Modelling the domestic poultry population in the United States: A novel approach leveraging remote sensing and synthetic data methods

Kelly A. Patyk,[1] Mary J. McCool-Eye,[1] David D. South,[1] Christopher L. Burdett,[2] Susan A. Maroney,[2] Andrew Fox,[1] Grace Kuiper,[2] Sheryl Magzamen[2]

[1]*United States Department of Agriculture, Animal Plant and Health Inspection Service, Veterinary Services, Strategy and Policy, Center for Epidemiology and Animal Health;* [2]*Colorado State University, Department of Environmental and Radiological Health Sciences, Fort Collins, CO, USA*

Abstract

Comprehensive and spatially accurate poultry population demographic data do not currently exist in the United States; however, these data are critically needed to adequately prepare for, and efficiently respond to and manage disease outbreaks. In response to absence of these data, this study developed a national-level poultry population dataset by using a novel combination of remote sensing and probabilistic modelling methodologies. The Farm Location and Agricultural Production Simulator (FLAPS) (Burdett *et al.*, 2015) was used to provide baseline national-scale data depicting the simulated locations and populations of individual poultry operations. Remote sensing methods (identification using aerial imagery) were used to identify actual locations of buildings having the characteristic size and shape of commercial poultry barns. This approach was applied to 594 U.S. counties with > 100,000 birds in 34 states based on the 2012 U.S. Department of Agriculture (USDA), National Agricultural Statistics Service (NASS), Census of Agriculture (CoA). The two methods were integrated in a hybrid approach to develop an automated machine learning process to locate commercial poultry operations and predict the number and type of poultry for each operation across the coterminous United States. Validation illustrated that the hybrid model had higher locational accuracy and more realistic distribution and density patterns when compared to purely simulated data. The resulting national poultry population dataset has significant potential for application in animal disease spread modelling, surveillance, emergency planning and response, economics, and other fields, providing a versatile asset for further agricultural research.

Introduction

The United States (U.S.) poultry industry is the world's largest producer and the second largest exporter of poultry meat (ERS, 2019). Poultry production is an important sector of the agricultural economy in the United States with 8.54 billion broilers, 99.8 billion eggs, and 238 million turkeys produced in 2014, for a combined value of USD $48.3 billion (NASS, 2015). Safeguarding the industry from infectious disease, especially high consequence pathogens, such as Highly Pathogenic Avian Influenza (HPAI), is a top priority due to the potentially devastating effects of outbreaks to producers, the poultry industry, and the U.S. economy. For example, the 2014-2015 H5N8 and H5N2 HPAI outbreak in the United States resulted in the death (from infection) or depopulation of approximately 48 million domestic poultry, amounted to USD $879 million in outbreak response costs, and incurred economic impacts to producers of USD $1.043 billion (Johnson *et al.*, 2016; Seitzinger and Paarlberg, 2016).

Complete and accurate data are needed to adequately prepare for, and efficiently respond to, manage, and eradicate poultry disease outbreaks. In the event of an outbreak, knowledge of the location, distribution, and density of operations is critical to quickly develop response and surveillance plans, estimate resource needs, and predict the potential number of impacted operations. Epidemiologic models are often used to support planning activities before an outbreak, aid in decision making during an outbreak, and investigate outbreaks retrospectively (Dent *et al.*, 2011; Patyk *et al.*, 2013; Backer *et al.*, 2015; Andronico *et al.*, 2019). The accuracy of model outputs relies on the quality of the underlying data. One key set of model inputs is population data, including the point or area locations of operations in the at-risk population. A national, comprehensive dataset of individual poultry oper-

ation locations is not currently available in the United States. While some actual data (*i.e.*, true latitude and longitude coordinates of poultry operations) may exist for specific geographic areas or poultry sectors, those data are typically not available or accessible due to the conditions under which they were collected (*i.e.*, emergency response) (Martin *et al.*, 2015). The United States Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) conducts a Census of Agriculture (CoA) every five years for all farms and ranches in the United States where USD $1,000 or more of agricultural products were produced and sold during the census year (NASS, 2019). While these data are the most comprehensive available in the United States, the data are collected under specific terms of confidentiality; therefore, NASS only publishes results in aggregate at the county or state level (NASS, 2019). In the absence of a national dataset that provides point locations of operations and operation attributes (*e.g.*, production type, numbers of birds), several efforts have been made to develop synthetic population datasets. These efforts have used smart-placement techniques (*e.g.*, dasymetric mapping, masking, statistical modelling) to generate operation locations from the county- or state-level summaries available from NASS (Melius *et al.*, 2006; Melius, 2008; Bruhn *et al.*, 2012; Burdett *et al.*, 2015). Synthetic datasets often meet the basic needs for preparedness and initial response; however, it is important to assess their fit for purpose given the potential impacts that population characteristics – such as location, number of animals, and the density and clustering of operations – have on the course of an outbreak and on model results (Boender *et al.*, 2007; Tildesley *et al.*, 2010; Werkman *et al.*, 2016; Bonney, *et al.*, 2018; Meadows, *et al.*, 2018; van Andel *et al.*, 2018). While synthetic datasets simulate operation locations by placing them in geographic areas considered suitable, techniques using remotely sensed data and artificial-intelligence methodology (*i.e.*, machine-learning algorithms) can identify the actual locations of buildings having the characteristic size and shape of a commercial poultry barn (Maroney *et al.*, 2020). Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance from the targeted area (USGS, 2019). Different surface types such as water, bare ground, vegetation, and buildings reflect radiation differently in various channels of the electromagnetic spectrum, referred to as the spectral signature of the features' surface (ESA, 2009).

Remote sensing classification methods use specialized software to identify (via spectral signatures) and map various features, such as agricultural facilities. Analyst input guides the process of correctly identifying and mapping the operations. However, because manually identifying poultry facilities using imagery can be labor intensive and time consuming over larger geographic areas, other methods are often needed to map commercial poultry operations at a multi-state or national scale. Newer remote sensing techniques use both spectral signatures in remotely sensed data and object identification methods, such as size and shape of buildings, to better classify and map human-made structures (*e.g.*, Freire *et al.*, 2014; Spröhnle *et al.*, 2014; Caggiano *et al.*, 2016).

The objective of this study was to use novel approaches to develop a national-level poultry operations and population dataset for the United States with improved locational accuracy suitable for preparedness and response activities in the event of an avian disease outbreak. Our goals were to combine simulated and remote sensing methods in a unique way that leverages the strengths of both methodologies: realistic poultry production types and inven-

tories and accurate location placements of operations. This novel approach utilized available baseline national-scale data generated by the Farm Location and Agricultural Production Simulator (FLAPS) and high-resolution imagery from the USDA National Aerial Imagery Program (NAIP). The approach and methodology are repeatable with moderate time and effort to allow for timely updates as new data become available.

Materials and methods

Overview

To develop a national-level poultry operations and population dataset, we first used a combination of remote sensing and probabilistic modelling methods to map the locations (latitude and longitude) and define the production type and number of poultry on individual commercial poultry operations throughout the coterminous United States (minimum bounding rectangle coordinates (WGS 1984): [-124.586288, 50.925292], [-66.808444, 47.875751], [-68.053058, 24.294786], [-125.830902, 27.344327]. In poultry-intensive counties (> 100,000 birds), we used a remote-sensing methodology and machine-learning algorithms to detect actual commercial poultry operations from aerial imagery. This output was integrated with probabilistic models and synthetic data from the FLAPS (Burdett *et al.,* 2015), which allowed us to identify erroneous operations and apply simulated demographic attributes to the actual operation locations. This is referred to as the "hybrid model". Next, in non-poultry-intensive counties (≤ 100,000 birds), we used synthetic location and demographic data from the FLAPS model. We also used FLAPS output to map backyard poultry operations (< 1,000 birds) in all U.S. counties. The final national-level domestic poultry operations dataset was developed by combining the hybrid model for poultry-intensive counties with FLAPS output for all non-poultry-intensive U.S. counties. Figure 1 provides an overview of the methods used to develop this national-level poultry operations and population dataset, described in detail below.

