Geospatial analysis and modelling in the prevention and control of animal diseases in the United States

Jerome E. Freier, Ryan S. Miller & Kenneth D. Geter

Summary

Geospatial analysis of disease investigation data improves data standardisation and validation and enhances pathogen detection. Grid-based surveillance systems for Newcastle disease in southern California and for bovine tuberculosis on Molokai Island, Hawaii, demonstrate the importance of this approach to operational planning. In addition, as shown by a bovine tuberculosis study in wildlife on Molokai Island, a lattice grid can be used to develop zonal strategies for disease regulatory actions. In risk mapping, disease risk distribution is compared with the distribution of risk factors to identify potential determinants of risk. This process is being applied to North American waterfowl migratory routes to identify geographic areas with high concentrations of migratory waterfowl so that a spatially targeted sampling strategy for use in surveillance operations can be designed. Finally, while farm location data are needed to model pathogen spread through susceptible animal populations, this information is generally unavailable to analysts and modellers. Recently, a farm location and animal population simulator application was developed in which agricultural census data is distributed to create a farm location dataset representative of an agricultural commodity within a specific geographic area.

Keywords

Bovine tuberculosis, Disease surveillance, Geographic information system, Geospatial analysis, Hawaii, Migratory birds, Newcastle disease, Risk mapping, Spatial epidemiology, Simulation models, United States of America, Waterfowl, Wildlife.

Analisi geo-spaziali e creazione di modelli nella prevenzione e nel controllo delle malattie animali negli Stati Uniti d'America

Riassunto

Le analisi geo-spaziali dei dati di diagnosi delle malattie contribuiscono a migliorare la validazione e la standardizzazione dei dati stessi ed incrementano le capacità di diagnosi delle malattie e di isolamento dei patogeni. Il sistema di sorveglianza per la malattia di Newcastle costruito su una griglia nel sud della California e quello per la tubercolosi bovina nell'isola Molokai, nelle Hawai, dimostrano l'importanza di un tale approccio nella pianificazione operativa. Inoltre, come dimostra uno studio sulla tubercolosi bovina negli animali selvatici nell'isola di Molokai, una griglia a reticolo può essere utilizzata per sviluppare strategie operative per il controllo di una stessa malattia che si diversificano da zona a zona. Nella mappatura del rischio, la distribuzione del rischio di una malattia viene comparata con la distribuzione dei fattori di rischio per la stessa malattia al fine di identificare potenziali determinanti del rischio stesso. Questo processo viene applicato alle rotte di migrazione degli uccelli acquatici del Nord America per identificare le aree geografiche con alta

United States Department of Agriculture, Animal and Plant Health Inspection Service, Veterinary Services Centers for Epidemiology and Animal Health, 2150 Centre Avenue, Building B Fort Collins, CO 80526, United States of America
jerome.e.freier@aphis.usda.gov

concentrazione di uccelli acquatici migratori così che si possa disegnare una strategia di campionamento spazialmente targettizzata da utilizzare nella sorveglianza. Infine, sebbene i dati di localizzazione delle aziende siano necessari per modellizzare la diffusione di patogeni all'interno di una popolazione animale suscettibile, questa informazione generalmente non è disponibile per gli analisti e per i modellatori. Di recente è stata sviluppata un'applicazione che simuli la localizzazione delle aziende e delle popolazioni animali nella quale sono distribuiti i dati di censimento agricoli al fine di creare un dataset di localizzazione degli allevamenti rappresentativo di un sistema agricolo in una specifica area geografica.

Parole chiave
Analisi geospaziale, epidemiologia spaziale, fauna selvatica, Hawaii, Malattia di Newcastle, mappatura del rischio, Modelli di simulazione, sorveglianza, Sistema informativo geografico, Stati Uniti d'America, Tubercolosi bovina, Uccelli acquatici, Uccelli migratori.

