In *Proceedings of WHO Consultation on Development and Application of Geographical Methods in the Epidemiology of Zoonoses*. Federal Research Institute for Virus Diseases of Animals, Institute of Epidemiology, Wusterhausen, Germany, May 30 to June 2, 1994.

Application of GIS in Animal Disease Control - Possibilities and Limits

D.U. Pfeiffer [a], R.S. Morris [a] and R.L. Sanson [b]

[a] Department of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand;

[b] Ministry of Agriculture and Fisheries, Batchelar Agricultural Centre, Palmerston North, New Zealand

Summary

Geographic information systems have become an important tool in modern animal disease control. Geographical data in digital form is now widely available, but its accessibility is still limited by the high costs involved. It is expected that costs will come down substantially over the next decade with increased use of tools such as Global Positioning Systems and satellite imagery. GIS software is becoming more sophisticated, but still lacks appropriate tools for advanced spatial analysis. In the field of veterinary epidemiology these geographic information systems have been used widely in field research for visual appraisal and to provide data for advanced spatial analyses. Animal disease information systems are being developed which incorporate GIS as a source of various types of spatial data to be used for epidemiological simulations, risk assessment as well as for production of maps.

Introduction

In recent years geographic information systems (GIS) have become widely used in a number of areas including urban and regional planning, utility management, land suitability assessment, environmental resource monitoring, emergency response management and ecological modelling. Mostly, GIS has been used for retrieval problems and only just are analytical and modelling tools being built into such systems. However, there are still only a few statistical methods for spatial data, and the explosion of geographical data has revealed major database storage problems. In the field of veterinary science the potential of GIS for animal disease control has been explored during the development of decision support systems such as EPIMAN (Sanson 1993). There have been a number of applications where GIS was used in investigations of the epidemiology of animal diseases (see Special Issue: Applications of Remote Sensing to Epidemiology and Parasitology, Preventive Veterinary Medicine 11 (3-4), 1991).

Geographic information systems

Definition of GIS

As quoted in Bailey (1990) the Association of Geographical Information has defined GIS in terms of a computerised data base management system for “capturing, storing, validating, maintaining, analysing, displaying and managing spatially referenced data ... with a primary function to integrate data from a variety of sources.” There are only few systems which would satisfy all these criteria.

Although GIS offers capacity to undertake procedures which were difficult or impossible in traditional mapping and geographical analysis approaches, and hence opens up many new opportunities, the technology is still at an early stage and has not yet come close to its full potential (Openshaw 1987, Goodchild *et al* 1992). There is general agreement that GIS technology needs stronger analysis and modelling capabilities, if it is to meet its potential as a general-purpose tool for handling spatial data.

Burrough (1992) says that “at present geographical information systems are powerful tool boxes in which many of the tools are as strange to the user as a robot-driven assembly plant for cars is to the average home handyman.” He calls for the development of intelligent GIS where a knowledge base is placed next to the GIS helping the user to choose the best set of procedures and tools to solve the problem within the constraints of data, data quality, cost and accuracy.

GIS consists of four major components for input, storage, manipulation and output of geographical data. They are used as tools for analysis, modeling and decision-making.

Components of GIS

When considering the use of a GIS there are number of issues in relation to the four major GIS components which have to be taken into account:

GIS Components	**Use**	**Issues**
Input	digitizing, scanning	cost, quality
Storage	data structures	volume vs. speed, raster vs. vector, layers vs. objects
Manipulation	analysis, modelling	functions, response time, menus vs. commands, error propagation
Output	plot, print, display	cartographic design, visualization

Input

Capture of spatial data is widely done through manual digitizing, despite its high cost and tediousness. The costs for digitizing or scanning can vary from US$ 1 to US$ 40 per polygon. Hence, a 500.000 polygon database would cost between US$ 500.000 to US$ 20 million. Depending on the amount of detail required, the size of the area and the number of layers or coverages data *storage* can become a major problem.

Modern data-capture technology (including Global Positioning Systems (GPS), remote sensing with improved spectral, spatial and temporal resolutions (SPOT-4 and Landsat-7), scan digitizing, field computers and possibly in the near future numerical cameras) can provide accurate and up-to-date geographical data which does not have to be digitized manually from maps (Muller 1993). But, the linkage of remote sensing information with GIS is still seen as a major challenge for the next plateau of GIS advancement (Dobson 1993).

