

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Spatial data requirements for  
animal disease management  
in New Zealand**

A dissertation presented  
in partial fulfilment of the requirements for the degree of  
Master of Veterinary Studies (Epidemiology)  
at  
Massey University

Graham Mackereth  
1998

## Acknowledgements

I would like to thank Prof. Roger Morris for his guidance and for helping me to focus on a single productive objective.

I deeply appreciated the zeal for epidemiology and the love of excellence that Dirk Pfeiffer has attained and models to all his students. His enthusiasm friendship and good teaching will remain with me throughout my career. Thank you for supervising and guiding my work.

I acknowledge and thank MAF Quality Management for their technical, managerial and financial support.

I am fortunate to have worked beside Robert Sanson and to have shared a common love for GIS development. I would like to thank Robert for his technical and moral support.

I am very grateful to Don Scott, who has done me the special service of believing in me, and investing time and money in my education.

I would like to thank Chris Carter for his professional guidance and for encouraging my professional development.

## Abstract

The science of geology has given rise to techniques for managing and analysing spatial data. The techniques often deal with samples that represent a continuum, such as mineral samples taken from various locations. Some animal health data is similar in nature to geo-statistical data, such as climate data or soil samples from various points on a farm. Animal health data is commonly discrete rather than continuous in space. Farms are represented as point or area features and attributes of the farm are attached to the features (Sanson R, 1993). Spatial analysis techniques were reviewed and comment made about their usefulness and validity in disease management.

The spatial data available in New Zealand for managing diseases was examined. Spatial data at a farm level is available in the national database management system Agribase, which records details of rural enterprises. The level of data completeness in Agribase was determined. The number of farms without spatial references varied from 10 to 18 percent, depending on the method used to update Agribase. Spatial data is available for cattle and deer herds in the National Livestock Database (NLDB). The number of herds without spatial data varied from 8 to 15 percent. Changes in the management of land information in New Zealand are resulting in an improvement in the quality and completeness of spatial data.

To determine the likely spatial data requirements for endemic disease management, bovine tuberculosis (TB) data from the NLDB was analysed. Possible applications of spatial data in TB management were developed and tested on point themes and polygon themes, both at different levels of completeness. All of the applications of spatial data that were tested required that a minimum of 85% of farms had spatial references.

The data requirements for sentinel surveillance were examined. A survey of farms in the Wairarapa and Hawke's Bay was undertaken to determine the contribution that slaughter surveillance makes to the TB testing program. The resulting slaughter surveillance information was combined with TB testing data and examined using point themes and polygon themes, and at different levels of data completeness. Polygon themes were found to be the most useful for the display of surveillance information. Gaps in surveillance were only visible using polygon themes. The relationship between vector control operations, testing zones, and TB surveillance could be assessed with a polygon theme.

The appropriate type of data for the detection of clusters of disease was determined. A cluster of farms with TB at Waipawa was used as an example. Polygon themes were preferred for cluster detection. Areas were superior to points in any spatial process that was concerned with contiguous farms. Measures of contiguity based on points were poorly sensitive and poorly specific, when compared with the actual contiguous properties. Some cluster detection processes were relatively insensitive to missing data. Spatial data is used in the response to exotic disease outbreaks. EpiMAN is a decision support system for the management of exotic disease outbreaks in New Zealand. The purpose is to provide rapid information on the location of people and animals and to automate many of the control activities. EpiMAN includes a simulation model called Interspread, which allows assessment of the effects of different control strategies. The performance of Interspread was tested comparing point and area data at different levels of data completeness. Point data was recommended for use with simulation modelling, as it required less processing time. It was found that for spatial processes that use distance, point themes and area themes could not be used interchangeably. Interspread performed



differently with area data than point data. Some adjustment was required when shifting from one data type to another. An adjustment process was developed to account for the difference in distance between two point features and two area features. Incomplete spatial data had an effect on the simulation outcome of Interspread. Adjustment for incomplete data was also possible.

In summary for the management of endemic and exotic diseases, farms should be represented as area features. Point coverage's can be generated from these area features and used in some applications, such as simulation models, and for labelling purposes. To function acceptably the applications tested required that 85% of farms or herds were represented spatially.