Introduction

Geospatial processes showing the occurrence of an infectious agent in an animal population are important for understanding the epidemiology and ecology of diseases. In geospatial analysis, spatial measurement, data aggregation processes and statistical methods are used to study population distributions, frequency and proximity of events, discern spatio-temporal patterns and identify spatial relationships among entities and processes within a given area. The use of geographic information systems (GIS) to track animal disease events within animal populations, modelling the spread of infection and planning control strategies has been summarised in detail by Norstrom (2), Boulos (1) and Ruankaew (5). Pfeiffer and Hugh-Jones (3) recognised the value of understanding spatial relationships between natural environments inhabited by wildlife and the pathogens maintained within wildlife populations. There is a need to consider geospatial analysis at the visual, statistical exploratory and spatial modelling levels. In advancing GIS as an analysis tool for epidemiologists, Rytkonen (6) discusses the need for a transition of GIS from visual analyses to a knowledge-based approach in which spatial statistics play a role in data interpretation. With an emphasis on spatial statistics, Robinson (4) has described methods useful for detecting spatial patterns in the distribution of a disease and in identifying causal factors that may account for an observed pattern. Information on spatial patterns of disease spread is important for designing an intervention strategy. An important outcome of exploratory geospatial analysis and spatial statistics is input variables and parameters that can be used in spatial models to predict the frequency of pathogen occurrence and its movement through animal populations (7). Recent efforts within the United States Department of Agriculture's Animal and Plant Health Inspection Service (Veterinary Services) (USDA/APHIS/VS) have focused on applying geospatial methods to address several important animal health issues. These applications include the following:

- use of a lattice grid-based strategy for Newcastle disease virus surveillance in southern California, during an outbreak in 2002-2003 to develop a comprehensive approach in managing this disease outbreak
- application of a similar grid-based approach to define control zones in the management of bovine tuberculosis within feral swine populations on Molokai Island, Hawaii
- design of a spatially targeted surveillance strategy to detect the possible introduction of pathogens into the United States by migratory waterfowl, using a spatial analysis of bird band recovery data
- simulation of agriculturally representative farm location data to use in applications when point-specific farm data are unavailable.

Materials and methods

Grid-based surveillance for Newcastle disease in California

During 2003 in the vicinity of Los Angeles, California, federal and state field teams used wide area augmentation system- (WAAS) enabled Garmin eTrex Legend (www.garmin.com) global positioning system (GPS) receivers to collect geographic coordinates for each Newcastle disease investigation site (including affected premises and all neighbouring premises with susceptible birds within 1 km of an infection site). To generalise point location data and make it easier to assess infection intensity, a 4 × 4 km lattice grid was created as a vector object using TNTmips® software, version 7.0 (MicroImages, Lincoln, Nebraska) and each cell was assigned a number. In addition, a 2 × 2 km inner grid was created and each quadrant was given a letter designation. Both lattice grids were converted to an Environmental Systems Research Institute (ESRI) shapefile format for use in analyses. Around each site containing one or more infected birds, geoprocessing methods were used to determine which grid cells were intersected by a 1-km buffer surrounding an affected premises, so that all nearest neighbouring cells ('queen' contiguity) were selected. Additional cells were added to create a regulatory zone for intensified surveillance. Data validation procedures were applied to geographic coordinate data to confirm positional accuracy by comparing coordinates obtained from GPS receivers with those determined from address geocoding, high-resolution aerial photographs, or land parcel data. Information from each investigation was entered into the USDA/APHIS/VS Emergency Management Response System (EMRS), a centralised, multi-user data system for case management, task assignments, and data analysis. Site specific data were transferred from the EMRS to a structured query language (SQL) Server 2000 (Microsoft Corporation, Redmond, Washington) geodatabase for additional processing and spatial analysis, using ArcSDE® version 8.3 and ArcGIS™ version 8.3 (ESRI, Redlands, California, www.esri.com).

