Veterinaria Italiana, **43** (3), 483-489

Geographic information system-based avian influenza surveillance systems for village poultry in Romania

Michael P. Ward

Summary

The analysis of surveillance data facilitates the planning, implementation and evaluation of disease control programmes. Geographic information systems (GIS) have several functions, including input (database functions), analysis (interpolation, cluster detection, identification of spatial risk factors) and output (sampling design, disease risk maps). This paper focuses on visualisation techniques that enable improved design and evaluation of surveillance data. Data generated within a pilot GIS-based surveillance programme for avian influenza in village poultry in the Romanian county of Tulcea is used as an example. The use of kriging helped highlight areas in the country where sampling potentially was sub-optimal, and error maps demonstrated the level of confidence that can be placed in serological surveillance results in different localities. Disease surveillance systems traditionally have not focused on the issues of disease risk and sample size visualisation. Standards need to be developed on how sampling and disease data generated within animal health surveillance systems are analysed and presented. This is particularly important for transboundary diseases such as avian influenza.

Keywords

Avian influenza, Epidemiology, Geographic information system, Romania, Surveillance.

Il sistema di sorveglianza basato su un sistema informativo geografico applicato all'influenza aviaria negli allevamenti avicoli pubblici in Romania

Riassunto

L'analisi dei dati di sorveglianza facilita la progettazione, l'implementazione e la valutazione dei programmi di controllo delle malattie. I sistemi informativi geografici (GIS) svolgono numerose funzioni che vanno dall'inserimento (funzioni dei database), all'analisi (interpolazione, riconoscimento di cluster, identificazione di fattori spaziali di rischio), all'output (disegno di piani di campionamento, costruzione di mappe di rischio). Questo lavoro focalizza l'attenzione sulle tecniche di visualizzazione che permettono di perfezionare i modelli e la valutazione dei dati provenienti dalla sorveglianza. Vengono portati come esempio i dati generati da un programma pilota di sorveglianza basato su GIS in un allevamento avicolo pubblico localizzato nella contea rumena di Tulcea. Mediante l'utilizzo del "kriging" è stato possibile evidenziare le aree del paese dove il campionamento risultava potenzialmente al di sotto dell'optimum e le mappe degli errori hanno evidenziato il livello di confidenza che poteva essere assegnato ai risultati di sorveglianza sierologia nelle diverse località. I sistemi di sorveglianza tradizionalmente non sono focalizzati sul rischio di insorgenza della patologia né sul disegno dell'opportuna dimensione del campionamento. E' necessario sviluppare degli

Texas A&M University College of Veterinary Medicine and Biomedical Sciences, MS 4458, College Station, TX 77843-4458, United States of America
mward@cvm.tamu.edu

standard di analisi e presentazione dei dati relativi alla sanità animale generati dai piani di sorveglianza e dalle osservazioni in caso di focolai malattie. Ciò è particolarmente importante per patologie transfrontaliere quali l'influenza aviaria.

Parole chiave
Epidemiologia, Influenza aviaria, Sistema informativo geografico, Sorveglianza, Romania.

Introduction

Disease surveillance has been defined as: 'the continued watchfulness over the *distribution and trends* of incidence through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant information and the regular dissemination of such data to all that need to know …' (7).

The analysis of surveillance data allows changes in the health status of populations over time and space to be detected, thus facilitating the planning, implementation and evaluation of disease control programmes. Surveillance and surveys enable disease control authorities to detect either the emergence of a new disease or an unusual increase (epidemic) in an endemic disease.

Geographic information systems (GIS) have several functions, including input (database functions), analysis (interpolation, cluster detection, identification of spatial risk factors) and output (sampling design, disease risk maps). These functions allow the design, implementation and assessment of surveillance systems.

Although GIS is a powerful tool for designing, assessing and implementing surveillance systems, many issues need to be considered prior to implementing a GIS-based surveillance system. The data that is collected in surveillance systems needs to be spatially accurate and timely. We need to consider the type of surveillance system used to collect the data (passive or active), location accuracy, spatial level of aggregation, the use of administrative units versus the aim of the system, edge effects and the modifiable unit area problem. Analytical issues include the detection of clusters versus trends and patterns, spatial versus temporal and spatio-temporal clusters, statistical power and multiple testing. Some of the issues to consider regarding GIS-based surveillance output include consistency, cartography and the choice of interpolation techniques.

