



# Representation of animal distributions in space: how geostatistical estimates impact simulation modeling of foot-and-mouth disease spread

Linda Highfield, Michael Ward, Shawn Laffan

## ► To cite this version:

Linda Highfield, Michael Ward, Shawn Laffan. Representation of animal distributions in space: how geostatistical estimates impact simulation modeling of foot-and-mouth disease spread. *Veterinary Research*, 2008, 39 (2), pp.1-14. 10.1051/vetres:2007055 . hal-00902897

**HAL Id: hal-00902897**

**<https://hal.science/hal-00902897>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Representation of animal distributions in space: how geostatistical estimates impact simulation modeling of foot-and-mouth disease spread

Linda HIGHFIELD<sup>1\*</sup>, Michael P. WARD<sup>1</sup>, Shawn W. LAFFAN<sup>2</sup>

<sup>1</sup> Department of Veterinary Integrative Biosciences, Texas A&M University College of Veterinary Medicine & Biomedical Sciences, College Station, TX 77845-4458, USA

<sup>2</sup> School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

(Received 13 February 2007; accepted 22 October 2007)

**Abstract** – Modeling potential disease spread in wildlife populations is important for predicting, responding to and recovering from a foreign animal disease incursion. To make spatial epidemic predictions, the target animal species of interest must first be represented in space. We conducted a series of simulation experiments to determine how estimates of the spatial distribution of white-tailed deer impact the predicted magnitude and distribution of foot-and-mouth disease (FMD) outbreaks. Outbreaks were simulated using a susceptible-infected-recovered geographic automata model. The study region was a 9-county area (24 000 km<sup>2</sup>) of southern Texas. Methods used for creating deer distributions included dasymetric mapping, kriging and remotely sensed image analysis. The magnitudes and distributions of the predicted outbreaks were evaluated by comparing the median number of deer infected and median area affected (km<sup>2</sup>), respectively. The methods were further evaluated for similar predictive power by comparing the model predicted outputs with unweighted pair group method with arithmetic mean (UPGMA) clustering. There were significant differences in the estimated number of deer in the study region, based on the geostatistical estimation procedure used (range: 385 939–768 493). There were also substantial differences in the predicted magnitude of the FMD outbreaks (range: 1 563–8 896) and land area affected (range: 56–447 km<sup>2</sup>) for the different estimated animal distributions. UPGMA clustering indicated there were two main groups of distributions, and one outlier. We recommend that one distribution from each of these two groups be used to model the range of possible outbreaks. Methods included in cluster 1 (such as county-level disaggregation) could be used in conjunction with any of the methods in cluster 2, which included kriging, NDVI split by ecoregion, or disaggregation at the regional level, to represent the variability in the model predicted outbreak distributions. How animal populations are represented needs to be considered in all spatial disease spread models.

**spatial modeling / epidemic modeling / deer density / sensitivity analysis**

## 1. INTRODUCTION

The aim of this research is to evaluate the sensitivity of epidemic model predictions to estimated spatial distributions of wildlife species. This research provides critical insight into the impact that the estimated spatial distribution has on modeling predictions. This is important because modeling predictions are

used to guide policy and evaluate mitigation strategies prior to an outbreak<sup>1</sup> [8, 37]. Modeling may also be used during an outbreak

<sup>1</sup> Taylor N., Review of the use of models in informing disease control policy development and adjustment, Department for Environment, Food, and Rural Affairs, London, 2003, available online at: <http://www.defra.gov.uk/science/documents/publications/2003/UseofModelsInDiseaseControlPolicy.pdf> [accessed 18 June 2007].

\* Corresponding author: [lhightfield@cvm.tamu.edu](mailto:lhightfield@cvm.tamu.edu)

to inform response strategies, particularly for wildlife populations. There are many questions that may be of interest to policy- and decision-makers either before or during an outbreak<sup>1</sup>. Two potential questions address the predicted average outbreak size and the 'best' or 'worst' case scenarios<sup>1</sup>. The former question relates to the mean or median predicted outbreak size, and this information is useful for investment decisions and ranking the importance of foreign animal disease outbreaks. The latter question addresses the minimum and maximum predicted outbreaks, information that can be used by decision-makers for potential resource allocation. Decision- and policy-makers may also be interested in specific values that can be selected from the distribution of predicted outbreaks<sup>1</sup>. Since model predictions are obtained via Monte Carlo simulations, the entire distribution of predicted outbreaks should be considered when making decisions, rather than focusing on just central tendency or even variability statistics<sup>1</sup>.

Given the above, this paper addresses the impact of 15 different geostatistical methods for estimating the spatial distribution of animal species distribution on the predicted outbreak distribution of foot-and-mouth disease (FMD) in white-tailed deer in southern Texas.

FMD is a highly contagious viral disease of cloven-hoofed animals, affecting both domestic and wild Artiodactyla species, including cattle, deer, and feral hogs. It is considered one of the most serious diseases of livestock [28] and the economic devastation caused by outbreaks can be enormous, as evidenced by some recent outbreaks [19, 38].

The severity of the disease in Cervidae spp. (deer) varies from unapparent or mild in some species to more severe in others [28]. Deer have been infected both naturally and experimentally<sup>2</sup> [14, 16, 28], and deer-to-deer and deer-to-cattle transmission has been observed [34]. Experimentally infected white-tailed deer exhibited intermediate disease

severity compared with susceptible livestock species (i.e., between cattle, sheep, and goats) and approximately 10% of those infected in a 1924 outbreak in California displayed typical signs of FMD infection<sup>2</sup>. Several species of deer are among the most commonly FMD-infected wildlife under field conditions, and deer are believed to play an important role in the epizootology of the disease [28].

The United States has been free of FMD since 1929, following a number of outbreaks in California and Texas during the 1920s. In 1924, an outbreak in California which started in pigs, spread to grazing cattle and subsequently infected deer across the central portion of the state. It required two years to eradicate the disease from the local deer population, and more than 22 000 deer were slaughtered<sup>2</sup> [21].

During the 2001 FMD outbreak in the UK, it was feared that a number of the deer species in the country (red, fallow) might become infected and potentially act as a reservoir for the disease [5, 34]. A similar concern was also expressed in the Netherlands during the 2001 FMD outbreak [11, 34]. However, evidence of infection in deer was not observed in either of these more recent outbreaks [11].

In areas of the United States where livestock are extensively grazed, the potential for interaction with susceptible wildlife species, such as white-tailed deer, is high [37]. Deer move through and forage in fields between farms, and enter premises with animal feed and slurry [34]. Additionally, supplemental feeding of white-tailed deer for hunting purposes is common [37]. Given the widespread distribution of wildlife species susceptible to FMD virus infection and the potential for interaction with livestock, modeling the spread of the disease in wildlife populations is an important resource in our ability to predict, respond to and recover from a foreign animal disease incursion.