FLAPS overview

FLAPS is a stochastic, spatially-explicit, microsimulation model that simulates the distribution and populations of individual poultry and livestock operations across the coterminous United States using USDA-NASS CoA data (Burdett *et al.*, 2015). CoA data represent the most comprehensive agricultural production data for the United States; however, to maintain confidentiality, CoA data are aggregated at the county- or state-level, and data are redacted from counties or states with few operations. FLAPS contains algorithms to simulate the redacted CoA and then disaggregates the county- or state-level data to produce synthetic datasets depicting the locations and populations of individual poultry and livestock operations. Initially developed for swine, the development of FLAPS modules for poultry is discussed in Supplement 1.

Hybrid model for poultry-intensive counties

Based on subject matter expertise, knowledge of the poultry industry, and historical avian influenza outbreak data, poultry-intensive counties were defined as those with > 100,000 birds (the total number of commercial layers, broilers, pullets, and turkeys) based on 2012 NASS CoA data (Figure 2).

Remote sensing to identify commercial poultry barns

The first step of the hybrid model was the application of remote sensing methods to NAIP imagery (collected 2014-2017) (USDA, 2016). Collected by USDA Farm Services Agency, the 3-band, true color NAIP aerial imagery has 1-m resolution sufficient to detect commercial poultry barns. A NAIP image was downloaded for each of the poultry-intensive counties which summed to a total of 484 GB of data. The ArcGIS Feature Analyst (FA) extension (Textron Systems) was used to classify and map commercial poultry operations within the NAIP imagery. We used a FA protocol developed in an earlier pilot study (Maroney *et al.*, 2020) where commercial poultry operations were identified from aerial imagery. The study area for the pilot covered thirty-five counties with varying poultry densities and production types across differ-
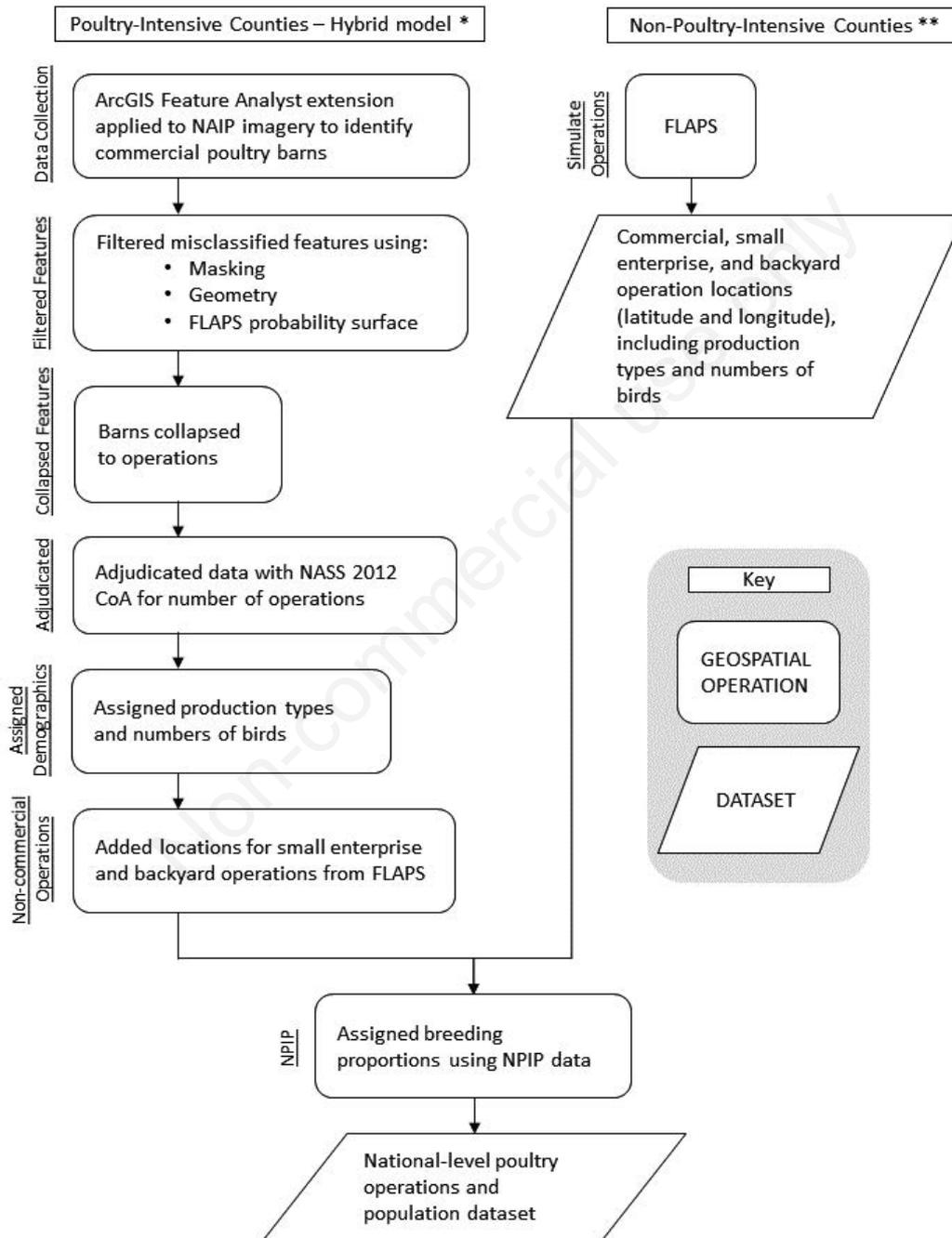
*Poultry-intensive counties are defined as having > 100,000 birds; **Non-poultry-intensive counties are defined as having ≤ 100,000 birds; NAIP = National Aerial Imagery Program; NASS = National Agricultural Statistics Service; FLAPS = Farm Location and Agricultural Production Simulator; CoA = Census of Agriculture; NPIP = National Poultry Improvement Plan.

Figure 1. Overview of methods to develop the U.S. national-level poultry operations and population dataset.

ing ecoregions in seven states in the southeastern United States. In the pilot study, FA used a supervised learning algorithm to identify objects in imagery, driven by user-defined samples and inputs (Blundell, 2008). This supervised learning routine was an iterative process that leveraged spectral inputs from imagery, training samples identified by the analyst, additional object parameters (length and width of buildings), and predefined software input patterns (Figure 3). An analyst performed iterative-processing runs in FA to improve results.

In the current study, we developed state-specific models in FA to expedite the remote sensing process for the poultry-intensive counties. Specifically, for each state having at least one poultry-

intensive county, a single county was selected to create a training set of commercial poultry barn features. Results from each state-specific training county were then applied in FA to the remaining poultry-intensive counties within each state. These state datasets included poultry barn features, which were combined for further processing. This dataset included many accurate poultry barn features, but also misclassified features, such as non-barn buildings, fallow fields, patches of bare ground, and river sand bars. These misclassified features, referred to as false positives, were filtered or removed from the dataset using the steps described below. Analogous processes were previously performed manually by an analyst in the pilot study (Maroney *et al.*, 2020). To improve the
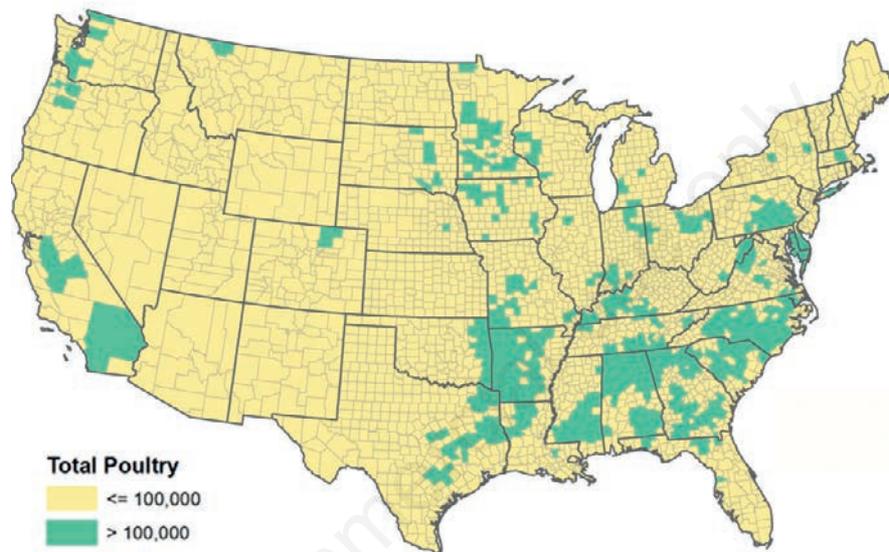
Figure 2. Map of counties in the coterminous United States with ≤ 100,000 (yellow) or > 100,000 (green) total commercial birds as reported in the NASS 2012 Census of Agriculture. Total poultry includes layers, broilers, pullets and turkeys.