To be successful, disease surveillance information must be disseminated regularly to all those that rely on this information to make decisions. It is important that data generated within such systems can be transformed into useful information that is interpretable by non-GIS experts – those who are likely to be responsible for making animal disease control decisions. Dot and choropleth maps are useful for visualising disease distributions. They are simple to construct, requiring little geostatistical expertise. However, it is easy to introduce bias into the interpretation of such maps. Isopleth maps, presenting smoothed estimates of disease risk, facilitate visualisation of latent risk by avoiding artificial administrative boundaries. To calculate disease risk, numerators (for example, clinical disease cases, serological test positive cases, cases of virus isolation) and denominators (number of animals at-risk for example in a herd, county, or zip code) are required. Thus, given the spatially discrete and irregular nature of animal health data derived from surveillance systems, interpolation methods are needed to produce such disease risk maps. Options include kernel density estimation, indirect distance weighted methods and kriging. An advantage of kriging is that predictions of disease risk are based on a parametric model of the empirical data (the semi-variogram). This process (variography) provides the analyst with a better understanding of the spatial structure of the disease data. More importantly, the use of more standardised and objective methods of analysing data derived from disease surveillance systems can provide confidence to the decision-maker when disease risk maps are interpreted.

This paper will concentrate on the analytical and output issues we face when designing and using GIS-based surveillance systems. In particular, methods of data analysis and data visualisation will be explored, through a

description of a pilot GIS-based surveillance system for avian influenza in village poultry in the Romanian county of Tulcea. Avian influenza has recently become an emerging issue for world health: the pathogenic H5N1 influenza strain circulating in Asia, Africa, the Middle East and Europe has caused numerous disease outbreaks in domestic poultry and wild bird populations, and threatens human health. As of 2 June 2007, 190 (61%) of 312 humans known to have been infected with H5N1 since 2003 and reported to the World Health Organization (WHO) have died in 12 countries in South-East Asia, China, the Middle East and Africa(13). There is a fear that H5N1 could become the next pandemic influenza strain. Avian influenza virus infection is endemic in a range of free-living bird species worldwide (1, 2, 8), particularly species associated with water (9). Waterfowl and shorebirds can be infected by all subtypes of type A influenza viruses with few or no symptoms (12). These species are probably responsible for the spread of viruses between regions (6). Research suggests that waterfowl and shorebirds maintain a separate reservoir of viral gene pools from which new virus subtypes emerge (11). In the northern hemisphere, influenza virus infection rates are highest during spring migration for shorebirds, whereas waterfowl infections peak in late summer and early autumn (6). Juvenile waterfowl are more susceptible to infection; when the birds are migrating south, a higher prevalence is expected than in the spring, when the juveniles have matured (5). Avian influenza outbreaks (both high and low pathogenic) in poultry are often assumed to occur from exposure to wild avian species.

Materials and methods

Data source

The data reported in this case study is part of a larger project aimed at assessing the effectiveness of the existing surveillance systems in Romania to detect foci of avian influenza virus transmission and at increasing the sensitivity of these existing surveillance systems. The surveillance site is Tulcea county, located in eastern Romania. It is bordered to the east by the Black Sea and to the north and west by the Danube River. The eastern part of the county consists of an extensive wetland system, part of the Danube River delta. It is a major breeding area and point of congregation for migratory birds on the Black Sea-Mediterranean fly path, which extends from West Africa to central Asia.

The first outbreak of H5N1 highly pathogenic avian influenza (HPAI) was detected in Tulcea county in early October 2005. Outbreaks of H5N1 were controlled by depopulation of poultry in affected villages, disinfection and surveillance of sentinel chickens in depopulated villages and serological surveillance in selected areas of the county.