There are a number of problems with regard to data quality. Any digital representation of geographic data introduces bias and error. The National Committee for Digital Cartographic Data Standards identifies six components of digital cartographic data: lineage, positional accuracy, attribute accuracy, logical consistency, completeness and temporal accuracy. Errors can be due to temporal changes since the data was captured, they can occur during capturing and through the process of discretisation required to represent within GIS real continuous or stochastic variation.

Given the availability of spatial data from various sources a major problem is still the conversion and translation between the different formats which are used for storage. This can result in loss of information during conversion or costs of duplicate data aquisition.

Storage

A GIS database captures real geographic variation in the form of a finite number of discrete, digital objects. As geographical variation is typically continuous and complex, the process of capturing reality must involve abstraction, generalization and approximation. The rules defining the objects and their relationships are termed the data model. Data models of GIS take two broad forms, one represents reality as an empty space populated by objects, the other as a set of layers or fields, each defining the spatial variation of one variable (Goodchild *et al* 1992). The layer view predominates in currently available GIS software but the object-oriented approach is emerging as an attractive alternative.

Both object and layer view produce sets of points, lines and areas which can be represented digitally using two types of data structures: *raster*- or *vector*-format. In *raster*-based systems, geographical variation is represented by dividing the world into discrete, uniform-sized elements, normally rectangular, called cells, pixels or grid cells, and specifying their content in a standard sequence, normally row by row from the top left corner. In *vector*-based systems, geographical variation is represented by identifying all point, line and area objects present and specifying each one's location. *Raster*-based systems are convenient for storage and manipulation of region-type features and information from remotely sensed image data is extracted efficiently. However, processing speed can be extremely slow. In order to achieve high resolution, cell size must be very small, and this may result in large data sets which are time-consuming to process. Line-type features are difficult to represent. With *vector*-based systems the resolution can be very high, because the actual coordinates of geographical features are stored. With arc- or polygon features the number of points used to define the shape of the feature determines the resolution. The disadvantage of *vector*-based systems is that manipulation of the information tends to be much more complex.

There is a great need for implementing a temporal dimension in GIS. At present different time periods have to be stored as different maps which are called up for comparison.

Manipulation

Manipulation of geographical information is one of the strengths of GIS technology. The simplest form of manipulation would be querying of data. Different layers of information can be overlayed and logical or arithmetic transformations between them can be performed. Measurements within and between layers can be performed to calculate the size of areas, the length of lines and the distance between features. Buffer zones can be generated around geographical features.

There is a wide range of specific GIS functions available. Goodchild (1992) listed 60 generic functions for data manipulation and analysis. Commercial software such as ARC/INFO (Environmental Systems Research Institute, Redlands, California, U.S.A.) already has about a 1000 different functions.

Despite this proliferation of GIS functions it is widely acknowledged that there still is a lack of integration between GIS and spatial analysis. The analytical functionality of GIS is still comparatively simple. It is limited to primitive geometric operations such as calculating the centroids of polygons, or building buffers around lines and may include more complex operations such as determining the shortest path through a network. There are currently developments where separate spatial analysis software is loosely coupled with GIS (SpaceStat, S-Plus), but there is no effective form of tight coupling where data can be passed from the GIS to the analysis module without loss of spatial structures, such as topology (Goodchild 1992).

Fotheringham and Rogerson (1993) describe a number of impediments to spatial analysis including the modifiable areal unit problem, boundary problems, spatial interpolation, spatial sampling procedures, spatial autocorrelation and goodness-of-fit in spatial modelling which all could lead to misuse of spatial analytical techniques, once they become available as standard functions in GIS. Decision analytical capabilities are just being integrated into GIS systems and will probably become more widely used in the future.

With the amount of data to be manipulated in data files growing all the time, the time required to carry out some spatial operations can become very large. This can be very unsatisfactory for the user and presents increasing demands on computer hardware.

Over the last couple of years GIS software has become more user-friendly, but most systems still require extensive training. More functionality typically results in more complex systems in the early stages of development of such new technologies, until the size of the user group begins to expand rapidly and attention of developers shifts to making the software useable to non-experts. This process is currently in its early stages for GIS.