## **Introduction**

Modern information systems provide new tools for disease surveillance, control, and eradication. Animal health data can be examined from a spatial viewpoint using a geographical information system. The improvements of computing software and hardware and data management systems now allow a spatial component to be incorporated in disease management. There are a variety of spatial themes available. A simple and relatively inexpensive way for representing farms is to use a point theme. Alternatively a polygon or combination of polygons can be used to represent a farm.

Spatial analysis techniques have been incorporated in to software, so that at a push of a button they can be applied to data. In this dissertation the spatial analysis techniques that are useful and valid for disease management were examined. The level of data quality and type of data that is required for disease management was assessed. The validity of using points or polygons to represent farms was explored in this dissertation.

Chapter 1 contains a review of spatial analysis techniques. Chapter 2 reviews the quality of farm spatial data available in New Zealand. Chapter 3 assesses the spatial data requirements for disease surveillance. Chapter 4 assesses the spatial data requirements for the simulation models. Chapter 5 assesses the data requirements for disease cluster detection. Appendix 1 contains a paper headed "A cross sectional survey of cattle and deer movements in the Hawke's Bay and Wairarapa Regions of New Zealand." Information gained from this survey was used in Chapter 3 to assess the data requirements for TB surveillance.

# List of Contents

<b>ACKNOWLEDGEMENTS</b>	<b>2</b>
<b>ABSTRACT</b>	<b>3</b>
<b>INTRODUCTION</b>	<b>5</b>
<b>CHAPTER 1      REVIEW OF SPATIAL DATA ANALYSIS TECHNIQUES</b>	<b>9</b>
INTRODUCTION	10
POINT DATA	10
AREA DATA	12
SPATIALLY CONTINUOUS DATA	13
SPACE TIME INTERACTION	14
DISCUSSION	14
<b>CHAPTER 2      EVALUATION OF THE QUALITY OF FARM SPATIAL DATA IN NEW ZEALAND</b>	<b>17</b>
INTRODUCTION	18
METHOD	19
RESULTS	19
DISCUSSION	20
<b>CHAPTER 3      SPATIAL DATA AND DISEASE SURVEILLANCE</b>	<b>24</b>
INTRODUCTION	25
METHOD	25
RESULTS	25
DISCUSSION	32
<b>CHAPTER 4      SPATIAL DATA REQUIREMENTS FOR SIMULATION MODELS</b>	<b>57</b>
ABSTRACT	58
INTRODUCTION	58
METHOD	58
RESULTS	59
DISCUSSION	60
<b>CHAPTER 5      SPATIAL DATA REQUIREMENTS FOR CLUSTER DETECTION</b>	<b>64</b>
ABSTRACT	65
INTRODUCTION	65
METHOD	65
RESULTS	66
DISCUSSION	67
<b>DISCUSSION</b>	<b>70</b>
<b>REFERENCES</b>	<b>72</b>
<b>APPENDIX 1</b>	<b>74</b>
<b>A CROSS SECTIONAL SURVEY OF CATTLE AND DEER MOVEMENTS IN THE HAWKE'S BAY AND WAIRARAPA REGIONS OF NEW ZEALAND</b>	<b>74</b>
ABSTRACT	75
INTRODUCTION	75
MATERIALS AND METHODS	75
RESULTS	76
DISCUSSION	78