Figure 3. Example of NAIP imagery of a commercial poultry operation in the United States.

remote sensing output, we developed a Python script to automate these processes (South, 2020a; 2020b).

Filtering remote sensing results

First, a masking process was implemented to remove poultry barn features from areas where they would not likely be found. Using 2012 and 2014 Environmental Systems Research Institute (ESRI) Base Data, masking layers were developed for urban areas (United States Defense Mapping Agency, 2014), streets (with a 20-m buffer) (TomTom Inc., 2012), lakes and reservoirs (USGS, 2014), and rivers (with a 10-m buffer) (USGS, 2014). Features within these masked areas were removed from the dataset. Second, length thresholds were used to remove misidentified commercial poultry barns from the datasets where the barn "footprint" (polygon) lengths were greater than 500 m or less than 50 m, substantially outside the standard length of a commercial poultry barn (Bell and Weaver, 2002; Fairchild, 2005).

We then used the probability surface from FLAPS that estimated the probability of occurrence of poultry operations to further eliminate false positives (Burdett *et al.*, 2015). In brief, the probability surface is a spatially-explicit raster dataset that maps a logistic regression model. The inputs to the logistic regression model included a binary response variable of places where poultry operations were both present and absent, and predictor variables that included distances to classified land-use (*e.g.*, forest, urban), distance to roads, and topography that were associated with the distribution of poultry operations. After the logistic regression results were mapped in a GIS, the output raster layer was standardized to a range of 0 to 1, which represented a gradient of increasing likelihood of poultry operation presence. A more detailed description of the probability threshold surface and its development are available in Supplement 1.

We assigned the value from the probability surface to a dataset of poultry operations from the 35 counties that were mapped with a manual methodology in the pilot study (Maroney *et al.*, 2020).

The value of this pilot dataset is that it identified the actual locations of poultry operations retained after manual review (*i.e.*, true positives), and the locations that the analyst removed during manual review (*i.e.*, false positives), and those that the analyst added during manual review (*i.e.*, false negatives, or barns that were not identified with the machine-learning algorithm). This allowed us to identify whether the probability surface values for actual poultry-operation locations differed from the set of false-positive locations that the machine learning algorithm identified as possible operation locations but were removed during the manual review because they were not barns. Each point (latitude and longitude) in the pilot dataset was assigned a number between 0 and 1 from the probability threshold surface described in Supplement 1. After plotting the probability values on a histogram, we found a large number of false-positive operations fell below a 0.1 threshold. Therefore, these locations were removed from our dataset because they were less likely to be actual poultry operations (Figure 4).

Aggregate barns into operations

After the removal of misclassified features from the remote sensing output, the Python script was used to combine multiple nearby features or barns into a single commercial poultry operation. The FA protocol established the location of individual barns or buildings rather than farms or operations, which often include multiple barns. Geographic centroids were calculated from features (individual barns) within 100 m of each other. The individual barns were converted into a single point feature where the output represented the approximate latitude and longitude of each operation (Figure 5). A barn count (from the remote sensing steps) was generated that calculated the number of barns present at the operation and added to the attribute table.

Adjudication of hybrid operations

The remote sensing methodology was only capable of identifying commercial poultry operations due to the characteristic size,
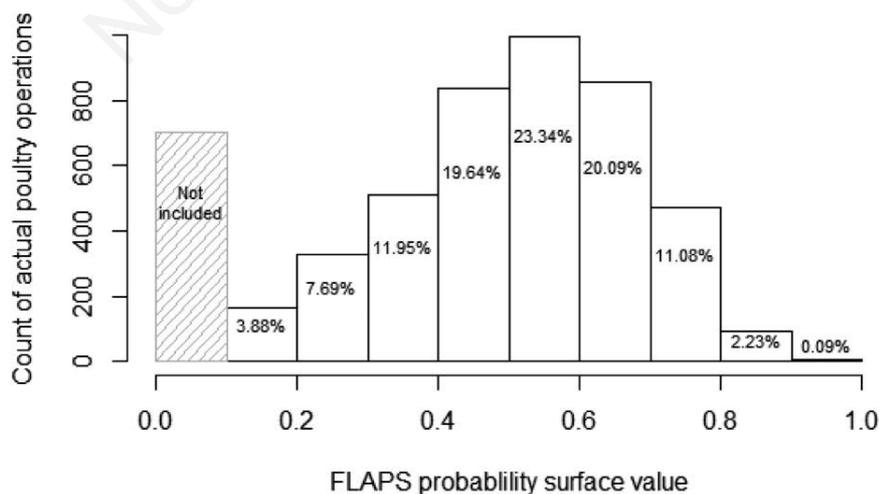
Figure 4. Distribution of values from the probability threshold surface for ground truth poultry operations.

shape, and spectral signature of commercial poultry barns. We defined commercial poultry operations as those with ≥ 10,000 broilers, layers or pullets, or ≥ 2,000 turkeys. These populations are the approximate number of birds housed in the commercial style poultry barns detected by the FA protocol and were based on industry standards that ensure humane conditions with adequate space per bird (Bell and Weaver, 2002; Aviagen Turkeys, 2018; Penn State Extension, 2018).

We adjudicated the remote sensing dataset with the number of expected commercial operations in each poultry-intensive county using the FLAPS probability surface and ground truth data derived from the 35 counties in the pilot study (Maroney *et al.*, 2020), utilizing the pattern previously found in the histogram of the pilot study dataset (explained in Burdett *et al.*, 2015). Whereas previously we were using the histogram to identify locations unlikely to be poultry operations, here we used the histogram to guide the selection of the same number of operations identified for a given county in the CoA. First the operations from the remote sensing dataset were sorted into a frequency distribution according to the value of the FLAPS probability surface at the location of each point. These data were further sorted into a series of 10 bins, ranging from 0 to 1.0 at intervals of 0.1. Points from the remote sensing

dataset were selected, sampled randomly without replacement, proportional to the number of ground truth points observed. The number of points selected for each bin from the remote sensing dataset was calculated with the following equation:

$$Y_b = E * P_b \qquad \text{(Eq. 1)}$$

where Y_b is the number of points to select for each bin, E is the expected number of commercial operations for each county, and P_b is the percentage of ground truth points observed for each bin (Table 1). Note that $P_1 + P_2 + \cdots + P_8 + P_9 = 100\%$. This process ensured that the hybrid dataset would retain the locational accuracy from the remote sensing dataset, while leveraging patterns in the ground truth data to ensure the hybrid dataset would match the expected number of operations in each county, as derived from the CoA.

Assigning poultry production types and population estimates

For poultry-intensive counties, the commercial poultry operations data (latitude and longitude) from the hybrid process were assigned demographic information (production type, number of birds) using output from the FLAPS model. Population sizes were allocated to the commercial operations based on the barn count field.

Figure 5. Individual barn features within 100 m of each other were converted to a single point representing the centroid of a poultry operation.