As accuracy of geographic coordinate data was essential in assigning surveillance tasks and in performing analyses, data validation procedures were applied. Data validation involved geocoding each address using a high-accuracy dataset provided by Tele Atlas® (www.teleatlas.com). The distance between a pair of GPS-collected coordinates and a pair of coordinates obtained through address geocoding was then compared. All discrepancies were investigated to verify the accuracy of every coordinate pair reported. Geocoding and processing of GIS data were performed with ArcGIS™ for ArcInfo™ versions 8.2 and 8.3.

Geodata processing and spatial analysis procedures were conducted with ArcGIS™ for ArcInfo® 8.3 or more recent versions. Animal sample locations were reported as latitude and longitude coordinates in decimal degrees with a World Geodetic System 1984 (WGS84) datum. For spatial analysis and mapping, data were converted to a universal transverse mercator (UTM) projection, using the North American Datum 1983 (NAD83).

Zonal strategy for management of bovine tuberculosis in feral swine on Molokai Island, Hawaii

This study was conducted on Molokai Island, which is located centrally in the Hawaiian archipelago. The island is 64-km long and 7-km wide, with an approximate land area of 671 km^2. Vegetation and terrain vary greatly: the eastern half of the island is mountainous, wetter, and characterised by dense vegetation; whereas the drier western half is less mountainous and has extensive open areas with less dense vegetation.

Hunters of feral swine were requested to bring recently killed animals to a central sample-collection station. Tissue samples from mediastinal and mandibular lymph nodes were obtained, along with tissue from any observed lung lesions, and sent to the USDA/APHIS National Veterinary Services Laboratory (NVSL) in Ames, Iowa. Histopathological examination of samples was performed and pathogen isolation in cell culture was attempted. Bacterial culture was necessary to provide sufficient material for assays that distinguished *Mycobacterium bovis* from other species of *Mycobacterium*. Hunters were asked to identify, as accurately as possible, the site where each pig was killed by using a detailed map of the area where the hunt took place. Hunter maps were sent to the Centers for Epidemiology and Animal Health (CEAH) in Fort Collins, Colorado, where kill sites were digitised and the information about each animal sampled was added to a database for spatial analysis. The infection status of tissues sampled was added to the database from diagnostic reports provided by the NVSL.

Risk mapping approach in identifying concentrations of migratory birds

Bird band recovery data on a continental scale were provided by United States Geological Survey, Patuxent Wildlife Research Center Bird Banding Laboratory in Patuxent, Maryland. The North American Bird Banding Program is jointly administered by the United States Department of Interior and the Canadian Wildlife Service. The recovery of bands was reported in 10-min blocks of degrees for longitude and latitude (~100 mile^2). For analysis, the data were summarised for the last 15 years (1991-2006). This resulted in a continental representation of band recoveries for the fall migration over the last 15 years, accounting for a total of 241 619 observations. Functional groups of migratory waterfowl included in the analysis were dabbling ducks (*Anas* spp.), dark geese (*Branta* spp.), light geese (*Chen* spp.) and swans (*Cygnus* spp.). The analysis was first conducted on data from waterfowl originally banded in Alaska and Asia and, secondly, for all waterfowl banded in northern locations of Alaska, Asia and Canada.

Spatial modelling to simulate farm locations

Data on the specific location of livestock and poultry operations in the United States are not available, except as generalised information provided by the USDA National Census of Agriculture (www.nass.usda.gov/Census_of_Agriculture). This report summarises the number of farms and the number of animals per farm for each commodity by county, at the finest spatial resolution available. Data on the number of farms for a livestock or poultry commodity within most postal zip codes are also available; however, the numbers of animals per farm are only reported at the county level.