## List of Tables

TABLE 1: SPATIAL ANALYSIS TECHNIQUES FOR UNIVARIATE AND MULTIVARIATE DATA.....	15
TABLE 2: SPATIAL ANALYSIS TECHNIQUES FOR FIRST ORDER EFFECTS BY DATA TYPE.....	15
TABLE 3: SPATIAL ANALYSIS TECHNIQUES FOR SECOND ORDER EFFECTS BY DATA TYPE.....	16
TABLE 4: THE KNOX CONTINGENCY TABLE.....	16
TABLE 5: TABLE OF DATA QUALITY FOR FARM TYPES IN AGRIBASE BY REGION.....	21
TABLE 6: TABLE OF FARM SIZE AND 'MONTHS SINCE DATA ENTRY' FOR FARMS WITH OR WITHOUT A SPATIAL REFERENCE.....	21
TABLE 7: SPATIAL REFERENCING BY NLDB HERD TYPES.....	22
TABLE 8: SPATIAL REFERENCING BY NLDB HERD TYPES.....	22
TABLE 9: CROSSTABULATION OF TESTS PER HECTARE PER YEAR BY ZONE AND TYPE OF FARM FOR THE HAWKE'S BAY.....	28
TABLE 10: SUMMARY STATISTICS OF TESTS NOT SPATIALLY REFERENCED BY SPECIES AND ZONE.....	28
TABLE 11: FINAL POISSON REGRESSION MODEL FOR PREDICTION OF THE EXPECTED NUMBER OF CATTLE SLAUGHTERED.....	30
TABLE 12: FINAL POISSON REGRESSION FOR PREDICTION OF THE NUMBER OF DEER SLAUGHTERED.....	30
TABLE 13: MORAN'S I VALUES FOR AREA AND POINT DATA IN THE WAIPAWA VECTOR RISK AREA AT THREE LEVELS OF PROXIMITY.....	31
TABLE 14: RESULTS FROM APPLYING CUZICK AND EDWARD'S TEST FOR CLUSTERING (AT 8 NEAREST NEIGHBOURS) OF CASES AT DIFFERENT LEVELS OF CASE DATA COMPLETENESS.....	32
TABLE 15: TABLE OF THE PREFERRED DATA TYPE FOR VARIOUS TB CONTROL APPLICATIONS.....	33
TABLE 16: CROSS-TABULATION OF NUMBER OF INFECTED FARMS RESULTING FROM INTERSPREAD MODELLING STRATIFIED BY SPREAD MECHANISMS AND CONTROL SCENARIO.....	60
TABLE 17: CROSS-TABULATION OF PERCENTAGE DROP IN NUMBER OF INFECTED FARMS AND IN SPREAD MECHANISMS STRATIFIED BY DATA COMPLETENESS, DATA TYPE AND CONTROL SCENARIO.....	61
TABLE 18: SUMMARY STATISTICS FOR FARM RADII IN THE COVERAGE USED WITH INTERSPREAD.....	61
TABLE 19: CROSS-TABULATION OF NUMBER OF INFECTED FARMS AND TRANSMISSION EVENTS BY ADJUSTED POINT DATA TYPE AND SPATIAL DATA COMPLETENESS FOR CONTROLLED EPIDEMIC SIMULATIONS.....	61
TABLE 20: SUMMARY STATISTICS FOR THE RADIUS OF FARMS AT WAIPAWA FOR CLUSTER DETECTION.....	68
TABLE 21: CROSS-TABULATION OF MORAN'S I VALUES FOR TB INFECTED FARMS IN THE WAIPAWA AREA USING 5 DIFFERENT DISTANCES TO DEFINE CONTIGUITY, STRATIFIED BY DATA TYPE.....	68
TABLE 22: CROSS-TABULATION OF NUMBER OF DEER OR CATTLE PURCHASED, BORN, SLAUGHTERED, OR SOLD TO MARKET IN THE YEAR ENDING 30 JUNE 1996 STRATIFIED BY HERD TYPE AND REGION.....	80
TABLE 23: CROSS-TABULATION OF THE CONTRIBUTION OF SLAUGHTER SURVEILLANCE TO THE TOTAL TB SURVEILLANCE IN THE YEAR ENDING 30 JUNE 1996, STRATIFIED BY HERD TYPE AND REGION BASED ON HERDS INCLUDED IN THIS STUDY.....	81
TABLE 24: CROSS-TABULATION OF RISKS OF TB DETECTION BASED ON TESTING AND SLAUGHTER SURVEILLANCE STRATIFIED BY REGION AND HERD TYPE.....	82

