–388. Available from http://www.who.int/wer/2005/wer8044.pdf.

Appendix

The US Census of Agriculture reports poultry operations by type in the following categories: broilers (chickens raised for meat); layers (chickens raised to lay eggs); pullets (young hens to replace layers); turkeys; and other poultry, including ducks, geese, and other miscellaneous species. Assigning the type of poultry operation to each synthetic farm is not straightforward because the Census of Agriculture counts the same farm in multiple categories if multiple types of poultry production occur on that farm. This means that summing the number of farms across all categories will not equal the total number of farms in the county.

To assign farm types, we ranked the 2002 Census of Agriculture poultry farm type/size categories in order of likelihood of being a part of a farm that contains multiple types of poultry production. We felt that farms that raised pullets, turkeys, ducks, geese, and others would be less likely to mix with broiler or layer production because of their different infrastructure needs. So, these farms were assigned to synthetic farms first. Also, based on the experience of the University of Pennsylvania researchers, the largest of the broiler and layer farms were not likely to have another type of poultry production occur on the same farm, so we assigned a higher assignment priority to these poultry farms next. The smallest layer and broiler farms are the most likely to occur with other types of poultry production on the same farm, so they were assigned last. If in the process of assigning farms types, the total number of poultry farms is reached, then no more farm types for that county will be assigned. In practice, this means that some counties did not have any farms that were assigned the lowest priority farm type (i.e., farms raising 1–49 layer chickens). The farm categories that had not yet been assigned to a synthetic farm location were assumed to be part of a farm already assigned to another poultry production type.

The order of assignment was as follows, from least likely to be part of a larger farm with a different poultry production type:

- Pullets
- Turkeys
- Ducks, Geese, and Other
- Broilers6 (500,000+ broiler chickens sold)
- Broilers5 (200,000–499,999 broiler chickens sold)
- Broilers4 (100,000–199,999 broiler chickens sold)
- Broilers3 (60,000–99,999 broiler chickens sold)
- Layers9 (100,000+ layer chickens)
- Layers8 (50,000–99,999 layer chickens)
- Layers7 (20,000–49,999 layer chickens)
- Layers6 (10,000–19,999 layer chickens)
- Broilers2 (2,000–59,999 broiler chickens sold)
- Layers5 (3,200–9,999 layer chickens)
- Layers4 (400–3,199 layer chickens)
- Broilers1 (1–1,999 broiler chickens sold)
- Layers3 (100–399 layer chickens)
- Layers2 (50–99 layer chickens)
- Layers1 (1–49 layer chickens)

Acknowledgments