To model the spread of FMD in a wildlife population, such as white-tailed deer, an estimate of the species spatial distribution is critical. A variety of spatial estimation methods have been used in the ecological literature for modeling the density and distribution of wildlife species, including dasymetric map-

<sup>2</sup> McVicar J.W., Suttmoller P., Ferris D.H., Campbell C.H., Foot and mouth disease in white-tailed deer: clinical signs and transmission in the laboratory, Proceedings of the 78th Annual Mgt. US Animal Health Association, 1974, pp. 169–180.

ping, regression-based approaches (including interpolation), and remotely sensed image analysis.

Dasymetric mapping, also known as surface based demographic data representation, redistributes the population from a set of areal units into either a vector or raster map using ancillary data such as land use or remotely sensed images [27]. Dasymetric mapping provides a means of visually representing a statistical distribution in a geographic information system (GIS) using discrete areal data [6]. This mapping procedure eliminates the artificial structure of political or arbitrary boundaries that are often placed on aggregate data and allows for representation of a more realistic distribution of the data [6, 24].

Regression approaches that have been used to describe the density and distribution of wildlife species include ordinary least squares, logistic, Poisson, and more recently, geographically weighted regression and kriging. Kriging predicts values at unsampled locations based on the autocorrelation structure in the measured observations and on the values of nearby observations, thereby taking into account both distance and geometry [3, 31].

Use of remote sensing in epidemiology is based on the development of a logical sequence linking measures of radiation from a sensor on an aircraft or a satellite, to measures of a disease and its corresponding vector or host [4]. One of the most commonly used types of remotely sensed data is the normalized difference vegetation index (NDVI). NDVI data is available from the United States geological survey (USGS) advanced very high resolution radiometer (AVHRR) database. AVHRR data are collected by the National Oceanic and Atmospheric Administration's (NOAA) polar earth-orbiting satellites, which collect data in the visible, near-infrared, and thermal infrared regions of the electromagnetic spectrum. NDVI is associated with photosynthetically active radiation, and is the most commonly used vegetation index [22].

Epidemics have historically been modeled using differential equations [1, 10, 33]. However, these models do not directly address the local character of disease spread and can-

not handle complex boundary conditions [33]. Geographic automata (GA) are a generalization of cellular automata models, capable of handling non-tessellated data (points). Both cellular and geographic automata provide an alternative to differential equation based epidemic models. These models treat time as discrete and interactions as localized [33] and have been applied to a wide range of disease spread problems [1, 2, 9, 17, 32, 36]. Susceptible-infected-recovered models are often built into geographic automata to examine the spatial and temporal propagation of epidemics [2, 13, 17, 23, 32, 33]. However, this approach has rarely been used to model infectious diseases in wildlife populations. Importantly, the impact of the methods used to derive the underlying wildlife distribution estimates on epidemiological modeling predictions has, to the authors' knowledge, not been evaluated.

The objectives of this study were: (1) to apply 15 commonly used geostatistical methods to estimate white-tailed deer distributions in southern Texas; (2) describe the predicted FMD outbreak distribution that might be observed, given the various geostatistical estimation methods used; and (3) compare the predicted FMD outbreak distributions for each of the geostatistical methods used.

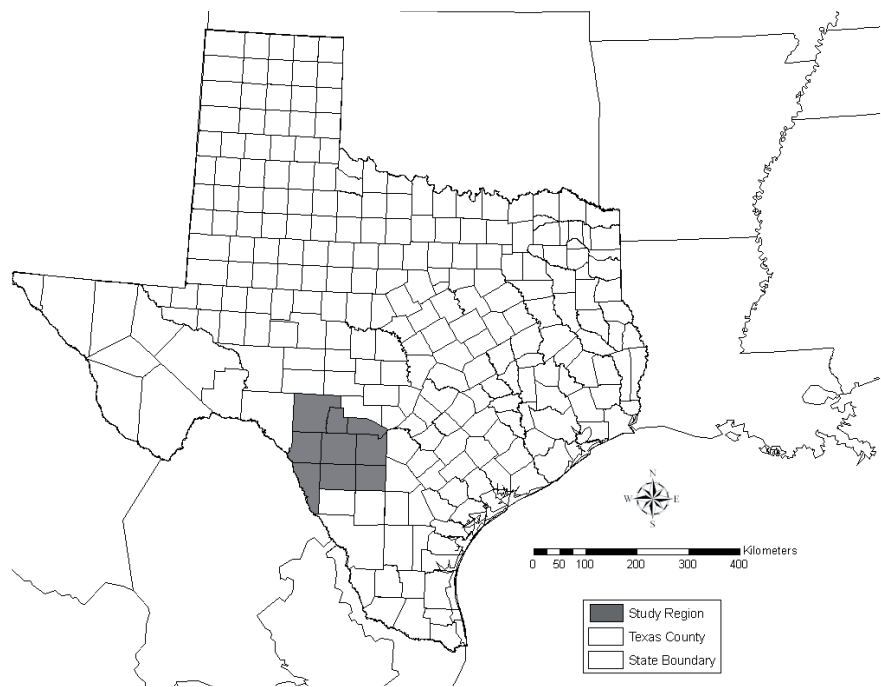
## 2. MATERIALS AND METHODS

### 2.1. Study site

The study site (approximately 24 000 km<sup>2</sup>) is an area of southern Texas, bordering Mexico, consisting of nine counties (Fig. 1). It consists of two ecoregions, the Edwards Plateau in the north and the South Texas Brush in the south. The annual rainfall ranges between 750 and 1 200 mm.

### 2.2. Data source

Data on the number of deer per reporting unit and per county were obtained from the Texas Parks and Wildlife Department (TPWD) [25]. The distribution of deer was estimated using geostatistical methods. All spatial processing was done using ArcGIS 9.0 (ESRI, Redlands, CA, USA) and all surfaces were estimated either on a square lattice with a 1 km<sup>2</sup> resolution or as a point representation at 1 km<sup>2</sup> resolution.



**Figure 1.** South Texas study region.

**Table I.** Count and percentage of suitable pixels per land use category.

Land use category	Count	Percentage
Forest	7 692 254	0.27
Shrub	16 774 976	0.14
Grassland	4 049 744	0.59
Total	28 516 974	1.00

**2.2.2.1. Disaggregation based on county numbers (DC)**

County-level deer populations were disaggregated based on suitable landuse and estimated carrying capacity within each land use category. Land use categories were derived from the 1992 national land cover dataset (NLCD). The NLCD grid was clipped to the study site and reclassified into suitable categories as follows: forest (land use classes 41, 42, and 43), shrub (classes 50 and 51), and grassland (class 71). The proportion of each land use category in the study area is shown in Table I. Estimated carrying capacity was derived from expert opinion and yielded values of 0.3 for forest, 0.3 for shrub, and 0.1 for grassland for this region. The

number of pixels per land use category was multiplied by the carrying capacity as a weighting factor. The number of deer per county was proportionally distributed within land use category based on the weighting factor for each category. The resulting fractional counts of deer at 30 meter resolution were aggregated to a 1 km<sup>2</sup> integer grid.