## List of Figures

FIGURE 1: FORM USED FOR DATA ENTRY INTO THE MAIN TABLE OF AGRIBASE.....	18
FIGURE 2: EXAMPLE OF THE MAPS POSTED ANNUALLY IN ORDER TO KEEP AGRIBASE UPDATED.....	23
FIGURE 3: TYPES OF SPATIAL DATA AVAILABLE .....	34
FIGURE 4: TESTING ZONES AT THE WAIPAWA VECTOR RISK AREA.....	35
FIGURE 5: CATTLE HERD TYPE DISTRIBUTION SHOWN WITH AREA DATA.....	36
FIGURE 6: CATTLE HERD TYPE DISTRIBUTION SHOWN WITH POINT DATA.....	37
FIGURE 7: THE DISTRIBUTION AND SIZE OF DEER HERDS DISPLAYED WITH AREA AND POINT DATA.....	38
FIGURE 8: COVERAGE OF CATTLE TB TESTING IN A ONE-YEAR AND THREE-YEAR PERIOD.....	39
FIGURE 9: THREE-YEAR SUMMARY TB TESTING DENSITIES FOR CATTLE SHOWN USING AREA DATA.....	40
FIGURE 10: FIGURE SHOWING THE DISTANCE TO FARMS THAT TEST.....	41
FIGURE 11: MAP OF THE WAIPAWA VECTOR RISK AREA SHOWING GAPS IN THE TB TESTING COVERAGE .....	42
FIGURE 12: VECTOR CONTROL METHODS AND HOURS WORKED IN THE WAIPAWA VECTOR RISK AREA.....	43
FIGURE 13: APPARENT GAPS IN THE COVERAGE OF POSSUM CONTROL WORK.....	44
FIGURE 14: REPRESENTING SURVEILLANCE WITH POINT DATA USING PIE CHARTS AND COLOURS.....	45
FIGURE 15: REPRESENTING SURVEILLANCE WITH POINT DATA USING BAR CHARTS AND DIFFERENT SIZED DOTS .....	46
FIGURE 16: REPRESENTING SURVEILLANCE WITH POINT DATA USING TESTING CONTOURS AND DEVIATION FROM THE MEAN TESTING LEVEL .....	47
FIGURE 17: VISUALISING MISSING DATA WITH POINT DATA .....	48
FIGURE 18: VISUALISING MISSING DATA WITH AREA DATA.....	49
FIGURE 19: THE EFFECT OF MISSING DATA ON DETECTING GAPS IN THE TB TESTING COVERAGE.....	50
FIGURE 20: FIGURE SHOWING THE CONTRIBUTION OF SLAUGHTER TO THE TB SURVEILLANCE PROGRAM.....	51
FIGURE 21: THE CONTRIBUTION OF FATTENING HERDS TO THE TB SURVEILLANCE IN THE WAIPAWA VECTOR RISK AREA.....	52
FIGURE 22: FIGURE SHOWING THE HIGH LEVELS OF TB SURVEILLANCE PER HECTARE IN DEER HERDS IN THE WAIPAWA VECTOR RISK AREA.....	53
FIGURE 23: TB INFECTED FARMS AT WAIPAWA SINCE 1990 AND A RANDOM SELECTION OF FARMS.....	54
FIGURE 24: POINT LOCATIONS OF ALL FARMS, THE TB INFECTED FARMS, AND THE RANDOM SELECTION OF FARMS AT WAIPAWA .....	55
FIGURE 25: CLUSTER ANALYSIS OF TB INFECTED HERDS AT WAIPAWA COMPARED WITH RANDOM CONTROLS .....	56
FIGURE 26: LINE CHART OF THE NUMBER OF INFECTED PROPERTIES RESULTING FROM A CONTROLLED EPIDEMIC UNDER DIFFERENT DATA ADJUSTMENT METHODS.....	62
FIGURE 27: FIGURE SHOWING THE RESIDUALS BETWEEN THE ACTUAL RADIUS OF A FARM LESS THE MEAN RADIUS OF ALL FARMS.....	63
FIGURE 28: THE AVERAGE NUMBER OF CONTIGUOUS NEIGHBOURS IN A GIVEN NUMBER OF METERS FROM A FARM CALCULATED WITH DIFFERENT DATA TYPES.....	69
FIGURE 29: BAR CHART OF CATTLE PURCHASE PATTERNS BY HERD TYPE.....	79

## List of Abbreviations

CSR	COMPLETE SPATIAL RANDOMNESS
DCDB	DIGITAL CADASTRAL DATABASE
DOSLI	DEPARTMENT OF SURVEY AND LAND INFORMATION
FMD	FOOT-AND-MOUTH DISEASE
ICS	INDEX OF CLUSTER SIZE
NLDB	NEW ZEALAND VETERINARY JOURNAL
TB	BOVINE TUBERCULOSIS
SUFI	STATIC UNIQUE FEATURE IDENTIFIER