Table 1. Percentage of ground truth operations assigned to each bin of the FLAPS probability surface.

Bin number	Range of FLAPS probability surface values	Percentage of operations observed in ground truth data
1	< 0.1	4.01*
2	> 0.1 - ≤ 0.2	3.88
3	> 0.2 - ≤ 0.3	7.69
4	> 0.3 - ≤ 0.4	11.95
5	> 0.4 - ≤ 0.5	19.64
6	> 0.5 - ≤ 0.6	23.34
7	> 0.6 - ≤ 0.7	20.09
8	> 0.7 - ≤ 0.8	11.08
9	> 0.8 - ≤ 0.9	2.23
10	> 0.9 - 1.0	0.09

* Features in the < 0.1 bin were excluded from the model.

Operations with larger numbers of barns were allocated a larger number of birds with randomization among operations with identical barn counts. Additionally, location and demographic data for small enterprise operations (broilers, layers, or pullets with < 10,000 birds, and turkeys with < 2,000 birds) and backyard operations (operations with ≤ 1,000 birds) were generated by FLAPS and added to the final population model for completeness.

Merging the hybrid model and FLAPS data to prepare the final dataset

For the remainder of the coterminous U.S. counties (those with ≤ 100,000 birds defined as non-poultry-intensive), the unaltered FLAPS output (locations, numbers of birds, poultry production types) for commercial, small enterprise, and backyard operations was used. The FLAPS output was then combined with the final output from the hybrid model to create a national-level poultry operations and population dataset for the coterminous United States.

Finally, all commercial and small enterprise operations were then assigned to either a production or breeding category. Data from the National Poultry Improvement Program (NPIP) were used to establish an overall percentage of breeding facilities in each state. These proportional data were subsequently distributed according to a frequency distribution of population sizes known to occur on poultry operations in the breeding sector. Full details of the assignment of operations into production and breeding sectors are discussed in Supplement 2.

Validation

The spatial accuracy, distribution, and clustering patterns of the output of both the hybrid and FLAPS models were assessed for a subset of 41 U.S. counties using ground truth commercial poultry operation location data manually digitized from aerial imagery or locally-available actual operation data. The validation methods were executed using a custom Python script in ArcGIS 10.5.1, and included three primary methods referred to as "buffer capture", "grid density", and "ellipse overlap". The county-level outputs generated by the custom Python script were imported into R 3.5.1 (R Core Team, 2015) for data cleaning and statistical analyses using a custom R script.[1]

Validation: buffer capture

The buffer capture method assessed the spatial location accuracy of the hybrid and FLAPS models in identifying commercial poultry operation locations relative to the ground truth data (Figure 6a). Spatial accuracy was represented by the proportion of true operation locations that were within a given distance of hybrid or FLAPS modelled points. Using ArcMap's *Buffer* tool, circular buffers were generated around each ground truth operation location at varying radii (100 m, 500 m, 1,000 m, 2,000 m, 5,000 m). The *Spatial Join* tool was then used to select and count the number of hybrid or FLAPS model points that were completely contained within a ground truth operations buffer. Buffers were assigned a binary "yes/no" outcome if they did or did not contain a hybrid or FLAPS model point. A ground truth operation was assigned a "yes" outcome if it captured a modelled point at the specified radius, indicating that the modelled point was spatially accurate at that resolution. A difference of proportions test was used to compare the percent of ground truth operations buffers that captured a hybrid model point to the percent that captured a FLAPS point at each of the buffer sizes generated.

Validation: grid density

To determine how the distribution patterns of the hybrid and FLAPS models' operation locations compared to those of the true operation locations, root mean square error (RMSE) values for the simulated points compared to true operations were calculated (Figure 6b). For each of the 41 counties, grids of 32 different sizes (varying from 500 m to 60 km) were generated using ArcMap's *Fishnet* tool. Ground truth, hybrid model, and FLAPS model points were counted within each grid cell using the *Spatial Join* tool. RMSE values were calculated using two different approaches, 1)

$$RMSE(r) = \sqrt{\sum_{i=1}^{n} \frac{(M_i(r) - T_i(r))^2}{n}} \qquad \text{(Eq. 2)}$$

where T_i is the number of true operations within a cell of a given size, M_i is the number of hybrid model or FLAPS points within the same cell, and n is the total number of cells of that size within the dataset (modified from Tildesley and Ryan, 2012); and 2)

$$RMSE(r) = \sqrt{\sum_{i=1}^{n} \frac{\left(\frac{M_i(r)}{M_T} - \frac{T_i(r)}{T_T}\right)^2}{n}} \qquad \text{(Eq. 3)}$$

where T_T is the total number of true operations within the dataset and M_T is the total number of modelled points within the dataset. The resulting RMSE values reflected the error in the difference of the proportion of modelled points (hybrid or FLAPS) and the proportion of ground truth points within grid cells of a given size. The RMSE values of the hybrid model were then compared to those of the FLAPS model for each grid cell size to determine the error associated with each model's distribution patterns relative to the distribution of true operations.

Validation: ellipse overlap

The ellipse overlap method quantified the differences in clustering patterns between the hybrid and FLAPS models' points and the ground truth point locations (Figure 6c). First, grid layers were developed at four different resolutions: 500 m, 1000 m, 2000 m, and 5000 m. For each model, operations were assigned and grouped by grid id. Within each grid, standard deviational ellipses were generated using the *Directional Distribution* tool in ArcMap. The *Directional Distribution* tool resulted in ellipses of varying sizes based on the clustering and spatial distributions of the underlying operations. Portions of modelled operations (hybrid and FLAPS) ellipses that overlapped with the ground truth ellipses were extracted using the *Intersect* tool in ArcMap. The proportions of the areas of the ellipses that overlapped between the hybrid or FLAPS models and the ground truth operations were calculated as:

$$P_O = \frac{A_I}{A_M} \times \frac{A_I}{A_T} \qquad \text{(Eq. 4)}$$

where P_O is a measure of the proportion of the areas of the ground truth operations ellipses and modelled ellipses that overlap; A_i is the area of the intersection of ground truth operations ellipses and modelled ellipses (hybrid or FLAPS) within grid cells of a given resolution; A_M is the area of the modelled ellipses (hybrid or FLAPS) at the same size; and A_T is the area of the ground truth operations ellipses

at the same resolution.

A simpler proportion of the area of the ground truth operations ellipses that overlapped with hybrid or FLAPS ellipses (*i.e.*, the second term of the P_O formula) was also calculated and used to conduct difference of proportions testing to compare the hybrid and FLAPS models at each of the varying grid resolutions.

Results

National-level poultry operations and population dataset

A total of 594 counties in 34 states, covering an area of approximately 1,062,365 km^2, were identified as poultry-intensive counties by our definition. Generally, these counties were concentrated in the southeastern and upper central United States. Very few poultry intense counties occurred in the western United States. The hybrid methods

were applied to all of the counties meeting the criteria as poultry intensive. The remaining 2,515 U.S. counties, about 6,727,358 km^2 in area, had ≤ 100,000 birds as reported in the 2012 CoA, and estimates of locations and population demographics in these counties relied on synthesized data from the FLAPS model. After applying FA to the NAIP imagery for the 594 poultry-intensive counties, we had over 1.2 million features (Table 2). A total of 784,378 features were removed by masking, 11,590 by geometry, and 222,760 by the probability surface (Figure 7). After removing false positives, 255,473 barn features remained. These were collapsed to 145,489 operations in comparison to 60,839 operations in the 2012 NASS CoA data. Following the adjudication step, the hybrid model output was reviewed and updated to include latitude and longitude locations for 60,839 individual operations. Demographic information (production type, number of birds) was then assigned to each hybrid operation using output from the FLAPS model. An anonymous example of hybrid model output is shown in Figure 8.

