

In G.J. Rowlands, M.N. Kyule and B.D. Perry (eds) Special Issue: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics, Nairobi, 15th-19th August, 1994. *The Kenya Veterinarian* 18(2), p 483-485, 1994.

SPATIAL ANALYSIS TECHNIQUES IN VETERINARY EPIDEMIOLOGY

Pfeiffer, D.U. and Morris, R.S.^a

Over the last five years geographical information systems (GIS) have become widely available as management tools for spatial data. These systems are mostly used for production of maps or simple descriptive analyses. The geographic pattern of disease is one of the basic factors in most epidemiological investigations as part of the triad “individual, time, space”. It is now also possible to perform advanced spatial data analysis using standard computer software.

Spatial analysis involves the analysis of data representing geographical features which have a locational attribute such as absolute location (coordinates) or relative positioning (distance). Spatial data can be expressed as points, lines or polygons representing features such as settlements, farms, counties etc.. Additional attribute information such as herd size, farm size, disease prevalence etc. is attached to each feature. The three major types of spatial data include point patterns, geostatistical and lattice data. The georelational database structure of a GIS is ideally suited to storing and manipulating this type of data.

In traditional statistical analysis it is assumed that observations are taken under identical conditions, and independently from one observation to another. The locational attribute of each observation gives rise to spatial effects including spatial dependence and heterogeneity. Ignoring these effects during statistical analysis violates the assumptions of independence and homoscedasticity. During analysis the objective may on one hand be to consider these effects as a nuisance and control for them, or alternatively to statistically describe the source of the spatial effect in order to improve the understanding of the underlying causal process.

ANALYSIS OF POINT PATTERNS

Disease occurrence produces a spatial point pattern where each point may represent the location of an individual case. Evidence of spatial clustering in such a point pattern could be the result of infectiousness or environmental factors. The objectives of spatial analysis are to identify areas of locally increased risk and of factors resulting in spatial interaction which cause for example increased transmission probabilities.

Spatial data analysis applied to point data typically involves the analysis of point distributions in space and the relationship between point distributions and other spatial features. The objective of the analysis would be to determine if the point pattern is “regular”, “random” or “clustered”. Two basic techniques are used: based on counting of cases within small squares (quadrat analysis) and measuring distances to the nearest case (nearest-neighbour analysis). Quadrat analysis summarizes the frequency of points per grid square and calculates the mean/variance ratio to indicate whether the pattern is regular, random or clustered. It is possible to identify particular quadrats which have more or fewer cases than expected by chance. Ripley (1987) concludes that summarizing the original point data into counts per quadrat results in the loss of information by ignoring the position of the quadrats. Nearest neighbour analysis uses the fact that if points are densely packed the distances to their nearest

^a Department of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand

neighbour will be small. The Clark-Evans test based on the average distance from each object to its nearest neighbour, has been popular, but it ignores edge effects and dependence between distances (Clark and Evans 1954). Ripley (1981) describes a number of methods which provide better approximations.

Distance to a geographical feature can be of importance for determining risk of environmental exposure. Pfeiffer (1994) used distance to herds infected with bovine tuberculosis and distance to an area where bovine tuberculosis infection is considered endemic in wildlife reservoir populations as risk factors in the multivariate analysis of a case-control study of tuberculosis breakdowns in New Zealand cattle herds.

ANALYSIS OF GEOSTATISTICAL DATA

Geostatistical data involves continuous variation of a feature attribute such as height above sea level in space. The basis of this type of data is quite often sample point information which is converted into geostatistical data using spatial interpolation techniques. Usually such data is used to produce surfaces showing the variation of the feature attribute in space.