## **Chapter 1**

### **Review of spatial data analysis techniques**

## Introduction

The science of geology has given rise to techniques for managing and analysing spatial data. These techniques often deal with samples that represent a continuum. Some animal health data is similar in nature to geo-statistical data, such as climate data or soil samples taken from various points on a farm. Animal health data is commonly spatially discrete in its nature. Farms are represented as discrete point or area features and attributes of the farm are attached to the features (Sanson R, 1993). In this Chapter spatial analysis techniques are reviewed and comment made about their usefulness and validity in disease management.

Spatial data analysis covers a range of techniques which can be used to explain or describe a process operating in space and its relationships to other spatial phenomena. Table 1 lists techniques for analysing spatial data. Simple point locations can only be analysed with respect to the presence of trends. Additional analyses can be conducted if attribute data is available for each location. The objective is to analyse the variation in attributes, conditional on their location. In the multivariate case, a vector of attributes is present at each location, one of which may be a temporal element.

Space can be represented using two *views*, the *entity view* and the *field view*.

The *entity view* describes space on the basis of discrete objects. Under this view spatial objects have features or attributes. These attributes can be nominal, ordinal or interval. The object is represented as a point, line or area feature. The feature used to represent an object such as a farm is significant. It affects the type of analysis that can be performed, and the degree in which the real world is being represented in the analysis.

The *field view* represents space as a continuum using discrete points, lines or areas. A value at a given point is part of a subset of a continuum of values in that area.

Variation in the mean value of the process in space is defined as a *first order* effect. It represents a trend or drift in the overall variation and in the mean value depending on location. A *second order* effect is the result of the tendency for neighbouring values to follow each other. There may appear to be an 'attraction' or 'repulsion' between local values. Table 2 and 3 show the techniques used for analysis of first and second order effects. In practice, a spatial process may involve first as well as second order effects. One component of variation may be dependent on the location, and another component on covariates.

## Point data

Point data can be purely the location of an event or object. Points can be distributed in a clustered, regular or random pattern. Their distribution can be associated with the population-at-risk or other factors. In all cases the nature of the point data should be considered. Does the point feature represent a true point in the real world?

Rainfall measurements can be collected at actual points. They represent a continuum of rainfall over the field, so they could be used to construct a surface depicting the variation in rainfall over a region.

Points are often used to represent the location of areas, such as farms or towns. For many purposes, such as labelling, this creates no difficulty. Point features are easier to capture, store and analyse than area features, but they can significantly distort the spatial characteristics of the objects that they represent. Representing proximity as the distance between points can distort the true proximity between the farms. A difficulty with animal health data is that farm size is often clustered (large farms are found together and small

farms are found together). The distance between a large farm and its immediate neighbour is greater than the distance between a small farm and its neighbour. Points can represent discrete objects, such as water troughs, that are to all intents and purposes points in *real life*. In this case analytical techniques that use discrete points can be used with confidence.

### **Point data and first order effects**

First order effects (trends across a region) can be explored using quadrature counts or kernel estimation. The region can be divided into squares and the number of events per quadrature counted. A three-dimensional histogram showing variation in intensity throughout the region can be constructed. This measure of intensity removes the exactness of the original locational data. To minimise this loss, a moving window approach can be used, where each estimate of intensity is calculated for the area of the moving window. The size of the aggregation affects the outcome. Kernel estimation uses the original point locations to produce a histogram of intensity. The smoothed value at any point is the weighted-average of all other values in the field. The weight arises from the probability distribution at that point and is referred to as the kernel (Diggle, 1985). This is essentially a risk surface for the occurrence of those events. Varying the bandwidth can control the degree of smoothing. Adaptive kernel density estimation involves varying the bandwidth to take into account the local intensity of events.

### **Point data and second order effects**

Second order properties (apparent local attraction or repulsion) for point data can be investigated using the distances between points. Disease or production problems may be clustered in space. The ability to detect real clusters of disease (perhaps associated with a trace element deficiency or a reservoir host), rather than clustering due to chance is important. Clustering can be assessed using general or focused methods. The focused method relates to the clustering of events around a particular fixed point (Alexander and Cuzick, 1992). General clustering techniques detect the presence or absence of clustering when there is a heterogeneous population-at-risk.