Buffer Capture
- ◆ Captured hybrid model operations
- ◆ Non-captured hybrid model operations
- · Ground truth operations
- ● Ground truth buffer (2 km)

Ellipse Overlap
- ● Area of ellipse overlap
- ● Hybrid model cluster ellipse
- ● Ground truth cluster ellipse
- ◆ Hybrid model operations
- · Ground truth operations
- ☐ 10 km grid

Grid Density
- ◆ Hybrid model operations
- · Ground truth operations
- ☐ 10 km grid

Figure 6. Spatial accuracy, distribution, and clustering patterns of the output of both the hybrid and FLAPS models were assessed by three methods. a) Buffer capture to quantify spatial accuracy of hybrid points relative to ground truth operation locations, 2000 m; b) Ellipse overlap to compare the clustering patterns of hybrid points relative to ground truth operation locations, 10,000 m; c) Grid density to compare the distribution of hybrid points relative to ground truth operation locations, 10,000 m.

Table 2. For each step of the hybrid model, counts of features in the dataset and percentages of features removed.

Hybrid model step	Count of features in the dataset	Percentage of features removed
Initial classification by FA	1,274,201	n/a
Filter: masking	489,823	61.6
Filter: length threshold	478,233	0.9
Filter: FLAPS probability surface	255,473	17.5
Collapse features	145,489	8.6
Hybrid model operations adjudicated with NASS 2012 CoA operations	60,839	6.6

Validation

Overall, validation showed improved accuracy in the spatial location, density, and clustering patterns of hybrid model operations relative to poultry operations modelled by FLAPS.

Validation: buffer capture

At all buffer sizes evaluated, the hybrid model showed statistically significant improvements in the spatial locational accuracy of operations relative to the FLAPS model (Figure 9a). On average, the hybrid model operations were twice as likely to be placed near
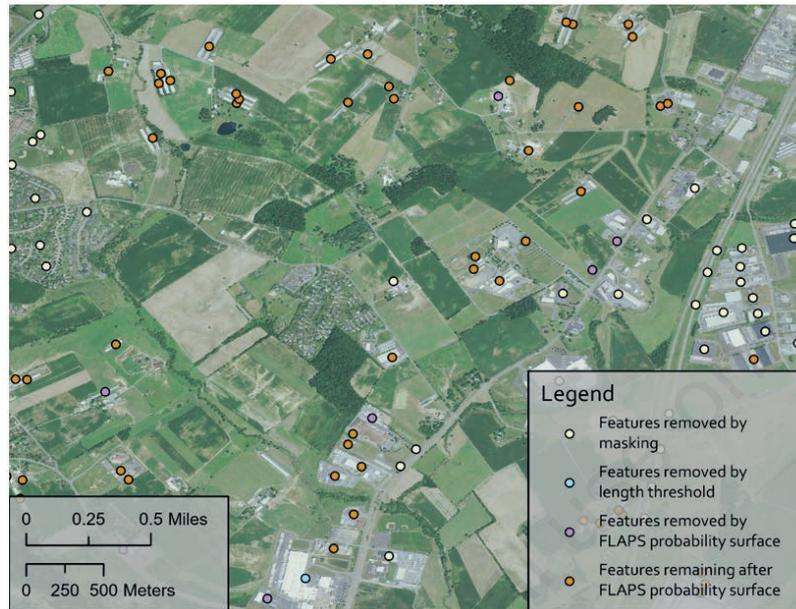
Figure 7. Anonymous example of filtering of the remote sensing results to remove misclassified features through masking, length thresholds, and the FLAPS probability surface.
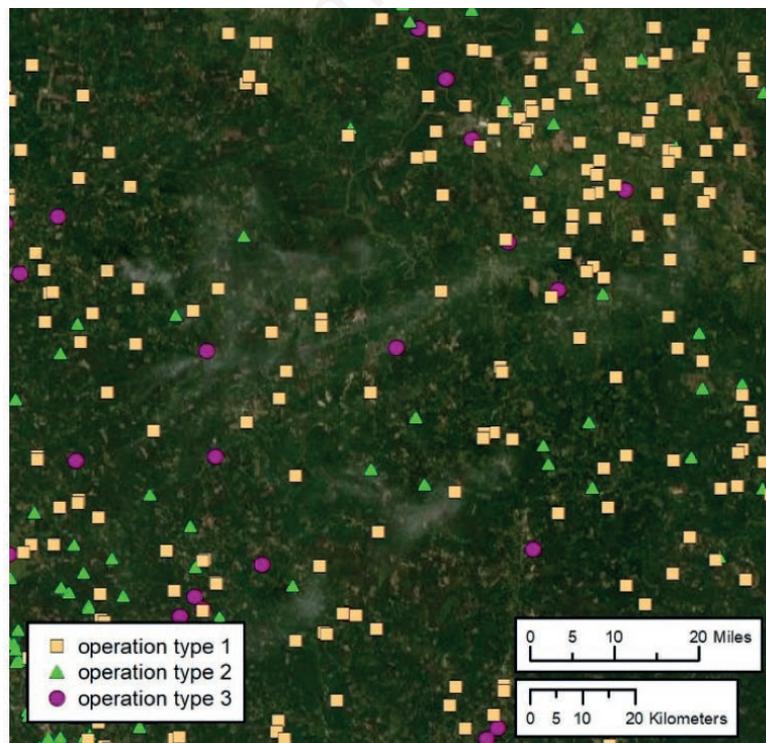
Figure 8. Anonymized example of output from the hybrid model. Imagery source: ESRI, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community.

the ground truth operation locations, evident in the number of operations "captured" in each buffer. Approximately 57% of the 1,000-meter ground truth buffers captured hybrid model operations, compared to a 29% capture of the FLAPS operations (P< 0.01). In short, at the 1,000-m resolution, the hybrid model, was twice as likely as the FLAPS model to place an operation in the same location as a ground truth operation.

Validation: grid density

At all grid cell sizes, the RMSE for FLAPS was greater than for the hybrid model. For example, at a grid size of 30 m, the RMSE for the hybrid model was 18.62, which was a 20.07% decrease in error from the FLAPS model's RMSE value of 23.29 (Figure 9b). Therefore, the distribution of the hybrid model points was more similar to that of the ground truth operations than the distribution of FLAPS model points.

Validation: ellipse overlap

At all ellipse sizes, the product of the proportion of overlap with the hybrid or FLAPS ellipses and the proportion of overlap with the ground truth operation ellipses (P_O) was higher for the hybrid model than the FLAPS model. For example, for the ellipses that were generated within a 10,000-meter grid, P_O for the hybrid model was 0.31, which was 42.2% greater than the P_O for FLAPS, which was 0.18 (Figure 9c). This suggests that the clustering patterns of hybrid model points are a closer representation of the ground truth operations clustering

patterns than those of the FLAPS model.

A difference of proportions test was conducted on the proportion of the ground truth operations ellipses that were overlapped with the hybrid or FLAPS ellipses (*i.e.*, the second term of the P_O formula). At all ellipse sizes, this test revealed that the hybrid model ellipses had significantly greater overlap with ground truth operation ellipses than the FLAPS model ellipses (P<0.01), which confirmed the trend that was observed for P_O.