From the collection of the original data to data analysis and manipulation, error propagation occurs which (depending on the quality of the original information and the procedures during processing) can produce completely erroneous results. It is now generally recognized that GIS systems should provide the user with confidence limits associated with the results of modelling (Burrough 1992).

Output

GIS systems are still predominantly used to replace paper maps with digital maps. Their advantage is clearly that highly specialized maps can be produced on demand either on screen or as printed output. GIS can also be the source of digital geographical data which is transferred to other computer analysis programs for further analysis or as input for decision making tools such as simulation or expert system software.

One of the strengths of GIS is that it is possible to produce 3-dimensional views of reality in combination with effective use of colour which are easier to understand for the user than 2-dimensional abstractions of reality. Map layers can also be turned on and off as required, or combined in various ways into a derivative layer using Boolean operators or more complex rules.

GIS in animal disease control

The potential applications for GIS in animal disease control range from use in epidemiological field studies and simulation to use in animal disease surveillance. The main two areas of use in epidemiological field studies include the visual display of geographical patterns and spatial analysis. GIS provides digital maps which can then be used in epidemiological simulation to contribute realistic geographical information. In the area of disease surveillance GIS can be used to produce maps of disease occurrence and it can be part of a sophisticated animal disease information system.

Reporting and visual display

The visual display of spatial phenomena provides a very effective descriptive analytical tool. Pfeiffer (1994) used this method to describe the spatial occurrence of different strains of *Mycobacterium bovis* in a wild animal population which allowed inferences on the importance of specific disease transmission paths. Lawrence (1991) used GIS to display the distribution of brown ear ticks in southern Africa, retrospectively comparing the ecoclimatic favourability of particular locations for *Rhipicephalus appendiculatus* with the occurrence of East Coast fever.

In the area of disease reporting GIS is a very useful tool for providing maps of the spatial distribution of diseases. Given the availability of an up-to-date database, new maps representing the most current situation can be produced almost instantly and the spatial dynamics of disease occurrence can be monitored over time. The level of aggregation of the original information is under the control of the user who can decide whether to present the location of actual cases or the incidence on a local, regional, country or international level.

Spatial analysis

Spatial analysis using GIS includes a wide range of operations. Typically they relate to analyses within or between layers of geographical data provided by the GIS. In spatial analysis three different types of spatial data can be analysed: *point patterns*, *geostatistical data* and *lattice data*.

Point patterns

The analysis of point patterns is important in veterinary epidemiology as it allows inferences on the occurrence of spatial clustering. The presence of clustering would suggest infectiousness or the presence of specific environmental risk factors. Statistical methodology allows to assess if a point pattern is regular, random or clustered. Based on this information further analyses can be conducted to identify areas of locally increased risk or even factors which influence transmission probability.

Kitron *et al* (1991) studied the occurrence of lyme disease and identified environmental risk factors. Perry *et al* (1991) used GIS to investigate the occurrence of *Rhipicephalus appendiculatus* in Africa to identify the factors controlling the distribution of the vector tick which transmits the parasite *Theileria parva* causing East Coast fever, Corridor disease and January disease in cattle. Pfeiffer (1994) used the GIS to provide for point locations (cases of disease) specific geographical variables such as height above sea level, aspect, slope and distance to features of interest which were then used as explanatory variables in multivariate statistical analysis. Clifton-Hadley (1993) used spatial descriptive measures, spatial autocorrelation and distance to particular features of interest to analyse patterns of occurrence of badger-related tuberculosis breakdowns of cattle herds in south-west England.

Geostatistical data

This type of data represents continuous variation of a feature attribute such as height above sea level in space. GIS cannot represent true continuous variation. An approximation can be achieved by increasing the resolution of a map layer. With geostatistical data techniques for spatial interpolation become important as this kind of data is collected using a sampling procedure and values between sampling points have to be interpolated. Some GIS provide the technique of Thiessen polygons (Voronoi or Dirichlet cells) which divides a region up into polygons around each sampling point assuming that the values for unvisited points within a particular polygon are likely to be similar to the value at the nearest sample point. Pfeiffer (1994) used this technique to generate polygons from point locations describing the areas where particular *Mycobacterium bovis* strains occurred. This technique can be appropriate for presence/absence type information. If the data at the sample points is measured on a continuous scale it is necessary to use a statistical model for the generation of the surface. Techniques such as trend surface regression and kriging are available for this purpose. Kriging is the preferable technique as it allows the interpolation error to be mapped and is considered from the statistical viewpoint the most satisfactory method for interpolation (Oliver and Webster 1990). Pfeiffer (1994) used the geostatistical spatial interpolation technique kriging to produce a surface of disease prevalence from prevalence estimates at sample points. Lessard *et al* (1990) used an inverse distance-weighted mathematical algorithm to interpolate climatic measurements between sample points.