**2.2.2. Focal smoothing (FS1-FS3)**

To eliminate demarcation lines that appeared in the DC method, this data was smoothed using the mean value of cells within a rectangular window around each cell (a focal mean). Three window sizes were used: 10 km<sup>2</sup> (FS1), 20 km<sup>2</sup> (FS2), and 30 km<sup>2</sup> (FS3).

**2.2.3. Disaggregation based on region (DR)**

The number of deer per county was summed to the regional level (nine county total) and disaggregated as described for the DC method, but without regard for county boundaries so that demarcation lines that appeared in the DC method were eliminated.

#### **2.2.4. Disaggregation based on reporting unit (DRU)**

Counts of deer per reporting unit were disaggregated using the *DC* method, except the disaggregation was done at the reporting unit scale rather than county scale.

#### **2.2.5. Kriging and deer redistribution based on land cover (K1-K4)**

The estimated counts for each county were divided by the total area in the county, to derive average densities per km<sup>2</sup>. The centroids were calculated for each county and the appropriate density value was joined to the centroid to prepare the data for kriging. An ordinary kriging model (*K1*) was used to create a raster surface of deer density.

The kriged surface was subsequently used to redistribute densities of deer based on land cover type using the NLCD data. The land cover data was reclassified to the following multipliers: shrublands were classified as 1.2, forests were classified as 1.0, and grasslands were classified as 0.8; all other land cover categories (including cropland, residential, and water) were classified as 0. The reclassified land cover raster was multiplied by the kriged surface. This resulted in a raster of deer density that was proportionally distributed within shrubland, forest, and grassland land covers (*K2*).

The redistribution of deer density resulted in areas of higher or lower density than that reported for ecoregions by TPWD. A correction factor was computed for each ecoregion using the ratio of the zonal sum of deer density for each ecoregion obtained from the raster described above to the deer density for each ecoregion (*K3*) reported by TPWD. Finally, to reduce the smoothness of the surface generated by kriging, the estimated deer from kriging were re-distributed based on the proportion of suitable land use within each county (*K4*).

#### **2.2.6. Disaggregation to farm-boundary (DFB)**

County-level deer population estimates (*DC*) were summed to farm boundaries and a centroid for each farm was used to represent the deer population. Farm boundaries were provided by the United States Department of Agriculture (USDA) Farm Services Agency (FSA) of Texas.

#### **2.2.7. Linear scaling based on NDVI (LS1-LS4)**

Maximum monthly composite NDVI values for the months of April (*LS1*) and December (*LS2*)

2004 were used to linearly scale deer density in the range 0–30 for each 1 km<sup>2</sup> pixel. April and December were selected because they represent the highest and lowest monthly precipitation, respectively. The estimated densities were summed and compared to the count of deer provided by TPWD. The initial scaling resulted in an overestimate in the number of deer in some areas and an underestimate in others. The study site was subsequently split by ecoregion and NDVI was used to linearly scale density in the range 0–45 deer per km<sup>2</sup> in the Edwards Plateau, and 0–15 deer per km<sup>2</sup> in the South Texas Brush ecoregions for both April (*LS3*) and December (*LS4*).

#### **2.2.8. Epidemic simulation model**

The same modeling scenario was used for all model comparisons: one cell was selected as infected to initiate the simulation and this cell was used as the starting point for all simulations. For every simulation of the model, each cell centroid was allowed to interact with other cells within a 2 000 m neighborhood. The model was simulated for a time period representing 100 days and 100 model runs were simulated for each dataset, yielding a total of 10 000 iterations. The median number of deer infected and median area affected (km<sup>2</sup>) were used to characterize each set of simulations at the 100th model day.

The population density, distribution, and habitat requirements of deer within the study area were explicitly incorporated in the model. We assumed the home ranges of deer in the study area were within a distance of 2 km and no interactions took place beyond this distance. The interaction probabilities between locations were weighted using a kernel defined by the inverse of the distance from the cell centroid, with the value being a fraction of a pre-specified bandwidth (1 000 m). The weights were reduced when neighbors were further away than the pre-specified bandwidth, and increased when they were closer.

In the model, deer could pass through four disease states: from susceptible to latent, from latent to infectious, from infectious to immune, and finally back to susceptible. Parameter values for the latent, infectious, and immune periods were based on the literature, predominantly lab based studies of FMD infection in deer<sup>1</sup> [14, 16, 37]. These transitions partially determined the dissemination rate of FMD between cells [15]. The first transition depended on contact rates between susceptible and infected deer cells in the previous time step. Homogenous mixing



**Table II.** Baseline model parameter values.

Parameter	Value
Latency, days (min, max)	3–5
Duration of infectiousness, days (min, max)	3–14
Duration of resistance to re-infection, days (min, max)	90–180
Maximum number of neighboring cells with which each infected cell can interact	8
Maximum distance of neighboring cells within which each infected cell can interact (m)	2 000
Density scaling parameters (min, max)	0–30

was assumed to take place within but not between cells.

The probability of interaction between neighboring locations also depended on the density of susceptible deer in the two locations, calculated as the product of their probabilities. Locations with more than a maximum threshold of deer were assigned a probability of 1.0. The remaining locations were linearly scaled into the interval 0 to 1 by dividing each location’s density by the maximum threshold value [37]. To incorporate stochasticity into the model, interactions between a susceptible location and an infectious neighbor occurred when a random number from a pseudo-random number generator (PRNG) using the Mersenne Twister mt19937 algorithm [26,35] was below the assigned probability threshold for that pair of locations [37].

Once a cell was infectious the second, third, and fourth transitions in the model depended on the length of the latent, infectious, and immune periods as assigned in the model parameterization<sup>1</sup> [14, 16,37]. The specific values were assigned randomly within the corresponding parameter ranges using a uniform distribution. The baseline model parameter values are summarized in Table II.

The geographic automata model framework is particularly suited to modeling foreign animal diseases in wild animal populations. Geographic variations are explicitly modeled in a simple manner and individual-level animal census data is not required, as long as an approximate statistical distribution is available [18]. In addition, the model does not require complex mathematical equations, but instead relies on local relationships between cells [18]. The assumption of local spread is a reasonable assumption for white-tailed deer populations: in the absence of disturbance, deer are unlikely to move outside their local home range [37].