Discussion

Effective planning and response to animal disease outbreaks relies on knowledge of the livestock and poultry industries in the area of interest (Sanson 1993; Tomassen *et al.*, 2002; Stevenson *et al.*, 2007; Porphyre, *et al.*, 2013). Population data are a critical input for epidemiologic models that are often used to facilitate decision making before and during outbreaks (*e.g.*, Stevenson *et al.*, 2013; Bradhurst *et al.*, 2015). In cases where operation demographic data are scarce, unavailable, or aggregated, synthetic datasets are needed to carry-out preparedness and management activities. The impacts of population data on model uncertainty and results are often not explicitly assessed. However, a few studies have reported that population demographics and the choice of a population dataset used in a model can result in statistically significant differences in model outputs and therefore have implica-
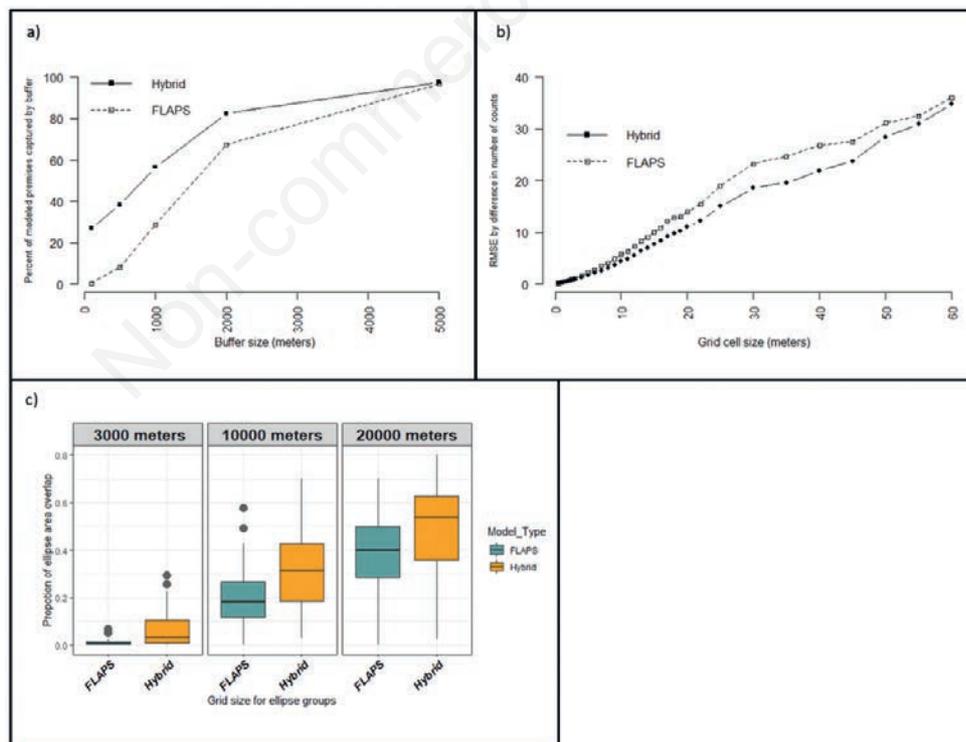
Figure 9. Validation results using three methods. a) Plot of the percent of buffers at varying sizes that capture a hybrid or FLAPS modelled operation and the number of modelled operation captured by a buffer using the buffer capture method to assess spatial location accuracy; b) Root mean square errors (RMSE) for the difference in operation distribution between the FLAPS or hybrid models and the ground truth poultry operation distributions; c) Plots show the proportion of standard deviational ellipse area overlap for the hybrid and FLAPS models with the ground truth data at 3000 m, 10,000 m, and 20,000 m.

tions for decision making (Reeves, 2012; van Andel *et al.*, 2018).

Several methods have been used to generate synthetic animal populations. In the United States, the aggregated CoA data have been used to estimate suppressed data, predict the locations of operations, and estimate operation types by rule-based and probabilistic modelling approaches (Melius *et al.*, 2006; Melius, 2008; Bruhn *et al.*, 2012; Burdett *et al.*, 2015). In New Zealand, counts of animals and livestock units have been assigned to known farm polygons using models of environmental and socioeconomic variables (van Andel *et al.*, 2017). Prosser *et al.* (2011) developed national-level poultry distribution maps for China based on statistical relationships between poultry census data and agro-ecological predictor variables, and Khan *et al.* (2018) modelled the domestic swine population in Ontario, Canada, based on publically available Canadian agricultural census data and geospatial information for factors thought to influence swine operation locations. Livestock operation locations have also been predicted in Europe and Australia using models based on land and climate factors (Emelyanova *et al.*, 2009; Neumann *et al.*, 2009). To our knowledge, this study is unique in its methods to combine remotely sensed and simulated data to generate a database of poultry operation locations, including production demographics, on a national scale.

The objective of this study was to combine two approaches to create an improved spatially-explicit database of individual poultry operations in the United States. We expanded the smart placement concept through the combined use of remote sensing and a probability surface. Our results demonstrate significant improvements in locational accuracy, distribution and density patterns as compared to use of the FLAPS probability surface alone (Table 1). While tools such as masking reduce the likelihood of placing operations in improbable locations, remote sensing classifies actual features characteristic of poultry barns, to identify operations based on reality. Utilizing our knowledge of actual operation locations from the 35 ground truth counties in combination with the probability surface to quantify the numbers and distributions of operations across the poultry-intensive counties improved the overall realism of the hybrid dataset. Further, validation demonstrated that our methods resulted in a comprehensive dataset with improvements in commercial poultry operation location accuracies, density estimates, and spatial distribution patterns relative to actual operation locations. The resulting detailed maps of poultry operations and animal densities improve our knowledge of high-risk poultry-producing areas within the United States. These outcomes can be used to improve our understanding of avian influenza transmission within domestic poultry populations, to evaluate risk at the interface of wild birds and domestic poultry, to improve our abilities to design risk-based surveillance, minimize the risk of disease introduction, and respond to and control future outbreaks. For example, the resulting dataset can be immediately applied in epidemiologic models upon first detection of infection of a reportable avian disease in the United States to allow for rapid estimation of spread and economic consequences for consideration in response measures.

Although the FLAPS model simulates operation locations by placing them in areas similar to true operation locations, remote sensing methods can identify the actual locations of buildings. Ideally, we would have performed remote sensing on all counties in the United States. The demand that this effort would have had on both personnel and computational resources is, however, still a limiting factor. Therefore, we chose to conduct remote sensing on

poultry-intensive counties given the historical outbreak risk associated with high-density poultry-producing regions. Several thresholds (bird numbers per county) to define poultry-intensive counties were explored. The 100,000 birds per county cut-off was chosen after consultation with subject matter experts and included key poultry producing regions and states while accounting for logistic constraints. Improved technologies and additional resources may make remote sensing across a larger geographical area more feasible in the future.

We acknowledge several limitations to our methods. Although remote sensing techniques can identify actual barn locations, it is important to consider the assumptions made in this study. Remote sensing approaches were used to identify buildings having the characteristic size and shape of commercial poultry barns; however, key operation characteristics such as production type and numbers of animals, cannot be directly derived from the imagery. Specifically, production practices, husbandry, and biosecurity measures vary among different production types. These differences as well as information about the number of animals on an operation are important because these factors influence susceptibility to disease, potential for introduction, frequencies and types of direct and indirect contacts, and transmission within and between operations (Bessell *et al.*, 2010; van Andel *et al.*, 2017; van Andel *et al.*, 2018). Further, production type differences and the numbers of animals on operations are considered in outbreak planning and response and are accounted for in epidemiologic and economic modelling. The hybrid approach taken in this study was critical in being able to apply the strengths of the two methods to generate a dataset with both location information (remote sensing) and information about production type and population size (FLAPS). Where remote sensing had knowledge gaps in operation characteristics, information from FLAPS provided those details in a non-random fashion.

We attempted to capture the breadth of possible sizes and shapes of barns in our input criteria for remote sensing, but it is possible that small operations or alternative poultry production systems such as those which allow birds to have outdoor access or where birds are pasture raised, were not identified by our methods. Additionally, while remote sensing can identify a barn, it is unable to distinguish barns in active production from those no longer in use, or to distinguish poultry barns from other agricultural structures. Finally, older imagery may not capture poultry production dynamics such as construction of new barns or demolition of older structures.