Lattice data

Lattice data represents discrete variation in space based on regular or irregular units. These units can be for example farm or administrative boundaries. GIS are particularly strong in the area of manipulation of lattice data. Yet, for further statistical analysis it is necessary to quantify spatial dependence in such data sets which requires generation of contiguity or spatial weights matrices. This is quite difficult to achieve in currently available GIS software and there are only few specialized spatial statistics software packages (e.g. SpaceSTAT) which can perform such operations.

Simple manipulations of lattice data include overlay operations which allow mathematical or logical operations between layers of data. It is possible to generate layers of data representing the smallest common geometry which contain the attribute information from all original layers. This data can be used for further analysis and evaluation (Haslett 1990).

Lattice data can be analyzed within the same layer of information by testing for the presence of spatial autocorrelation. The Moran and Geary autocorrelation coefficients measure the average relationship between areal units. Hungerford (1991) analyzed the spatial distribution of cattle anaplasmosis between counties within the state of Illinois using second-order analysis and detected significant spatial clustering within the state.

If spatial autocorrelation has been found further analyses are required to describe the underlying spatially stochastic process using spatially autoregressive and/or moving average processes. Explanatory analyses can be conducted between layers of information using simple methods such as multivariate spatial correlation or more sophisticated spatial regression procedures. Hungerford (1991) analyzed the relationship between cattle density and anaplasmosis prevalence on a county basis in Illinois using measures of spatial correlation.

Epidemiological simulation

GIS can provide geographical data which allows computer simulations of the dynamics of infectious diseases for specific geographical locations. Spatial heterogeneity can be represented in simulation models resulting in more realistic representations of reality. There are only few examples where this approach has been used in veterinary epidemiology. Sanson (1993) described a model of foot-and-mouth disease which represents inter-farm spread of the disease on a true geographical area, using various transmission mechanisms. Pfeiffer (1994) developed a geographic simulation model of the dynamics of bovine tuberculosis infection in wild possum populations. The geographical component is a major feature of this model. The model uses vegetation maps to represent the ecological conditions of particular environments.

Animal disease information systems

Disease information systems are beginning to replace largely manual systems which have been used by decision makers for the control of endemic and epidemic diseases. The large amounts of data which can be processed easily, their objectivity and the quickness of response are some of the advantages of computerised animal disease information systems. GIS provides an essential component of such systems. An example of an animal disease information system is EpiMAN which was developed in New Zealand for the management of an outbreak of foot-and-mouth disease (Morris *et al* 1992). The system incorporates a database management system, a GIS, a simulation model of foot-and-mouth disease and expert system elements. It allows rapid integration of important information specific to the geographical setting where the emergency is ocurring. Geographical data used by the system includes property boundary maps, topography, locations of dairy- and meat-processing plants, sale yards and other animal congregation points, and wild animal distribution maps. Simulated plumes of the air-borne spread of FMD virus can be overlayed over property boundaries using the GIS to identify properties at different levels of risk given certain farm characteristics. The system acts as a decision support tool by permitting the expected effects of various control options to be simulated. Output in map form outlining farm boundaries can be produced to be used in the field during farm visits.

The system is currently being extended to produce versions for other purposes. These include endemic disease control and product quality assurance in international trade. They are intended to operate at various administrative levels to suit the nature of the particular disease problem. For example, the endemic disease system can handle disease control decisions down to the district level, and in the system for control of wildlife tuberculosis being developed in New Zealand district managers will be able to assist farmers in developing individual farm control plans. Vegetation / land use data will be derived from satellite photographs, and linked to farm boundary information plus information on farm and animal ownership, in order to plan control programs.