2.3. Data analysis

The results of the model simulations were analyzed using agglomerative clustering to identify

groups of geostatistical methods that produce similar results at the 100th model day. This was done using the agnes algorithm within the Cluster package in R [20,30].

The similarity matrix used in the clustering analysis was developed by calculating the difference between each pair of geostatistical methods as the square root of the average squared difference between each pair of model runs:

$$similarity = \sqrt{\frac{1}{n^2} \sum_{i=1..n} \sum_{j=1..n} (m_{1i} - m_{2j})^2}$$

where  $n$  is the number of model runs (100 in this case),  $m_{1i}$  is the  $i$ th value for methodology 1, and  $m_{1j}$  is the  $j$ th value for methodology 2.

3. RESULTS

3.1. Predicted distribution and density of deer for each individual method

Each of the estimation methods were compared with respect to the predicted number of deer per county as provided by TPWD. The number of deer in the study region predicted from each estimation method is summarized in Table III. Estimates ranged from 385 939 to 768 493. In addition, the geostatistical methods resulted in different spatial representations of the distribution of deer in the study region. Examples of some of the distributions and the corresponding predicted spatial epidemic spread are shown in Figures 2 and 3, respectively. Figures for all of the spatial representations and predicted spatial epidemic spread are available online (Figs. A and B) at [www.vetres.org](http://www.vetres.org). Summary statistics for the predicted number of deer infected and area affected for each of the estimation methods (10 000 iterations each) is shown in Table IV. A graphic depiction of the median number

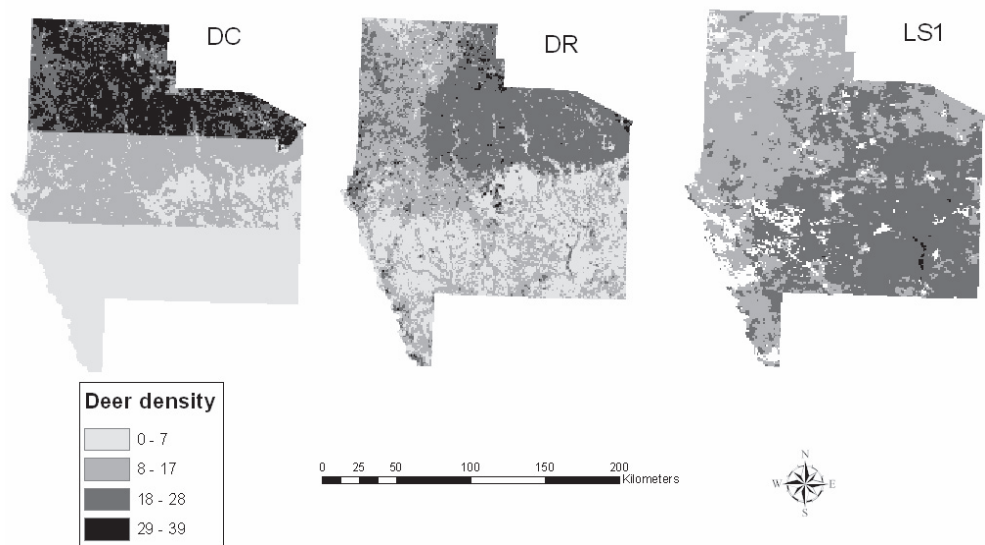


Figure 2. Example surfaces (*DC*, *DR*, and *LSI*) from geostatistical estimates of deer distribution and density.

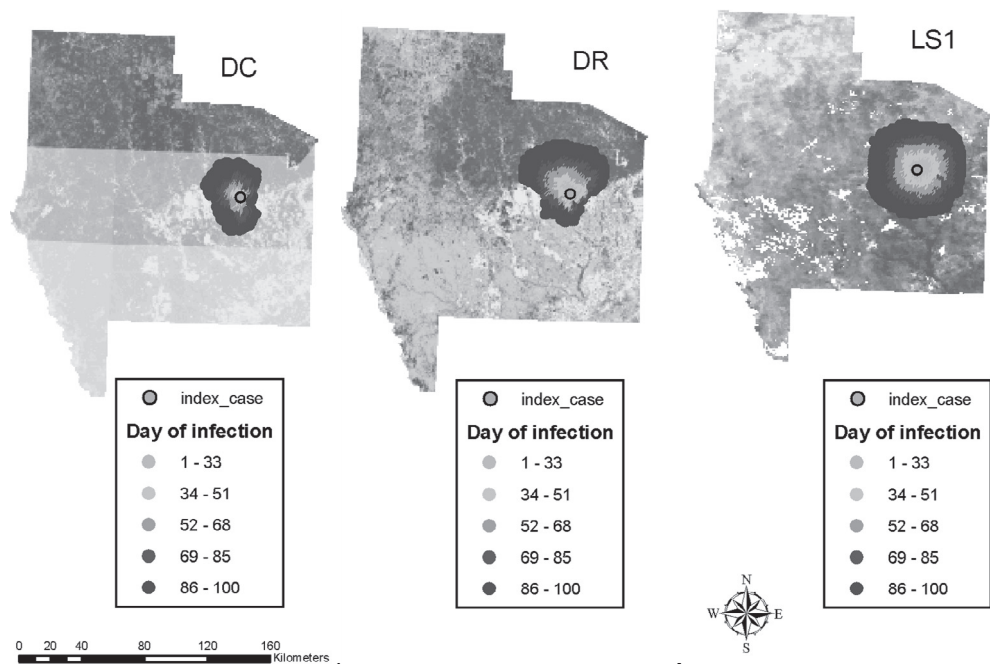


Figure 3. Epidemic progression (*DC*, *DR*, and *LSI*).



**Table III.** Predicted number of deer by estimation methodology as compared to Texas Parks and Wildlife Department (TPWD) provided county estimate.

Method	Number of deer	
	Estimated	TPWD ( <i>n</i> = 427 292) % estimated
<i>DC</i>	412 770	97
<i>FS1</i>	395 281	93
<i>FS2</i>	395 828	93
<i>FS3</i>	396 269	93
<i>DR</i>	410 624	96
<i>DRU</i>	768 493	179
<i>K1</i>	403 027	94
<i>K2</i>	439 169	103
<i>K3</i>	536 042	125
<i>K4</i>	385 939	90
<i>DFB</i>	385 939	90
<i>LS1</i>	491 566	115
<i>LS2</i>	415 426	97
<i>LS3</i>	470 428	110
<i>LS4</i>	464 680	109

of deer infected and the median area affected by method is shown in Figure 4. Boxplots of the predicted outbreak distribution for each method are shown in Figure 5.