The hybrid model focused on commercial poultry operations, as those are most readily identified by remote sensing. Several studies and retrospective analyses of HPAI outbreaks have reported that backyard flocks have a marginal role in HPAI outbreaks, but their contributions to outbreaks should be considered (Terregino *et al.*, 2007; Bavinck *et al.*, 2009; Smith and Dunipace, 2011; Souvestre *et al.*, 2019). The number of pet poultry alone has been estimated at 15.4 million in 1.4 million households in 2016 in the United States, although actual numbers are difficult to determine especially given the dynamic and often transient nature of the backyard population (AVMA, 2018). While some demographic data may be available on a smaller scale (*e.g.*, state, city or county level) particularly following response to an animal disease outbreak, data on backyard flocks does not exist on a national scale in the United States. The CoA captures information from any place from which ≥ $1,000 USD of agriculture products were produced or sold, or normally would have been sold during the census year

(NASS, 2012). Therefore, the CoA captures commercial and small-scale poultry operations and incidentally some hobby poultry (premises generating agricultural revenue from other sources which also have poultry). We included backyard and small enterprise poultry operations from FLAPS (*i.e.*, CoA) in our final dataset for completeness, but these estimates likely greatly underestimate the true numbers of these flocks.

Despite the constraints and potential limitations discussed, validation demonstrated that our methods resulted in a comprehensive dataset with improvements in commercial poultry operation location accuracies, density estimates, and spatial distribution patterns. Given these results, we consider the hybrid approach an improvement over FLAPS and other fully synthetic population datasets, yet this comprehensive approach still does not reflect the locations and distributions of commercial poultry operations with complete certainty. Additionally, we only had the opportunity to do the hybrid approach in poultry-intensive counties. We did not evaluate hybrid model performance in non-poultry-intensive counties and therefore performance in low poultry-dense areas is unknown. If additional data for true poultry operation locations become available in the future, these data would provide opportunities for further validation.

Opportunities for future work include updating data sources as new information is released. For example, the NASS CoA is conducted every five years. At the time this study was conducted, the 2012 CoA data were the most current. New CoA data will reflect any changes in the industry such as the addition or loss of operations, changes in the size of operations, or changes in the geographic locations of operations. In particular, there may be some industry changes as a result of the recent HPAI events in the United States which merit updates to our model. Also, given the known importance of wild birds in the etiology of avian influenza, this national hybrid model allows us to build on previous efforts to explore transmission dynamics at the wildlife domestic poultry interface (Stenkamp-Strahm *et al.*, 2020). Finally, these methods could be applied in other countries lacking comprehensive commercial poultry population data given the availability of appropriate imagery.

Conclusions

Poultry disease outbreaks have the potential to result in substantial losses and impacts to producers, poultry and allied industries, and economies, at local, regional, and national levels. Knowledge of the susceptible population is required for disease managers to adequately prepare for outbreaks and to respond during a disease event; however, a comprehensive poultry population database is lacking in the United States. We developed a novel approach to meet this need which makes use of current technologies and the best available data. This approach could be repeated as new data become available, and the resulting dataset can be applied in epidemiologic models, for planning surveillance activities, estimating resource needs, and for effective decision-making during an outbreak.

References

American Veterinary Medical Association (AVMA), 2018. AVMA Pet Ownership and Demographics Sourcebook. 2017-2018 Edition. Available from: https://www.avma.org/sites/default/files/resources/AVMA-Pet-Demographics-Executive-Summary.pdf

Andronico A, Courcoul A, Bronner A, Scoizec A, Lebouquin-Leneveu S, et al., 2019. Highly pathogenic avian influenza H5N8 in south-west France 2016-2017: A modeling study of control strategies. Epidemics 28:100340.

Aviagen Turkeys, 2018. Management Guidelines: Raising Commercial Turkeys. Available from: https://www.aviagen-turkeys.us/uploads/2015/12/21/Aviagen%20Commercial%20Guide.pdf. Accessed: November 2, 2018.

Backer JA, van Roermund HJW, Fischer EAJ, van Asseldonk MAPM, Bergevoet RHM, 2015. Controlling highly pathogenic avian influenza outbreaks: An epidemiological and economic model analysis. Prev Vet Med 121:142-50.

Bavinck V, Bouma A, van Boven M, Bos MEH, Stassen E, Stegeman JA, 2009. The role of backyard poultry flocks in the epidemic of highly pathogenic avian influenza virus (H7N7) in the Netherlands in 2003. Prev Vet Med 88:247-54.

Bell DD, Weaver WD (eds), 2002. Commercial Chicken Meat and Egg Production. Kluwer Academic Publishers, 819-827, 965-71.

Bessell PR, Shaw DJ, Savill NJ, Woolhouse MEJ, 2010. Statistical modeling of holding level susceptibility to infection during the 2011 foot and mouth disease epidemic in Great Britain. Int J Infect Dis 14:e210-5. doi:10.1016/j.ijid.2009.05.003

Blundell S, Opitz D, Morris M, Rao R, 2008. Feature Analyst V5.0. ASPRS 2008 Annual Conference, 1-9. Available from: http://www.asprs.org/a/publications/proceedings/portland08/0012.pdf

Boender GJ, Hagenaars TJ, Bouma A, Nodelijk G, Elbers ARW, et al., 2007. Risk maps for the spread of highly pathogenic avian influenza in poultry. PLoS Comput Biol 3:e71. doi:10.1371/journal.pcbi.0030071

Bonney PJ, Malladi S, Boender GJ, Weaver JT, Ssematimba A, et al., 2018. Spatial transmission of H5N2 highly pathogenic avian influenza between Minnesota poultry premises during the 2015 outbreak. PLoS One 13:e0204262.

Bradhurst RA, Roche SE, East IJ, Kwan P, Garner MG, 2015. A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. Front Environ Sci 3:17. doi:10.3389/fenvs.2015.00017

Bruhn MC, Munoz B, Cajka J, Smith G, Curry RJ, et al., 2012. Synthesized population databases: A geospatial database of US poultry farms. Methods Rep RTI Press doi:10.3768/rtipress.2012.mr.0023.1201

Burdett CL, Kraus BR, Garza SJ, Miller RS, Bjork KE, 2015. Simulating the distribution of individual livestock farms and their populations in the United States: An example using domestic swine (*Sus scrofa domesticus*) farms. PLoS ONE 10:e0140338.

Caggiano MD, Tinkham WT, Hoffman C, Cheng AS, Hawbaker TJ, 2016. High resolution mapping of development in the wildlife-urban interface using object based image extraction. Heliyon 2:e00174.

Dent JE, Kiss IZ, Kao R, Arnold M, 2011. The potential spread of highly pathogenic avian influenza virus via dynamic contacts between poultry premises in Great Britain. BMC Vet Res 7:59.

Emelyanova IA, Donald GE, Miron DJ, Henry DA, Garner MG, 2009. Probabilistic modelling of cattle farm distribution in

Australia. Environ Model Assess 14:449-65.

Economic Research Service (ERS), 2019. Poultry and Eggs. United States Department of Agriculture. Available from: https://www.ers.usda.gov/topics/animal-products/poultry-eggs/. Accessed: May 8, 2019.

European Space Agency (ESA), 2009. Spectral signatures. Available from: http://www.esa.int/SPECIALS/ Eduspace_ EN/SEMPNQ3Z2OF_2.html. Accessed: June 26, 2019.

Fairchild BD, 2005. Basic Introduction to Broiler House Environmental Control. Available from: http://www.thepoultrysite.com/articles/386/basic-introduction-to-broiler-housing-environmental-control/. Accessed: February 16, 2018.

Freire S, Santos T, Navarro A, Soares F, Silva JD, et al., 2014. Introducing mapping standards in the quality assessment of buildings extracted from very high resolution satellite imagery. ISPRS J Photogramm Remote Sensing 90:1-9.

Johnson KK, Seeger RM, Marsh TL, 2016. Local economies and highly pathogenic avian influenza. Choices, Quarter 2. Available from: http://www.choicesmagazine.org/choices-magazine/theme-articles/economic-consequences-of-highly-pathogenic-avian-influenza/local-economies-and-highly-pathogenic-avian-influenza.

Khan SU, O'Sullivan TL, Poljak Z, Alsop J, Greer AL, 2018. Modeling livestock population structure: a geospatial database for Ontario swine farms. BMC Vet Res 14:31.

Maroney S, McCool-Eye MJ, Fox A, Burdett CL, 2020. Identifying commercial poultry operations from high resolution imagery to support animal health emergencies. Geospat Health 15 [this issue].