The first step in determining where farms might be located within a geographic area is to remove all locations where a livestock or poultry operation would not be expected to be situated. Examples include urban areas, lakes, parks and selected federal lands. A processing mask was created in ArcGIS™ for ArcInfo™ version 9.1, which excluded unlikely locations for farms. After unlikely farm locations were excluded, a processing mask was created based on a 300-m buffer of major roads in likely farm areas. The series of geoprocessing steps used in simulating farm locations was managed by the Model Builder Extension of ArcGIS™ for ArcInfo™ version 9.1.

Results and discussion

Grid-based surveillance for Newcastle disease in California

In establishing a lattice grid to use in surveillance for Newcastle disease in southern California, the first step was to determine a grid cell size that reflected a neighbourhood-level scale containing a population of poultry where virus transmission could occur. In the case of backyard poultry in suburban and rural areas, a grid cell size of 4 × 4 km was considered to represent neighbourhood populations relative to the density of premises with susceptible poultry. With respect to other locations, cell size will vary depending on the density of the population area under investigation and whether surveillance is being conducted in an urban, suburban or rural area. An example of a 4 × 4 km major grid with an inner 2 × 2 km minor grid is shown in Figure 1. Grid cells may be subdivided into smaller surveillance task areas in locations with a high density of animals or livestock operations. An important advantage in using grid cells is that the surveillance area is completely enclosed by the cells, providing for a more thorough coverage than if surveillance were based on circular zones. In areas where a more rounded grid is desired, hexagon-shaped grid cells might be used.

In creating zones based on grid cells, Figure 2 shows an example of a 4 × 4 km grid cells that

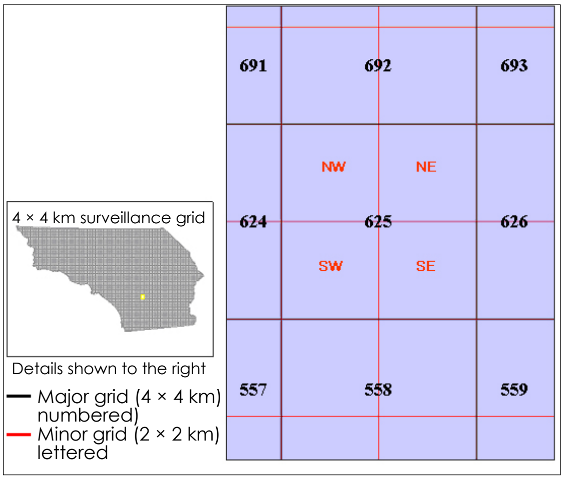

Figure 1
Lattice grid for spatially targeted surveillance of areas for Newcastle disease in California in 2003 with a 4 × 4 km major grid and an inner 2 × 2 km minor grid

contained one or more premises with birds infected with Newcastle disease virus. Cells serving as buffers around infected cells were added, as shown in Figure 3. Grid cells in the region may be further combined, as shown in Figure 4, to create and prioritise zones for control of infection and disease surveillance.

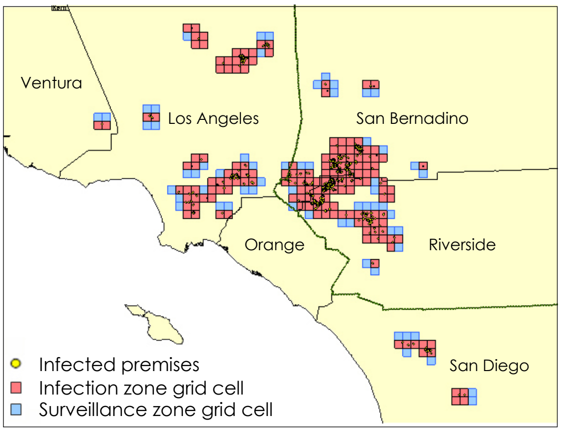

Figure 2
Grid cell selection based on proximity to affected premises: a first step in creating zones of infection and surveillance
An example of targeted surveillance for Newcastle disease in California in 2003 is shown

An important benefit of grid-based approaches to surveillance and incident planning is the division of geographic areas into manageable units that represent the neighbourhood or community structure in which an event is occurring.