3.2. Cluster Analysis

The clustering algorithm for the predicted distribution of deer infected identified two distinct clusters of methods, and one outlier (Fig. 6). The first cluster included the *DC* and *DFB* methods, both of which were methods constrained within county boundaries. The second cluster included all other methods except *LS1*: the difference between this and other methods is apparent in Figure 2.

4. DISCUSSION

We found substantial differences in the estimated number of deer in the study region based on the geostatistical estimation procedure used: the total deer population ranged from 385 939 to 768 493. Substantial differences were also observed in the median predicted magnitude of the outbreak, which ranged from 1 563 to 8 896 deer infected. This

variability in the predicted median outbreak size, as a result of using different geostatistical methods to describe the population at-risk, supports the argument that reporting only summary statistics from simulation models can be misleading. It is important that an attempt be made to consider the entire predicted outbreak distribution when summarizing modeling results, especially if these results are to be presented to policy-makers or to be used by decision-makers in the face of an outbreak.

Species predicted spatial distributions should be compared to known data, where available, and consideration should be given to how well the spatial estimate matches the species distribution reported from reputable data sources. Reports available from TPWD for this study region (for example, *DC* and *DR*, Fig. 2) indicate that the highest density of deer is found in the northern half of the study region [25]. Some of the spatial estimation methods better reflected this distribution than others. This is highlighted in Figure 2, where the predicted deer distribution for three of the methods is shown. *DC* predicts the highest density in the uppermost three counties, but fails to extend far enough south. *DR* appears to best represent the known density of deer in the study region, with the highest density in the northern half of the study region. *LS1* is clearly inconsistent with the deer estimate from TPWD, showing the highest density in the southwest portion of this region. Even though some of the spatial distributions do not accurately reflect the TPWD estimate, using these geostatistical methods still resulted in similar estimates of the total disease outbreak impact. Thus, the choice of geostatistical method for representing animal species distribution is probably secondary to the objectives of the study. If the aim is to estimate the overall impact of an FMD disease outbreak, results from this study suggest that the choice of geostatistical method is not critical. However, from the perspective of spatial analysis and predicting the likely spatial distribution of infected areas, the choice of geostatistical method becomes more important.

Study results indicate that the simpler methods for predicting white-tailed deer density

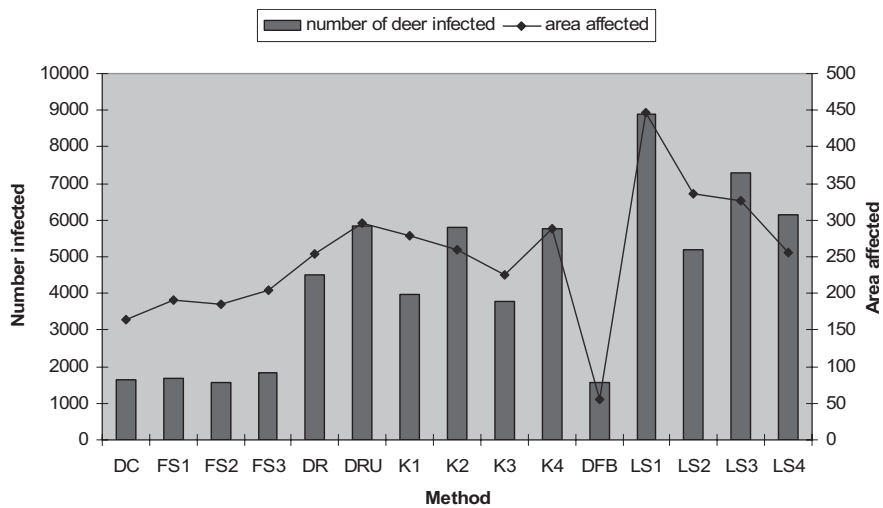


Figure 4. Median number of deer infected and median area affected in km² by estimation methodology.

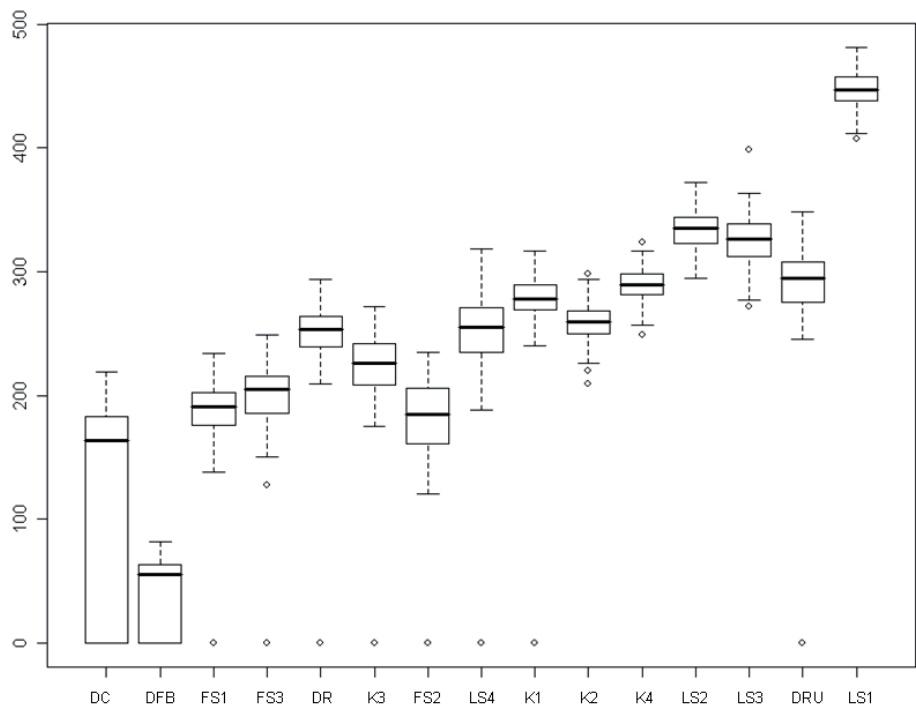
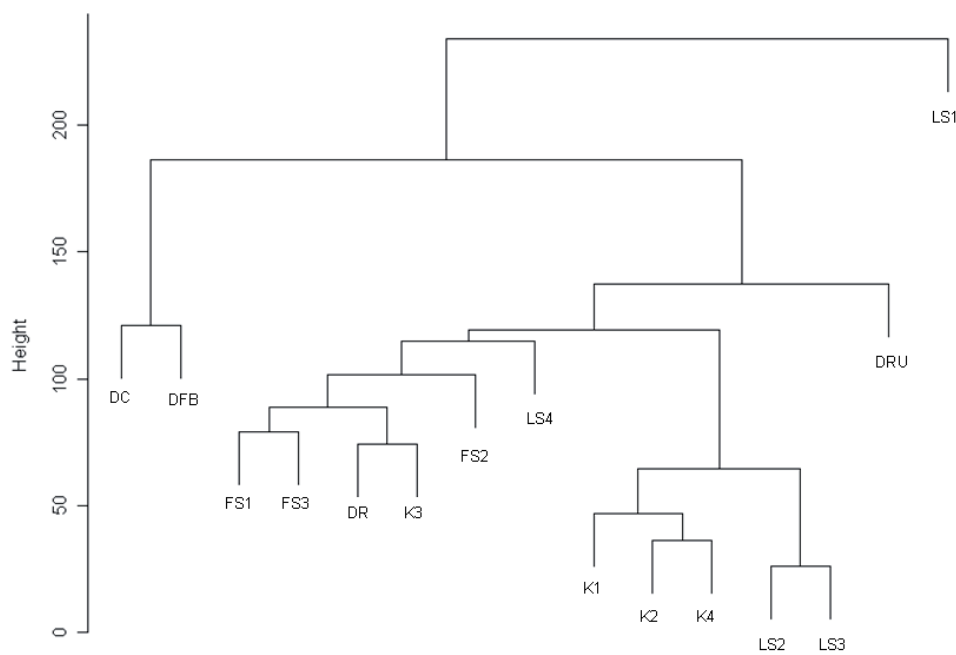