Data analysis

Variography was used to investigate serological surveillance for avian influenza antibodies in Tulcea county between January and August 2006. Variography is the process of constructing a semi-variogram of empirical data (for example, discrete locations and estimates of disease risk at those locations) and modelling the resulting distribution. Thus, the spatial structure of the data can be described by a small number of parameters (in this case, the nugget, range and sill). The location of all villages (n=141) in Tulcea county were identified by longitude and latitude coordinates. The total number of domestic poultry from which sera were collected was calculated during the study period. All possible unique pairs of village locations (n=9 870) were formed. Distances and sample sizes were calculated and a semi-variogram was formed. Semi-variance is a measure of average dissimilarity between observations as a function their separation in distance and direction. A semi-variogram is a plot of the semi-variance of all pairs of locations at a series of defined distances (lags). For locations close to each other, values (for example, sample size) are expected to be similar and the semi-variance will be low (values will be highly correlated). As locations get farther apart, values are expected to become more dissimilar and thus the semi-variance increases. The rate of increase in semi-variance

as distance increases, and the distance at which locations can essentially be considered independent, characterised the spatial pattern of the event-of-interest. Estimating the parameters of the line of best fit of an empirical semi-variogram allows the distribution to be modelled and interpolated with techniques such as kriging. A range of lag numbers and lag spacings were chosen to produce a semi-variogram which could be described by one of a number of *a priori* models. Using a line-of-best-fit approach, the parameters of the selected model (exponential, spherical, Gaussian) were estimated. Variography was performed using the freeware program, Variowin 2.2 (Yvan Pannatier, www.springer-ny.com/supplements/ variowin.html). The semi-variogram parameters (nugget, range, sill) were used to produce an interpolated map of sample size in Tulcea county (Spatial Analyst: ArcGIS™ 9.0. Environmental Systems Research Institute [ESRI], Redlands, California) and an error (sample variance) map for this interpolated surface. These maps were overlaid on the location of villages to identify localities where surveillance sampling appeared sub-optimal to identify avian influenza virus antibodies.

Results

Between January and August 2006, sera were collected from a total of 12 172 domestic poultry species. No samples were collected from 35 villages. In the remaining villages, the number of samples collected ranged from 2 to 1 030 (Fig. 1). The median number of samples collected in these villages was 58 (interquartile range, 21-247). The number of samples collected in Tulcea villages did not show strong evidence of clustering (Moran's autocorrelation statistic 0.026, P = 0.012), but villages from which samples were collected were clustered (Cuzick and Edwards test Bonferroni P = 0.010), compared to those villages where sampling was not conducted during the period. Most samples were collected during March (21.2%), June (19.5%) and July (16.5%).

The semi-variogram of sampling intensity, using 15 lags and a lag size of 3.5 km, is shown in Figure 2. A Gaussian model best fit this