Martin MK, Helm J, Patyk KA, 2015. An approach for de-identification of point locations of livestock premises for further use in disease spread modeling. Prev Vet Med 120:131-40.

Meadows AJ, Mundt CC, Keeling MJ, Tildesley MJ, 2018. Disentangling the influence of livestock vs. farm density on livestock disease epidemics. Ecosphere 9:e02294.

Melius C, 2008. Developing poultry facility type information from USDA agricultural census data for use in epidemiologic and economic models. Lawrence Livermore National Laboratory. Available from: https://digital.library.unt.edu/ark:/67531/metadc894967/

Melius C, Robertson A, Hullinger P, 2006. Developing livestock facility type information from USDA agricultural census data for use in epidemiologic and economic models. Lawrence Livermore National Laboratory. doi:10.2172/1036849

NASS, 2012. 2012 Census of Agriculture. United States Department of Agriculture. Available from: https://www.nass.usda.gov/Publications/AgCensus/2012/

NASS, 2015. USDA Poultry Production Data. Fact Sheet. United States Department of Agriculture. Available from: https://www.usda.gov/sites/default/files/documents/nass-poultry-stats-factsheet.pdf.

NASS, 2019. 2017 Census of Agriculture. Summary and State Data. Volume 1. Geographic Area Series. Part 51. AC-17-A-51. United States Department of Agriculture. Available from: https://www.nass.usda.gov/Publications/AgCensus/2017/Full_Report/Volume_1,_Chapter_1_US/usv1.pdf

Neumann K, Elbersen BS, Verburg PH, Staritsky I, Perez-Soba M, et al., 2009. Modelling the spatial distribution of livestock in Europe. Landscape Ecol 24:1207-22. doi:10.1007/s10980-009-9357-5

Patyk KA, Helm J, Martin MK, Forde-Folle KN, Olea-Popelka FJ,

et al., 2013. An epidemiologic simulation model of the spread and control of highly pathogenic avian influenza (H5N1) among commercial and backyard poultry flocks in South Carolina, United States. Prev Vet Med 110:510-24.

Penn State Extension, 2018. Small-flock turkey production. Available from: https://extension.psu.edu/small-flock-turkey-production. Accessed: November 2, 2018.

Porphyre T, Auty HK, Tildesley MJ, Gunn GJ, Woolhouse MEJ, 2013. Vaccination against foot-and-mouth disease: Do initial conditions affect its benefit? PloS One 8:e77616.

Prosser DJ, Wu J, Ellis EC, Gale F, Van Boeckel TP, et al., 2011. Modelling the distribution of chickens, ducks, and geese in China. Agric Ecosyst Environ 141:381-9.

R Core Team, 2015. What is R.? R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.r-project.org/about.html

Reeves A, 2012. Construction and evaluation of epidemiologic simulation models for the within- and among-unit spread and control of infectious diseases of livestock and poultry. PhD thesis. Colorado State University, Fort Collins, CO.Available from: https://mountainscholar.org/bitstream/handle/10217/71580/Reeves_colostate_0053A_11528.pdf?sequence=1&isAllowed=y

Sanson R, 1993. The development of a decision support system for an animal disease emergency. Massey University, Palmerston North, NZ. Available from: https://www.massey.ac.nz/massey/fms/Colleges/College%20of%20Sciences/Epicenter/docs/RobertSansonPhD.pdf?F9BBE59A68A5FF147272D78FF0231CF7

Seitzinger AH, Paarlberg PL, 2016. Regionalization of the 2014 and 2015 Highly Pathogenic Avian Influenza Outbreaks. Choices, Quarter 2. Available from: http://www.choicesmagazine.org/choices-magazine/theme-articles/economic-consequences-of-highly-pathogenic-avian-influenza/regionalization-of-the-2014-and-2015-highly-pathogenic-avian-influenza-outbreaks.

Smith G, Dunipace S, 2011. How backyard poultry flocks influence the effort required to curtail avian influenza epidemics in commercial poultry flocks. Epidemics 3:71-5.

South DD, 2020a. Standard Operation Procedures: Hybrid Model. Available from: https://github.com/VS-DavidSouth/RS_Python_Scripts_GitHub/blob/master/SOP.md

South DD, 2020b. Hybrid Model: Feature Analyst Automated Review Script. Available from: https://github.com/VS-DavidSouth/RS_Python_Scripts_GitHub/blob/master/Automated_Review.py

Souvestre M, Guinat C, Niqueux E, Robertet L, Croville G, et al., 2019. Role of backyard flocks in transmission dynamics of highly pathogenic avian influenza A (H5N8) clade 2.3.4.4, France, 2016-2017. Emerg Infect Dis 25:551-4.

Spröhnle K, Tiede D, Schoepfer E, Füreder P, Svanberg A, Rost T. 2014. Earth observation-based dwelling detection approaches in a highly complex refugee camp environment – a comparative study. Remote Sens 6:9277-97.

Stenkamp-Strahm C, Patyk K, McCool-Eye MJ, Fox A, Humphreys J, et al., 2020. Using geospatial methods to measure the risk of environmental persistence of avian influenza virus in South Carolina. Spat Spatiotemporal Epidemiol 34:100342. doi:10.1016/j.sste. 2020.100342

Stevenson MA, Sanson RL, Miranda AO, Lawrence KA, Morris RS, 2007. Decision support systems for monitoring and main-

taining health in food animal populations. N Z Vet J 55:264-72.

Stevenson MA, Sanson RL, Stern MW, O'Leary BD, Sujau M, et al., 2013. InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. Prev Vet Med 109:10-24.

Terregino C, De Nardi R, Guberti V, Scremin M, Raffini E, et al., 2007. Active surveillance for avian influenza viruses in wild birds and backyard flocks in Northern Italy during 2004 to 2006. Avian Pathol 36:337-44.

Tildesley MJ, House TA, Bruhn MC, Curry RJ, O'Neil M, et al., 2010. Impact of spatial clustering on disease transmission and optimal control. PNAS 107:1041-6. doi:10.1073/pnas.0909047107

Tildesley MJ, Ryan SJ, 2012. Disease prevention versus data privacy: using landcover maps to inform spatial epidemic models. PLoS Comput Biol 8:e1002723.

Tomassen FHM, de Koeijer A, Mourits MCM, Dekker A, Bouma A, et al., 2002. A decision-tree to optimize control measures during the early state of a foot-and-mouth disease epidemic. Prev Vet Med 54:301-24.

TomTom Inc., 2012. Street Centerline GIS dataset [data file]. [producer]. Redlands, CA: Environmental System Research Institute (ESRI) [distributor].

United States Defense Mapping Agency, 2014. Urban Areas GIS Dataset, Digital Chart of the World [data file]. Washington, D.C. [producer] Redlands, CA: Environmental System Research Institute (ESRI) [distributor].

USDA, 2016. Geospatial Data Gateway: Direct NAIP Download. Available from: https://datagateway.nrcs.usda.gov/. Accessed: July 25, 2018.

United States Geological Survey (USGS), 2014. National Hydrography GIS dataset [data file]. Washington, D.C. [producer] Redlands, CA: Environmental System Research Institute (ESRI) [distributor].

United States Geological Survey (USGS), 2019. What is remote sensing and what is it used for? Available from: https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used?qt-news_science_products=7#qt-news_science_ products. Accessed: June 20, 2019.

van Andel M, Jewell C, McKenzie J, Hollings T, Robinson A, et al., 2017. Predicting farm-level animal populations using environmental and socioeconomic variables. Prev Vet Med 145:121-32.

van Andel M, Hollins T, Bradhurst R, Robinson A, Burgman, M, et al., 2018. Does size matter to models? Exploring the effect of herd size on outputs of a herd-level disease spread simulator. Front Vet Sci 5:78.

Werkman M, Tildesley MJ, Brooks-Pollock E, Keeling MJ, 2016. Preserving privacy whilst maintaining robust epidemiological predictions. Epidemics 17:35-41.