Figure 5. Boxplots of the predicted outbreak distribution for each estimation methodology.



**Figure 6.** Resulting clusters predicted by hierarchical agglomerative clustering algorithm for the distribution of deer infected.

and distribution form one cluster, and the more complex methods form a second cluster (Fig. 5). For these two clusters of methods, the mean predicted median outbreak size and median area affected were 1 599 deer and 110 km<sup>2</sup> and 4 578 deer and 263 km<sup>2</sup>, respectively. Thus, broadly these two clusters resulted in a two to three-fold difference in predicted outbreak size and distribution, indicating that the estimation methodology used to distribute deer has a significant impact on model predictions. Specifically in the present study, two representations of the population are necessary to capture the variability in the predicted outbreak size and distribution. The values in the dendrogram (Fig. 6) are measures of the variation in the number of predicted infected locations. Cluster 2 values span a range of approximately 140. This characterizes the variability that one might expect in model predictions, depending on the method used to characterize the spatial distribution of an animal species of interest. Certain methods are closer together than others; for example K1, K2, K4, LS2, and LS4 are separated by a dis-

tance of only 70. The distance cut point used for cluster identification is somewhat arbitrary. For example, cluster 2 could be further subdivided if the variation of 140 is considered too large. This would result in the *DRU* method – a method that produced a distribution that was inconsistent with the TPWD report of deer distribution (highest densities in the southern portion of the study region) – being an outlier. The underlying assumptions of the various geostatistical estimation methods should be considered, in addition to how well they predict the known data. Dasymetric mapping assumes that the data (wildlife density and distribution over the landscape) has an underlying spatial pattern which can be characterized using ancillary data, such as habitat and carrying capacity. This assumption is reasonable for wildlife data. Dasymetric mapping methods (such as *DC*, *DFB*, and *DRU*) that were applied within a political (county) or ecological (reporting unit) boundary further assumed that these ancillary attributes were captured within the bounds placed on the data. For these methods, it was assumed that habitat and carrying

capacity attributes were adequately captured at the sub-county or reporting unit level. This assumption may not be entirely valid. We know that wildlife, especially white-tailed deer, view and select habitat at the patch level and that patch dynamics do not necessarily follow political boundaries, that is, a suitable patch could easily cross county boundaries leading to an inaccurate estimate of the density or distribution. Based on the deer distribution data available from TPWD, the reporting unit scale appears too coarse to adequately model deer distribution using the ancillary attributes of habitat and carrying capacity. In addition, methods that were forced to distribute within county boundaries lead to demarcation lines (horizontal bands) along these boundaries in the resulting spatial estimate of the species distribution. Although we know these boundaries are not realistic, they often represent the source of the only census data available.

Kriging has an underlying assumption of spatially continuous data. Wildlife distributions are typically not considered continuous and therefore kriging may not be the most appropriate method to use. Estimation methods that depended on remotely sensed imagery (*LS1–LS4*) assume there is a relationship between NDVI values and deer density and distribution. NDVI measures vegetative greenness, and it was assumed that there is a linear relationship between deer density and NDVI value. However, this assumption has not been validated. The distribution derived from NDVI did not always adequately characterize the known deer distribution. For example, the outbreak distribution produced by the *LS1* method (Fig. 5) was an outlier in the cluster analysis. How this method might be applied requires further investigation.

Demarcation lines in the data are an artifact of the artificial (administrative and political) boundary that is placed on the data (county or reporting unit). Dasymetric mapping methods create a statistical deer distribution designed to remove the effect of artificial boundaries [6,24,29]. However, when these methods are used within a boundary the resulting distribution tends to suffer from demarcation at that boundary. Methods that do not explicitly include

boundaries in the estimation procedure (*DR*, *K1*, *LS2*) did not suffer from demarcation in the resulting estimated distribution. Because we are interested in modeling disease spread in a multi-county region, a clear demarcation line (for example, high density to zero or extremely low density across a single 1 km<sup>2</sup> pixel) presents a problem for epidemic modeling. Given that geographic automata models operate at the local level, the distribution and density of surrounding cells is very important in determining whether and how the disease will spread. An ideal method is one that results in both realistic and suitable spatial animal distributions for spatial modeling of disease spread.

The results of this study demonstrate that to eliminate demarcation in the spatial distribution data at the level of aggregation available it is necessary to use a regionalized interpolation method (kriging) or a method involving individual pixel level data (NDVI). However, these methods that smooth population distributions result in much larger estimates of the magnitude of the outbreak and the spatial distribution of infection (Fig. 3). The actual estimate of the overall population density appears to play little, if any, role in the resulting magnitude of the predicted outbreak. The distribution and, more specifically, the smoothness and spatial continuity of the distribution appear to have a major role in the predicted outbreak size (Fig. 3). This is highlighted in Figures 3, 5, 6, and Table IV. The group with low spatial continuity: *DC*, *FS1–FS3*, *DR*, *DRU*, *DFB*, and *LS4* all have zeros in their distributions where outbreaks failed to occur. Their fifth percentiles (Tab. IV) are all zero, indicating that outbreaks did not occur in at least 5% of the model runs. All of the surfaces with low spatial continuity had zeros in their fifth percentiles for 98 of the 100 runs indicating that in only 2 of the 100 runs did an outbreak start. The group with high spatial continuity: *K1–K4* and *LS1–LS3* all have much higher values for their fifth percentiles indicating that an epidemic always occurred for these surfaces. This result is to be expected given the spatial formulation of the model. The more continuity in the spatial distribution, the greater the opportunity for

**Table IV.** Predicted size of an outbreak of foot-and-mouth disease in a population of deer in southern Texas for each estimation method. Results shown are from 100 simulations of the geographic automata model for each deer surface.