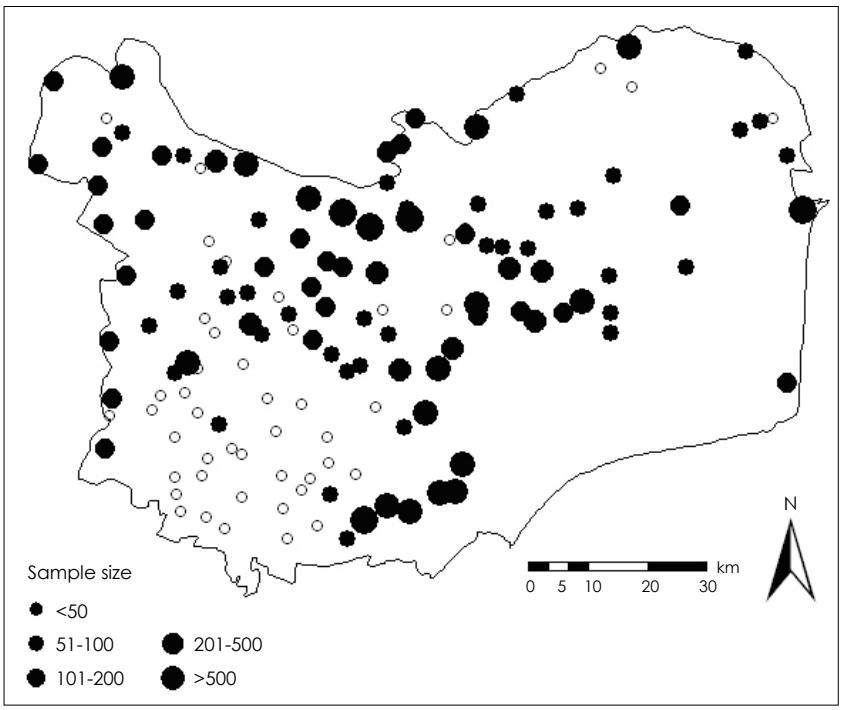

Figure 1
Number of poultry sampled for antibodies to avian influenza type A viruses in 141 villages in Tulcea county, Romania, between January and August, 2006
Open circles represent villages from which samples were not collected

semi-variogram. The estimated nugget, range and sill were 13.32, 13.25 and 23.40 km, respectively. Interpolated sampling intensity, using an ordinary kriging model, is shown in Figure 3 and the variance of predicted sampling intensity is presented in Figure 4. Areas in the south-west, south-east and north-east of Tulcea county were identified in which the variance of sampling was relatively high. Although few villages are located in the eastern part of the county, explaining the high sampling variance, villages are located in southwest Tulcea county. In this area, 28 (49%) of the 57 villages were not sampled, compared (P<0.001) to 35 (25%) villages that were not sampled from the entire county.

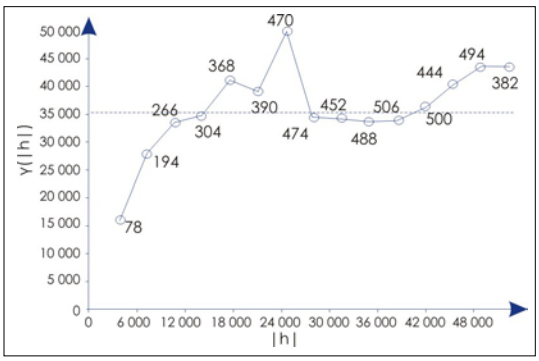

Figure 2
Semi-variogram of the number of poultry sampled for antibodies to avian influenza type A viruses in 141 villages in Tulcea county, Romania, between January and August, 2006 (15 lags, lag size 3 500 m [3.5 km])
The model of best fit (Gaussian, nugget 13.32, range 13.25, sill 23.40 km) is shown

Discussion

The aim of surveillance is to detect the occurrence of disease in time and space. Surveillance relies on repeated observations of the population-of-interest and a sufficient sample size to detect incursions of a foreign animal disease or changes in the incidence of an endemic disease. Analysis may or may not involve the use of spatial statistics, but whatever form of analysis is used, the design and implementation of sampling is critical to the success of the surveillance system.

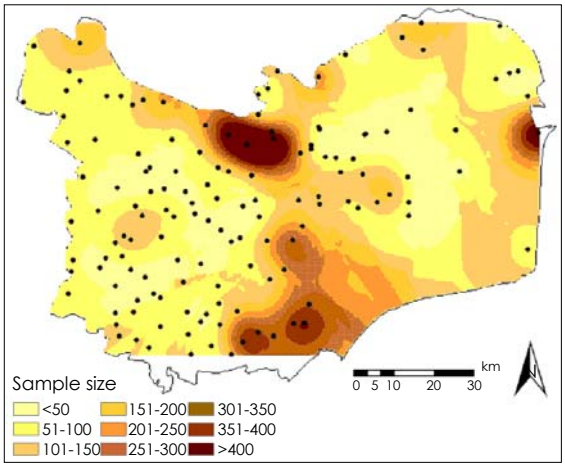

Figure 3
Interpolated number of poultry sampled for antibodies to avian influenza type A viruses in 141 villages in Tulcea county, Romania, between January and August, 2006, using an ordinary kriging model

Functions available with GIS packages can be useful and sometimes are essential in the design, implementation and assessment of surveillance systems, and the analysis of surveillance data. Some of these functions include buffering, overlay, zonal statistics and network analysis.