A point pattern from a random spatial process should follow a homogeneous Poisson distribution with no first or second order effects. This model of complete spatial randomness (CSR) can be used to assess if a point pattern is clustered, regular or random. Tests for this include: the index of dispersion test based on the quadrature counts; nearest neighbour tests; the Clark-Evans event to event test; and the Hopkins random point to event test. The distribution function of the nearest neighbour distances can be used. Clustering can be visually examined by comparing the probability distributions of the observed nearest neighbour distances with the distribution functions of CSR or other hypothesised models.

The K function (Ripley, 1977) looks at all inter-event distances rather than just the nearest neighbours.  $K(d)$  is the expected number of events within a distance ( $d$ ) of an arbitrary event. This is graphically or statistically compared with hypothesised explanatory models or spatially random models.

When assessing clustering it is necessary to adjust for underlying variation in the population-at-risk. The kernel estimate for the event (which is an event per unit area) can be divided the population density at that point; the result is a kernel estimate of events per unit of population. It is possible to establish a kernel density estimate for the underlying population density. The estimate of events per unit population can be replaced by the ratio between the two estimates of kernel intensity. It is also possible to use another process that is representative of population variation.



If a set of controls is available for each case a different approach can be used. The controls can be used to represent the distribution of the population-at-risk. Where there is no clustering in the cases relative to the controls then the cases are a random sample of both cases and controls. The hypothesis tested is that of random labelling of cases and controls rather than that of independence (Kingham, Gatrell and Rowlingson, 1995). The Cuzick and Edwards's test, which is available in the software "STAT!" is a case control test. The principle is that assuming disease clustering, the nearest neighbour to a case should be a case, rather than a control. The count of the number of cases whose nearest neighbours are cases rather than controls is the test statistic  $T(k)$ . The expected value  $E(t)$  of  $T(k)$  can be calculated under the null hypothesis, which is that cases and controls are sampled from a common spatial point distribution.

## Area data

Area data can be associated with arbitrary administrative boundaries or with natural area features such as lakes or alpine zones. The data may be events associated with the area, or data values for the area. The area can be represented textually as hectares, or spatially as complex polygons, surveyed parcels, or grids.

Proximity can be measured using a  $(n \times n)$  spatial proximity matrix  $W$ , where  $w_{ij}$  represents the spatial proximity of areas  $i$  and  $j$ . The matrix can be calculated using a range of methods. If two objects are considered to be in close proximity  $w_{ij} = 1$ . The element  $w_{ij}$  can be set to 1 under a variety of conditions:

- 1 where the centroid of  $i$  is one of the  $k$  nearest centroids to the centroid of  $j$ ;
- 2 where the length of the common boundary with  $j$  divided by the perimeter of  $i$  is greater than a certain proportion;
- 3 where the centroid of  $i$  is within a specified distance to the centroid of  $j$ ;
- 4 where  $j$  shares a common boundary with  $i$ .

The element  $w_{ij}$  can equal 0 where the above conditions are not met. A series of proximity matrices can be used to represent different spatial lags.

## Area data and first order effects

A simple way to assess variation in an attribute value globally is to use weighted averages of the attribute. The average is weighted by neighbouring values. The proximity matrix provides a method for determining a suitable set of weights.

For data on a regular grid, summing the grid rows and columns can show broad spatial trends. The row and column means could be examined using ANOVA. Median polish is a technique that uses medians, not means, and thereby overcomes the problem of extreme values. The difficulty with median polish is that the direction of the grid bears no relationship to trends. Banding effects are seen due to the row and column basis of the method.

## Area data and second order effects

Moran's  $I$  is closely related to the covariogram (discussed later) and Geary's  $C$  is related to the variogram. Both,  $I$  and  $C$  can be used to estimate the spatial auto-correlation at different lags. A correlogram can be used to show these correlation's. Local indication of spatial association can be seen using the Moran scatterplot and spatial lag pies. Quantitative estimates can be obtained using the  $G$  statistic by Getis and Ord (1992) or other indicators (Anselin, 1995). Large first order effects affect all of the above approaches.