Method	Number of deer					Area (km <sup>2</sup> )				
	Median	IQR	5%, 95%	Skewness	Kurtosis	Median	IQR	5%, 95%	Skewness	Kurtosis
<i>DC</i>	1 626	1 830	0, 2 051	−0.847	−1.095	164	182	0, 205	−0.849	−1.095
<i>FS1</i>	1 661	263	0, 1 935	−2.59	6.438	191	27	0, 219	−0.271 1	−6.907
<i>FS2</i>	1 563	439	0, 1 931	−1.512	1.011	185	45	0, 226	−1.601	−1.165
<i>FS3</i>	1 839	344	0, 2 155	−2.198	4.4	205	30	0, 229	−2.505	5.538
<i>DR</i>	4 510	429	0, 4 991	−3.065	8.415	253	25	0, 281	−3.072	8.426
<i>DRU</i>	5 829	724	0, 6 570	−1.59	0.719	295	31	0, 326	−1.61	0.745
<i>K1</i>	3 978	297	3 515, 4 373	−5.265	32.385	278	20	248, 302	−5.574	34.95
<i>K2</i>	5 798	492	5 110, 6 331	−0.678	1.625	259	17	234, 286	−0.098	0.792
<i>K3</i>	3 791	689	3 028, 4 416	−2.894	17.993	226	33	188, 255	−3.938	27.66
<i>K4</i>	5 752	562	5 225, 6 351	0.059	−0.773	289	17	266, 311	−0.16	−0.09
<i>DFB</i>	1 572	1 899	0, 2 340	−0.436	−1.329	56	63	0, 76	−0.551	−1.399
<i>LS1</i>	8 896	374	8 493, 9 381	−0.075	0.173	447	19	427, 471	−0.088	0.198
<i>LS2</i>	5 190	349	4 762, 5 443	−0.227	−0.39	335	21	303, 353	−0.317	0.026
<i>LS3</i>	7 304	589	6 639, 8 039	−0.267	1.088	326	27	291, 354	−0.047	1.107
<i>LS4</i>	6 149	1 231	0, 7 080	−1.773	1.701	255	38	0, 294	−1.862	1.922

interactions between locations and therefore more interactions will occur even when there are lower interaction probabilities.

The need to use spatially-explicit models to simulate the spread of FMD has been recognized [15, 19], and spatial heterogeneity has been identified as possibly the greatest challenge to realistically representing FMD spread through a landscape [12]. In addition to capturing the spatial heterogeneity of the population across the landscape, wildlife distributions need to be seasonally-dynamic, since these species are particularly affected by variations in climate and natural resources. Such temporal dependency may have a significant impact on the spread of disease within wildlife populations, and further, into domesticated animal populations of interest [7]. Temporal dependency should be incorporated in future studies of disease spread in potential wildlife reservoirs. While we have included some level of temporal dependency with the NDVI surfaces, a more detailed analysis is necessary in the future.

The model used in this study has been used previously to investigate wildlife-domestic species interactions between feral pigs and cattle [7, 37] and between wild deer and cattle [37]. In the current study, our focus was on

the potential spread of FMD in wild deer populations. We made the simplifying assumption that because of relatively low grazing densities in this extensively managed livestock system, cattle do not contribute greatly to disease spread. Also, we focused on the initial stages of disease spread ( $\leq 100$  days); assuming a minor role for domestic livestock during this initial phase of an outbreak is likely to be valid. The duration of resistance to FMD virus reinfection was assumed to be 90–180 days. Although this assumption may be unrealistically low, it probably had little impact on study results because of our focus on the initial stages of disease spread. The model predictions are likely to be sensitive to temporal fluctuations in the population densities (for example, seasonal or annual population trends, particularly if these fluctuations occur differentially across geographical areas of the study region). Thus, study results should be viewed as the average effect of different representations of animal densities. More research is needed to determine if the methods of representing animal densities, or temporal fluctuation of those densities, are more important in determining the outcome of a disease incursion such as FMD. Caution should be exercised when using the same epidemiological parameters on different

spatial landscapes. This is even more problematic when epidemiologic parameters are estimated from a disease outbreak that occurs within a given spatial landscape. Given that FMD has not occurred in the USA since 1929, it is virtually impossible to estimate the epidemiologic parameters, should FMD virus be introduced into the deer population. However, the model system does incorporate uncertainty by using parameter ranges [37]. Regarding the role of spatial heterogeneity on parameter estimation, we feel that the model is robust even in the absence of detailed parameter estimates. Spatial heterogeneity has been implicitly included in the model by the use of density to adjust disease transmission. Furthermore, by using landscape variability (key habitat features) in the distribution methodologies and density to control interaction in the simulation model, we have incorporated heterogeneity of transmission via a 'self-adjusting' model that varies across the landscape. We have captured variation in both the distribution of susceptible hosts and contact rates over the landscape: this is the primary underlying cause of the differences between model results.

To our knowledge, this is the first study to define the range and distribution of estimates of outbreak magnitude generated by various methods of processing aggregate data of white-tailed deer. Depending upon the geostatistical methods used, we found high variability in the predicted spatial distribution of white-tailed deer and in the predicted magnitude and size of the outbreak. At least two representations of the population may be necessary to adequately capture this variability in the predicted outbreak distributions, and certain geostatistical estimation methods may perform better than others from a spatial analysis perspective. Modelers must be aware of the underlying assumptions of the various estimation methods that might be used, and the potential impact of these on model predictions. In addition, the goal of estimation needs to be considered – whether it is solely to represent the spatial distribution of an animal species of interest, or to provide data for a simulation modeling exercise. It is also of critical importance to consider available data, meth-

ods of distribution estimation and assumptions prior to developing modeling scenarios and generating epidemic predictions. The importance of critically considering these factors cannot be overstated, especially if models are used to inform policy and epidemic response. How animal populations are represented needs to be considered in all spatial disease spread models.

**Acknowledgements.** Partial funding for this study was provided by the Foreign Animal and Zoonotic Disease Defense Center (FAZD), a department of Homeland Security National Center of Excellence at Texas A&M University. We would like to acknowledge the efforts of our collaborators at the FAZD Center for their contributions to this manuscript (Dr. Raghavan Srinivasan, Dr. Jay Angerer, and Jennifer Jacobs). This study represents research by the senior author in partial fulfillment for the degree of Ph.D. in Biomedical Sciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University.