Spatial interpolation techniques range from drawing boundaries around the sample points to using statistical models for estimation of the values between the sample points. The technique of Thiessen polygons (Voronoi polygons or Dirichlet cells) divides a region up into polygons around each sample point assuming that values for unvisited points within a particular polygon are likely to be similar to the value at the nearest sample point. This technique can be quite useful for displaying presence/absence type information collected during a spatial sampling exercise. Pfeiffer (1994) used this technique to visualize the spatial distribution of particular strains of *Mycobacterium bovis* in an infected wild possum population as this was more effective than displaying the original point location data such as trap sites and radio-located den sites. If the values at the spatial point locations represent data on a continuous scale it is preferable to use statistical models in order to represent continuous variation between sampling points. Trend surface regression techniques have been used extensively for this purpose, but have the disadvantage that they are susceptible to outliers and the assumption of independence between data points is clearly violated. The geostatistical spatial interpolation technique kriging is a method of local weighted averaging derived from regionalized variable theory which uses a stochastic approach to take account of spatial variation (Oliver and Webster 1990). One of its advantages is that it provides estimates of interpolation error which can be mapped and is from a statistical viewpoint the most satisfactory method of interpolation. Pfeiffer (1994) used this technique to produce maps of tuberculosis prevalence in a wild possum population.

ANALYSIS OF LATTICE DATA

Lattice data represents discrete variation in space based on regular or irregular units such as disease prevalence per farm. Often this type of data is based on information aggregated within areas defined by administrative boundaries. Spatial dependence in this type of data is quantified by producing a contiguity or spatial weights matrix. This matrix is the basis for analysis of spatial correlation and multivariate spatial correlation as well as for spatial principal components and regression analysis incorporating spatial effects.

Spatial autocorrelation is a measure of the degree of association between data units at a given spatial lag. The Moran and Geary autocorrelation coefficients measure the average relationship between each areal unit and its contiguous neighbours. In the context of disease risk positive spatial autocorrelation identifies areas of high or low risk. Negative

autocorrelation is rather unlikely to occur as it implies that areas of high risk are adjacent to areas of low risk. Tests for spatial autocorrelation are designed to quantify the extent of clustering and allow statistical inference. Getis and Ord (1992) used a distance-based G statistic to investigate the spatial association in the occurrence of sudden infant death syndrome by county in North Carolina.

If spatial autocorrelation has been found, similar to time-series analysis it becomes necessary to describe the underlying spatially stochastic process using spatially autoregressive and/or moving average processes. Inclusion of explanatory variables is referred to as a mixed regressive-spatial autoregressive specification. This type of model can be used either to investigate the spatial effect while controlling for the influence of the explanatory variables or the other way round. This leads to spatial regression models taking account of spatial dependence in the error term or in the dependent variable.

ANALYSIS IN TIME AND SPACE

Often there are temporal as well as spatial factors which influence the occurrence of disease. An infectious disease process may be clustered in space and as well as in time.

If point information on the spatial and temporal distribution of cases of disease is available, techniques such as Knox's test and Mantel regression have been used to investigate space-time clustering. Both methods use the distance in space and time of all pairs of observations in the analysis. Mantel regression was used by Pfeiffer (1994) to examine the occurrence of different strains of *Mycobacterium bovis* in subgroups of a wild possum population for time-space clustering.

Lattice data can be analyzed in space and time using standardized principal components analysis. Eastman (1993) used this technique to identify cyclic seasonal elements of change and to isolate particular change events. The technique can be used to investigate animal disease problems which are subject to seasonal and spatial variation.

REFERENCES

- Clark, P.J. and F.C. Evans 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35 (4), 445-453.
- Eastman, J.R. and Fulk, M. 1993. Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering & Remote Sensing* 59 (6), 991-996.
- Getis, A. and Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24 (3), 189-206.
- Pfeiffer, D.U. 1994. The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished Ph.D. Thesis, Massey University, Palmerston North, New Zealand, 496pp.
- Oliver, M.A. and Webster, R. 1990. Kriging: a method of interpolation for geographical information systems. *Int. J. Geographical Information Systems* 4 (3), 313-332.
- Ripley, B.D. 1981. *Spatial Statistics*. Wiley, New York, 252pp.
- Ripley, B.D. 1987. Spatial point pattern analysis in ecology. In P. and L. Legendre (eds) *Developments in Numerical Ecology*. NATO ASI Series, Springer Verlag, Berlin. G14, 407 - 429.