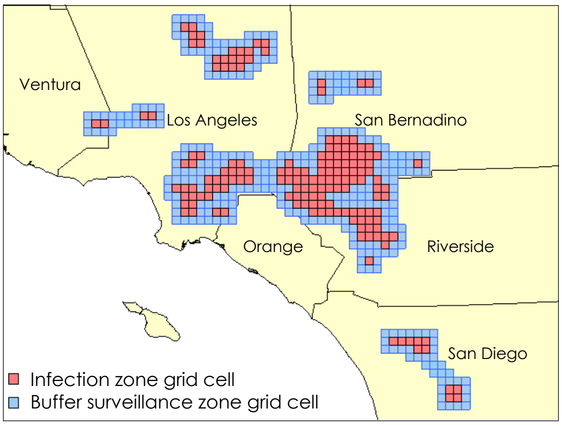

Figure 3
Intermediate step in building infection and surveillance zones by adding additional cells to each area to create discrete geographical units
An example of targeted surveillance for Newcastle disease in California in 2003 is shown

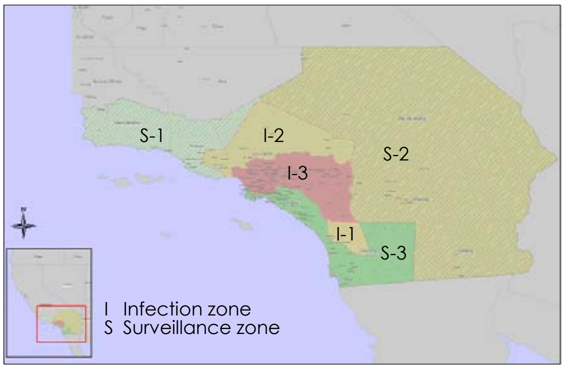

Figure 4
Newcastle disease control in southern California in 2003: aggregation of infection and surveillance zones into separate areas prioritised by level of difficulty in conducting inspections or applying disease control measures
Numbers represent the priority order of response efforts, beginning with areas having the lowest host population density (surveillance zone) and lowest infection intensity (infection zone)
(1 = lowest difficulty, 2 = medium difficulty and 3 = highest difficulty)

Zonal strategy for management of bovine tuberculosis in feral swine on Molokai Island, Hawaii

A minimum of three types of zones are recommended in managing an infectious disease, namely: infection, surveillance and free zones. In developing zone-specific control strategies for an infection zone, it is important to identify all sources of infection and determine the location of animal populations that are susceptible to infection. In a potentially intermixing population of animals at risk of infection, distances between infected animals and the most distant susceptible animal should be calculated. To avoid underestimating, a buffer area should be extended around the infection area originally calculated. Typically, a buffer extension should add ≥20% to the total area, depending on local conditions and confidence that infection sources have been identified. Figure 5 shows an example of an infection zone determined with information on the distribution of