## REFERENCES

- [1] Ahmed E., Agiza H.N., On modeling epidemics, including latency, incubation and variable susceptibility, *Physica A* (1998) 253:347–352.
- [2] Benyoussef A., Boccara N., Chakib H., Ez-Zahraouy H., Lattice three-species models of the spatial spread of rabies among foxes, *Int. J. Mod. Phys. C* (1999) 10:1025–1038.
- [3] Buckland S.T., Elston D.A., Empirical models for the spatial distribution of wildlife, *J. Appl. Ecol.* (1993) 30:478–495.
- [4] Crombie M.K., Gillies R.R., Arvidson R.E., Brookmeyer P., Weil G.J., Sultan M., Harb A.M., An application of remotely derived climatological fields for risk assessment of vector-borne disease: a spatial study of Filariasis prevalence in the Nile delta, Egypt, *Photogramm. Eng. Remote Sens.* (1999) 65:1401–1409.
- [5] Davies G., Foot and mouth disease, *Res. Vet. Sci.* (2002) 73:195–199.
- [6] DeMers M., *Fundamentals of geographical information systems*, John Wiley and Sons, New York, USA, 2000.
- [7] Doran R.J., Laffan S.W., Simulating the spatial dynamics of foot and mouth disease outbreaks in feral pigs and livestock in Queensland, Australia, using a susceptible-infected-recovered cellular automata model, *Prev. Vet. Med.* (2005) 70:133–152.
- [8] Durand B., Mahul O., An extended state-transition model for foot and mouth disease epidemics in France, *Prev. Vet. Med.* (2000) 47:121–139.
- [9] Duryea M., Caraco T., Gardner G., Maniatty W., Szymanski B., Population dispersion and equilibrium infection frequency in a spatial epidemic, *Physica A* (1999) 132:511–519.



- [10] Edelstein-Keshet L., Mathematical models in biology, Random House, New York, USA, 1988.
- [11] Elbers A.R.W., Dekker A., Dekkers L.J.M., Serosurveillance of wild deer and wild boar after the epidemic of foot-and-mouth disease in the Netherlands in 2001, *Vet. Rec.* (2003) 153:678–681.
- [12] Ferguson N.M., Donnelly C.A., Anderson R.M., The foot and mouth epidemic in Great Britain: pattern of spread and impact of interventions, *Science* (2001) 292:1155–1160.
- [13] Filipe J.A.N., Gibson G.J., Studying and approximating spatio-temporal models for epidemic spread and control, *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* (1998) 353:2153–2162.
- [14] Forman A. J., Gibbs E. P. J., Studies with foot-and-mouth disease virus in British deer (red, fallow and roe), *J. Comp. Pathol.* (1974) 84:215–220.
- [15] Garner M.G., Lack M., An evaluation of alternate control strategies for foot and mouth disease in Australia: a regional approach, *Prev. Vet. Med.* (1995) 23:9–32.
- [16] Gibbs E.P.J., Herniman K.A.J., Lawman M.J.P., Sellers R.F., Foot-and-mouth disease in British deer: transmission of virus to cattle, sheep and deer, *Vet. Rec.* (1975) 96:558–563.
- [17] Johansen A., A simple model of recurrent epidemics, *J. Theor. Biol.* (1996) 178:45–51.
- [18] Kao R.R., The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK, *Trends Microbiol.* (2002) 10:279–286.
- [19] Kao R. R., The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease, *Proc. Biol. Sci.* (2003) 270:2557–2564.
- [20] Kaufman L., Rousseeuw P.J., Finding groups in data: an introduction to cluster analysis, Wiley, New-York, USA, 1999.
- [21] Keane C., The epizootic of foot-and-mouth disease in California, *Calif. Dept. Agric., Special Pub.* 65, 1926, p. 54.
- [22] Kitron U., Kazmierczak J., Spatial analysis of the distribution of Lyme disease in Wisconsin, *Am. J. Epidemiol.* (1997) 145:558–566.
- [23] Kleczkowski A., Grenfell B.T., Mean-field-type equations for spread of epidemics: the ‘small world’ model, *Physica A* (1999) 274:355–360.
- [24] Langford M., Unwin D.J., Generating and mapping population density surfaces within a geographical information system, *Cartogr. J.* (1994) 31:21–26.
- [25] Lockwood M., White-tailed deer population trends. Performance report as required by federal aid in wildlife restoration act, W-127-R-13, 2005, Texas Parks and Wildlife, Austin, TX, USA, pp. 1–31.
- [26] Matsumoto M., Nishimura T., A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Trans. Model. Comput. Simul.* (1998) 8:3–30.
- [27] Mennis J.; Generating surface models of population using dasymetric mapping, *Prof. Geogr.* (2003) 55:31–42.
- [28] Pinto A.A., Foot-and-mouth disease in tropical wildlife, *Ann. NY Acad. Sci.* (2004) 1026:65–72.
- [29] Poulsen E., Kennedy L.W., Using dasymetric mapping for spatially aggregated crime data, *J. Quant. Crim.* (2004) 20:243–262.
- [30] R Development Core Team, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2005.
- [31] Rossi R.E., Dungan J.L., Beck L.R., Kriging in the shadows: geostatistical interpolation for remote sensing, *Remote Sens. Environ.* (1994) 49:32–40.
- [32] Rousseau G., Giorgini B., Livi R., Chante H., Dynamical phases in a cellular automata model for epidemic propagation, *Physica A* (1997) 103:554–563.
- [33] Sirakoulis G.C., Karafyllidis I., Thanailakis A., A cellular automaton model for the effects of population movement and vaccination on epidemic propagation, *Ecol. Model.* (2000) 133:209–223.
- [34] Suttmoller P., Barteling S.S., Olascoaga R.C., Sumption K.J., Control and eradication of foot-and-mouth disease, *Virus Res.* (2003) 91:101–144.
- [35] Van Neil K., Laffan S.W., Gambling with randomness: the use of pseudo-random number generators in GIS, *Int. J. Geogr. Inf. Sci.* (2003) 17:49–68.
- [36] Vlad M.O., Schonfisch B., Lacoursiere C., Statistical-mechanical analogies for space-dependent epidemics, *Physica A* (1996) 229:365–401.
- [37] Ward M.P., Laffan S.P., Highfield L.D., The potential role of wild and feral animals as reservoirs of foot-and-mouth disease, *Prev. Vet. Med.* (2007) 80:9–23.
- [38] Yang P.C., Chu R.M., Chung W.B., Sung H.T., Epidemiological characteristics and financial costs of the 1997 foot-and-mouth disease epidemic in Taiwan, *Vet. Rec.* (1997) 145:731–734.