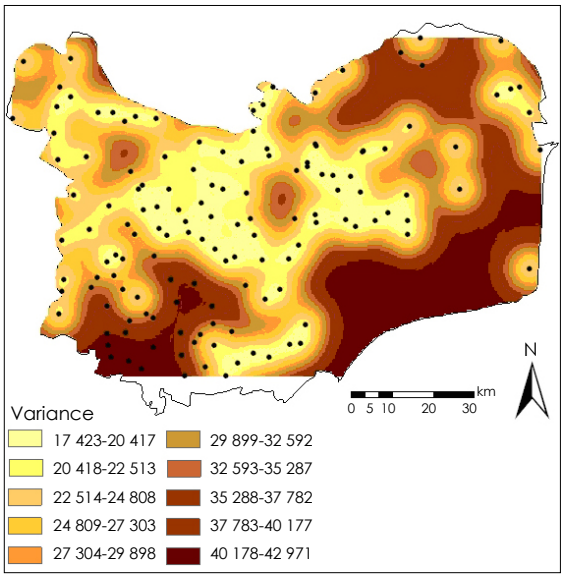

Figure 4
Interpolated variance of the number of poultry sampled for antibodies to avian influenza type A viruses in 141 villages in Tulcea county, Romania, between January and August, 2006, using an ordinary kriging model

An important tool that GIS offers is data visualisation. The sampling points used in a surveillance programme may simply be plotted as a point or polygon map. Production of maps in which points are drawn in proportion to the sample size (proportional symbol map) provides more information to those designing and implementing surveillance systems. Choropleth maps – in which polygons (usually administrative units) are shaded according to some disease occurrence value – explain spatial variation in disease distributions. However, this may prevent adequate visualisation and hypothesis generation and testing, and lead to inefficient resource allocation. Smoothing and interpolation methods reduce artificial effects of boundaries and facilitate identification of patterns by estimating disease occurrence at a given location using data from surrounding locations. A recent development in the assessment of surveillance programmes for livestock diseases is risk mapping. The need to represent disease distributions is driven by a desire to identify disease foci, visualise disease spread, identify risk factors and resource allocation, surveillance planning, disease control and prevention.

Isopleth maps – in which a continuous surface map of the disease or process under study is represented – aid visualisation of latent risk and address problems associated with choropleth mapping, specifically visual bias resulting from the uneven shape and size of administrative units in most countries and regions, whereby large areas may visually dominate maps. They are especially well suited to help the viewer see the geographical distribution of latent risks by overcoming visual bias and representing the continuous spatial variation in disease risk or in sampling intensity.

Although interpolation techniques have been used for some time to produce smoothed maps of disease risk (incidence or prevalence), these techniques have rarely been used to interpolate sampling schemes and assess the spatial sampling coverages within disease surveillance systems. For example, the use of kriging in the current study has helped highlight areas of intensive serological sampling for avian influenza viruses in Tulcea county from January to August 2006. More importantly, areas where sampling intensity was potentially sub-optimal are apparent when viewing such interpolated maps.

Valid modelling of the semi-variogram allows interpolation and disease risk mapping using kriging techniques. However, many decisions are involved in the computation of semi-variograms: it is not an automatic procedure, and should be considered exploratory data analysis. Researchers need to know their data and keep in mind the study objectives. Kriging was developed in 1971 by Matheron from an idea of Krige, and has been used extensively in the earth sciences (mining, meteorology, petrology, hydrology). More recently, it has been used to describe geographic variations in disease occurrence (3, 4, 10). Alternative methods of representing disease distributions include inverse distance-weighted methods, trend polynomial surfaces and splines. Limitations of these methods include the use of fixed constants and *a priori* assumptions that do not take advantage of the spatial structure of the disease distribution, and the inability to easily estimate the error of interpolation.

Kriging is a geostatistical method that allows optimal spatially continuous prediction of the latent risk surface from observed regional risk estimates. It is based on a parametric spatial model that may be specified by a spatial dependence function such as a semi-variogram. A weighted average is calculated from the whole data sample, where weights depend on the spatial dependence structure of the data. Regional disease risk estimates have greater influence on predictions the closer they are to the prediction sites.