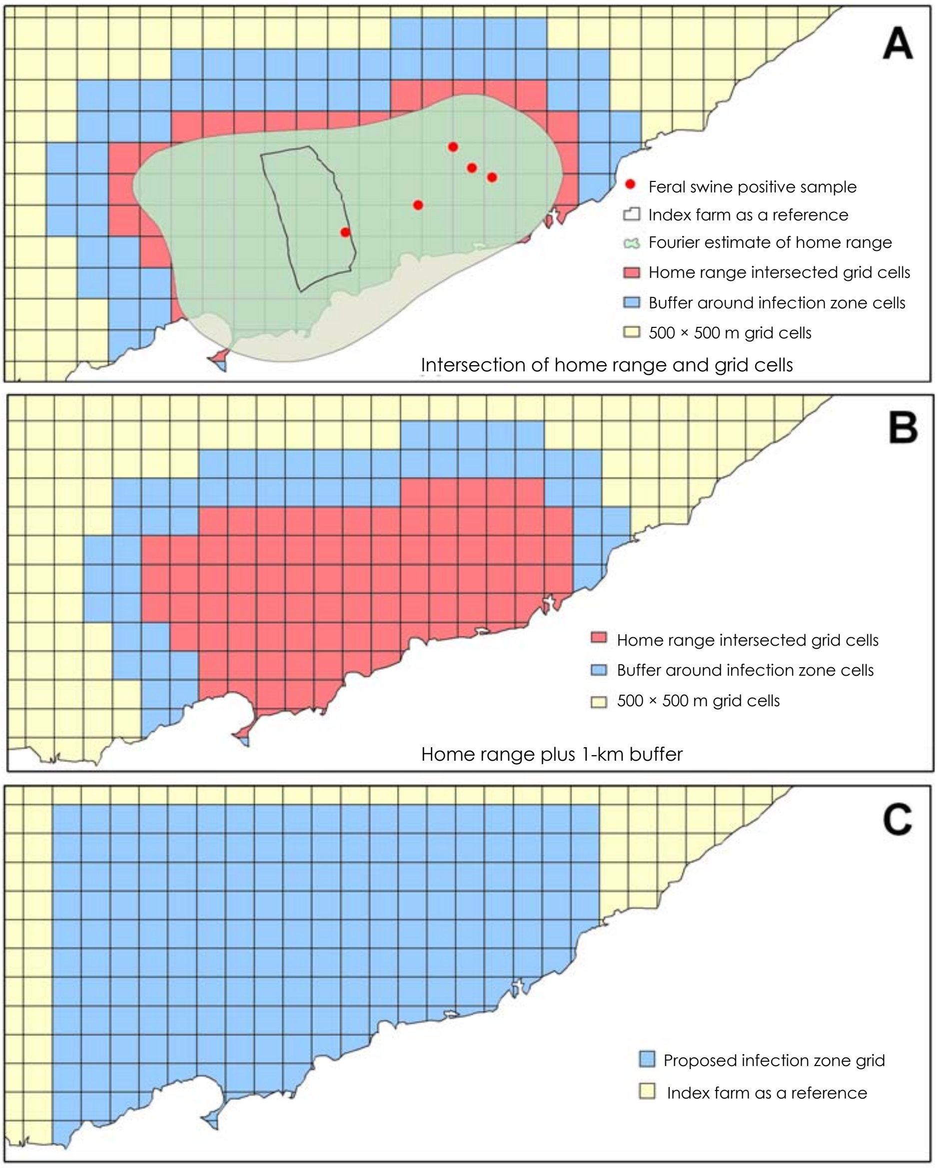

Figure 5
Use of buffer zones (A) around test-positive samples from feral swine and a home range estimate of movement by infected animals combined with (B) a lattice grid (500 × 500 m) to establish (C) a zone of infection for bovine tuberculosis on Molokai Island, Hawaii

infected feral swine relative to grid cells used to divide the region. Grid cells (500 × 500 m) intersecting the home range distribution of infected feral swine were used to define the initial infection area. A 1-km buffer area surrounding the initial infection zone was added. Finally, additional cells were included in the infection zone to establish a more regular and uniform array of cells to use for regulatory purposes.

A surveillance zone should be based on susceptible animal populations located in proximity to an infection zone, but may be at a lower risk of infection based on population isolation, or other factors. Areas outside either an infection or surveillance zone, where there is no evidence of infection in susceptible hosts, should be considered free zones. An example of proposed infection, surveillance and free zones for bovine tuberculosis in feral swine on Molokai Island is shown in Figure 6. This figure also shows that once zones are established it is useful to display sample and at-risk populations relative to the zones.