GIS in veterinary public health

GIS will find a variety of applications in veterinary public health, particularly for diseases where environment or habitat factors influence disease occurrence. Rabies is one obvious example where the technique has particular application, but there are numerous examples among other diseases in various countries. GIS will mainly be adopted first in research on zoonoses, but will gradually be adopted for disease control. Most of the management applications of GIS will be in form of integrated decision support systems which make field disease control programs more effective, and target them more precisely to control needs.

References

Bailey,T.C. 1990: GIS and simple systems for visual, interactive, spatial analysis. *The Cartographic Journal* 27, 79-84.

Burrough,P.A. 1992: Development of intelligent geographical information systems. *Internationl Journal of Geographic Information Systems* 6(1), 1-11.

Clifton-Hadley,R.S. 1993: The use of geographical information system (GIS) in the control and epidemiology of bovine tuberculosis in south-west England. In M.V. Thrusfield (ed) *Proceedings of Annual Meeting held by Society for Veterinary Epidemiology and Preventive Medicine*, University of Exeter, 31. March - 2. April 1993, 166-179.

Dobson,J.E. 1993: Commentary: A Conceptual Framework for Integrating Remote Sensing, GIS, and Geography. *Photogrammetric Engineering and Remote Sensing* 59 (10), 1491-1496.

Fotheringham,A.S. and Rogerson,P.A. 1993: GIS and spatial analytical problems. *International Journal of Geographical Information Systems* 7(1), 3-19.

Goodchild, M.F., Haining, R and Wise, S (+ 12 others) 1992: Integrating GIS and Spatial Data Analysis: Problems and Possibilities. *International Journal of Geographical Information Systems* 6 (5), 407-423.

Goodchild,M.F. 1992: *Spatial analysis using GIS.* National Center for Geographic Information and Analysis, Santa Barbara, California, U.S.A.

Goodchild, M.F. 1992: Geographical Information Science. *International Journal of Geographical Information Systems* 6(1), 31-45.

Haslett,J.R. 1990: Geographic information systems: a new approach to habitat definition and the study of distributions. TREE 5(7), 214-218.

Hungerford, L.L. 1991: Use of spatial statistics to identify and test significance in geographic disease patterns. *Preventive Veterinary Medicine* 11, 237-242.

Kitron,U., Bouseman,J.K. and Jones,C.J. 1991: Use of the ARC/INFO GIS to study the distribution of Lyme disease ticks in an Illinois county. *Preventive Veterinary Medicine* 11, 243-248.

Lawrence,J.A. 1991: Retrospective observations on the geographical relationship between *Rhipicephalus appendiculatus* and East Coast fever in southern Africa. *Veterinary Record* 128, 180-183.

Lessard,P., L`Eplattenier,R., Norval,R.A.I., Kundert,K., Dolan,T.T., Croze,H., Walker,J.B., Irvin,A.D. and Perry,B.D. 1990: Geographical information systems for studying the epidemiology of cattle diseases caused by *Theileria parva*. *Veterinary Record* 126, 255-262.

Morris,R.S., Sanson,R.L and Stern,M.W. 1992: EPIMAN - A Decision Support System for Managing a Foot-and-Mouth Disease Epidemic. *Proceedings Fifth Annual Meeting of the Dutch Society for Veterinary Epidemiology and Economy,* Wageningen, 1-35.

Muller,J.C. 1993: Latest Developments in GIS/LIS. *International Journal of Geographical Information Systems* 7 (4), 293-303.

Oliver,M.A. and Webster,R. 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems* 4 (3), 313-332.

Openshaw,S. 1987: Guest editorial: an automated geographic analysis system. *Environment and Planning A* 19, 431-436.

Perry, B.D., Kruska,R., Lessard,P., Norval,R.A.I. and Kundert,K. 1991: Estimating the distribution and abundance of *Rhipicephalus appendiculatus* in Africa. *Preventive Veterinary Medicine* 11, 261-268.

Pfeiffer,D.U. 1994. *The role of a wildlife reservoir in the epidemiology of bovine tuberculosis.* Unpublished PhD Thesis, Massey University, Palmerston North, New Zealand, 496pp.

Sanson,R.L. 1993: *The development of a decision support system for an animal disease emergency.* Unpublished PhD Thesis, Massey University, Palmerston North, New Zealand, 264pp.