There are two fundamental assumptions implicit in risk mapping, namely:

- the disease surface is continuous
- no local trends exist in the disease surface (the intrinsic hypothesis).

The assumption of stationarity – that spatial autocorrelation depends only on distance (isotrophy) and not distance and direction (anisotropy) – is often violated in epidemic disease investigations. Ignoring anisotrophy

may result in failure to adequately model spatial dependence and may invalid geostatistical predictions. However, in the case of ordinary kriging, Carrat and Valleron (4) suggest that the estimator is virtually unbiased for interpolation where the location to be estimated is surrounded by data on all sides and is within the range of influence of these data. The validity of kriging for epidemic and endemic disease visualisation should be investigated further.

An additional extension of the kriging functionality is error mapping. Using a kriging model, the variance of estimates can be calculated and displayed, so that regardless of the sample used to derive the estimate, the variance associated with the estimate can be visualised. In a disease surveillance system, areas with a high variance of sampling intensity need to be considered in the interpretation of surveillance data and the design or modification of future surveillance. As this area of visualisation is relatively new, standards need to be determined regarding how sampling data generated within animal health surveillance systems are analysed and presented. This is particularly important for transboundary diseases such as avian influenza.

Acknowledgements

Support for the acquisition and analysis of the data used in this study was provided by the Foreign Agricultural Service of the United States Department of Agriculture. The Romanian National Animal Health and Food Safety Authority and the Institute for Diagnostics and Animal Health, are thanked for providing access to outbreak data.

References

1. Alexander D.J. 2000. A review of avian influenza in different bird species. *Vet Microbiol*, **74**, 3-13.
2. Alexander D.J. 2001. Orthomyxoviridae-avian influenza. *In* Poultry diseases 5th Ed. (F. Jordan, M. Pattison, D. Alexander & T. Flanagan, eds). WB Saunders, London, 281-290.
3. Berke O. 2005. Exploratory spatial relative risk mapping. *Prev Vet Med*, **71**, 173-182.
4. Carrat F. & Valleron A.-J. 1992. Epidemiologic mapping using the 'kriging' method: application to an influenza-like illness epidemic in France. *Am J Epidemiol*, **135**, 1293-1300.
5. Delogu M., De Marco M.A., Donatelli I., Campitelli L. & Catelli E. 2003. Ecological aspects of influenza A virus circulation in wild birds of the western palearctic. *Vet Res Comm*, **27**, 101-106.
6. Krauss S., Walker D., Pryor S.P., Niles L., Chenghong L., Hinshaw V.S. & Webser R.G. 2004. Influenza A viruses of migrating wild aquatic birds in North America. *Vector-borne Zoon Dis*, **4**, 177-189.
7. Langmuir A.D. 1963. The surveillance of communicable diseases of national importance. *New Engl J Med*, **268**, 182-192.
8. Rosenberger J.K., Krauss W.C. & Slemmons R.D. 1974. Isolation of Newcastle disease and type-A influenza viruses from migratory waterfowl in the Atlantic flyway. *Avian Dis*, **18**, 610-613.
9. Stallnecht D.E. & Shane S.M. 1988. Host range of avian influenza virus in free-living birds. *Vet Res Comm*, **12**, 125-141.
10. Ward M.P. 2006. Spread of equine West Nile virus encephalomyelitis during the 2002 Texas epidemic. *Am J Trop Med Hyg*, **74**, 1090-1095.
11. Webster R.G., Bean W.J., Gorman O.T., Chambers T.M. & Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol Rev*, **56**, 152-179.
12. Woebser G.A. 1997. Avian influenza, Newcastle disease, and other paramyxoviruses in diseases of wild waterfowl. Plenum Press, New York, 29-41.
13. World Health Organization (WHO) 2007. Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO, 12 June 2007. WHO, Geneva (www.who.int/csr/disease/avian_influenza/country/cases_table_2007_06_12/en/index.html accessed on 16 June 2007).