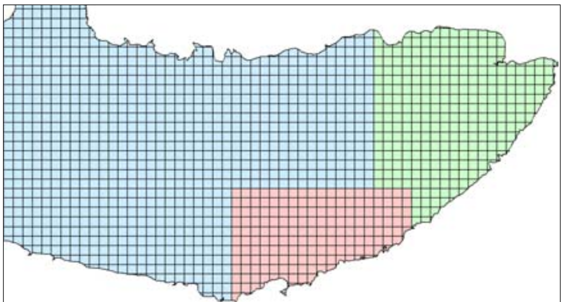

Figure 6
Application of a lattice grid strategy to define zones of infection (red), surveillance (green) and disease freedom (blue) for bovine tuberculosis in feral pigs on Molokai Island, Hawaii

Risk mapping approach in identifying concentrations of migratory birds

An analysis of bird band data for birds banded in northern Alaska or north-east Asia and recaptured in the United States (Fig. 7) shows that most recaptured occur in the Pacific Northwest, the central valley of California, and the Texas Gulf Coast. Birds banded in other parts of Alaska and Canada are recaptured mostly in areas of the lower Mississippi River, the Gulf Coast and the central Atlantic coast. After determining the number of bird bands recovered within each county and categorising every band by its geographic origin, maps were created to show which counties in the United States received populations of migratory birds from each flyway crossing the northern North American continent. These risk maps are being used by wildlife officials to prioritise sample sites for surveillance to detect pathogens that may be introduced by migratory birds. Risk mapping strategies are important in developing spatially targeted surveillance systems to optimise the likelihood of finding an event or incident case with a low frequency of occurrence.

Spatial modelling to simulate farm locations

Specific information about the location of livestock and poultry operations in the United States is not available, except as summarised data provided by the USDA Census of Agriculture. This report contains county-level information about the number of farms and the number of animals per farm for each commodity category. It is common to use agricultural census data to create dot density maps showing the location of randomly placed farms within a county. However, areas exist within each county where a farm is unlikely to be situated. Therefore, it is possible to create exclusion zones that avoid randomly placing farms in unsuitable areas. In addition, other factors (for example, transportation networks and proximity to urban areas or markets) may be used to weight the distribution of random points when they are added to a map. Figure 8 shows an example of points added back to a processing mask for counties in North Carolina where exclusion zones and transportation routes were used to create a farm distribution representative of commercial chicken broiler operations in that state. Information, representing the estimated distribution of an animal commodity, may be

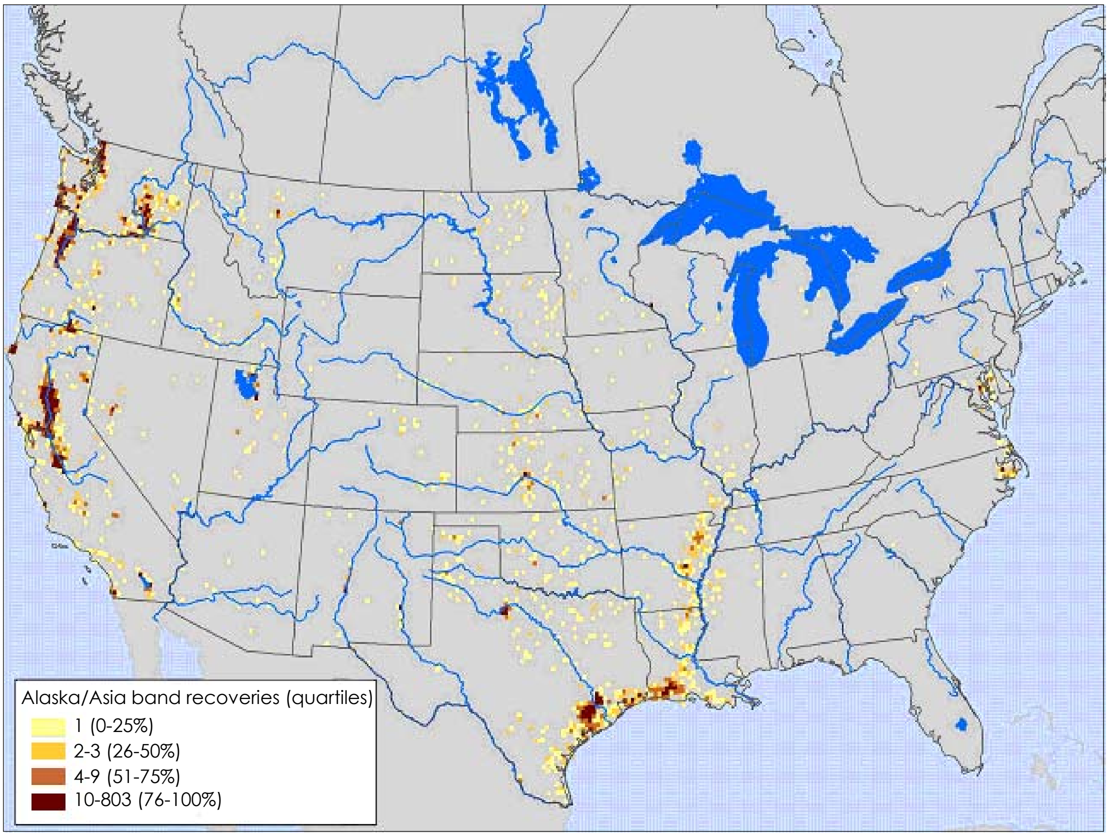

Figure 7
Location of sites in the United States from which bird bands were recovered from migratory waterfowl originally banded in Alaska or northern Asia between 1991 and 2005

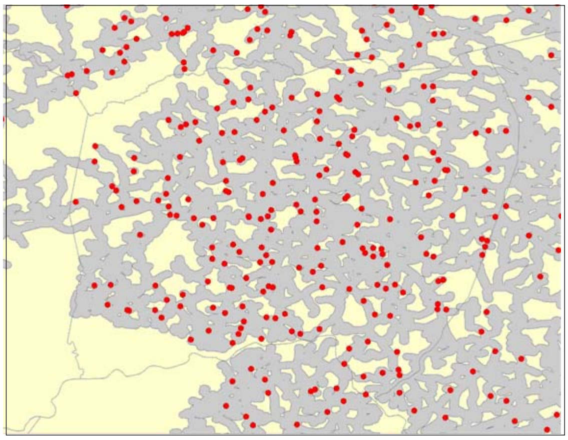

Figure 8
Mask layer (grey polygons) created to restrict random placement of farms (red dots) to areas where farms are likely to be found in central North Carolina, United States

used in simulation models to predict pathogen spread among populations of specific farm animals and the effect of mitigation strategies on disease spread.

References

1. Boulos M.N.K. 2004. Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. *Int J Health Geogr*, **3**, 1-50.
2. Norstrom M. 2001. Geographical information system (GIS) as a tool in surveillance and monitoring of animal diseases. *ActaVet Scand Suppl*, **94**, 79-85.

3. Pfeiffer D.U. & Hugh-Jones M. 2002. Geographical information systems as a tool in epidemiological assessment and wildlife disease management. *Rev Sci Tech*, **21**, 91-102.
4. Robinson T.P. 2000. Spatial statistics and geographical information systems in epidemiology and public health. *Adv Parasitol*, **47**, 81-128.
5. Ruankaew N. 2005. GIS and epidemiology. *J Med Assoc Thai*, **88**, 1735-1738.
6. Rytkonen M.J. 2004. Not all maps are equal: GIS and spatial analysis in epidemiology. *Int J Circumpolar Health*, **63**, 9-24.
7. Schroder W. 2006. GIS, geostatistics, metadata banking, and tree-based models for data analysis and mapping in environmental monitoring and epidemiology. *Int J Med Microbiol*, **40**, 23-36.