Clustering can be detected using Moran's  $I$ . Moran's  $I$  for the observed value is compared with  $n$  permutations of the possible data value arrangements. If the observed

value of  $I$  is an extreme value in the permutation distribution, this shows significant autocorrelation. The observed Moran's  $I$  can also be compared to  $n$  random permutations in a Monte Carlo approach.

## **Spatially continuous data**

The objective in analysing spatially continuous data is to understand the spatial distribution of values over the whole study region, given a set of values at sample points. This may enable predictions of values at other points or predict values in a variety of circumstances.

Continuous data can be mapped using a variety of devices such as symbols, columns, colours and pies. A surface can be created using interpolation or triangulated irregular networks. Proportional symbols are often used, with the size of the symbol reflecting the data value. In the cases of circular symbols, the human mind perceives a circle of twice the area, as being less than twice the size. Overlapping circles make it difficult to visualise trends. It is often helpful to place the values within a range of classifications. The number of classifications recommended is  $(1+3.3\log n)$  where  $n$  is the number of observations. Other adjustments may be required to deal with outliers. It is possible to show continuity such as isobars on a weather map.

### **Continuous data and first order effects**

The main techniques for examining global trends are spatial moving averages, tessellation and kernel estimation. Global trends can be described by a spatial moving average that interpolates values between a given number of neighbouring points. This smoothes the data and thereby reveals overall trends. A three point moving average is the unweighted average of the nearest three points. If more points are used the result is smoother. The process can also be adjusted for the distance between points.

A tessellation (tiling) of observed sample points could be used to show trend. The most common technique is to use Delauney triangulation (triangulated irregular network or TIN). This method assigns to each sampling point a territory in which each point is closer to the sampling point than any other. The resulting polygon map is called a Dirichlet tessellation and the tiles are known as Voronoi or Thiessen polygons. The Delaunay triangles are a set of non-overlapping triangles with data points at the vertices. The data values represent the height of the vertices. The values at other locations can be taken from the sides of the triangle. Isolines can be drawn through points of equal value in order to generate a contour map. Repeated tessellation can be used in a process called natural neighbour interpolation.

Polynomial regression techniques are utilised to fit successive groups of data points and create a smooth surface. This is the basis of spline smoothing. Kernel estimation also allows conversion of sampling points into a surface, using the value at the point rather than the number of points in a given area. This is in effect a more sophisticated moving average.

### **Continuous data and second order effects**

Spatial dependence between attribute values measured at sampled locations taking into account their distance from each other is described using the covariance function (covariogram). In the presence of second order effects there is strongly positive covariance between observations a small distance apart and less covariance between values further apart. The covariance function describes how the variation in deviations from the mean value varies depending on the distance between sampling locations. The function assumes stationarity. A covariogram describes the covariance for varying

distances  $h$  between sample points. The correlogram shows the corresponding correlations at a given distance.

A semi-variogram is a graphical representation of the variation between sampling points separated by a given distance and direction. It assumes stationarity and isotropy. For stationary spatial processes all three describe similar information. Examination of the sill, range and nugget enables the degree of stationarity to be assessed.

An isotropic covariogram can be used to reveal the appropriate lags. Variograms at these lags could be run in two or three directions to explore the possibility of directional effects.

Kriging is a weighted-moving-average technique for estimating the value of a spatially distributed variable from adjacent values while considering the interdependence expressed as a variogram (Isaaks and Srivastava, 1989). It allows the interpolation error to be mapped (Oliver and Webster, 1990).

## **Space time interaction**

Interaction can be present in time as well as space. Space-time relationships can be assessed using three main methods. They all involve matrices of the spatial and temporal distance. The Knox method (Knox, 1964) compares the contingency table shown in Table 4 with that obtained from the Poisson distribution. The problems associated with this process include determining the classification thresholds and the non-independence of events. (Glick, 1979). Mantel's test (Mantel, 1967) uses numerical, rather than categorical, measures of closeness in time and space. Interest can be restricted to pairs that are meaningfully close.

## **Discussion**

Techniques for analysing spatial data have been developed and are available for use with animal health data. The validity of using them depends on the nature of the data and the technique being applied. Care should be exercised when applying “push-button” spatial techniques to animal health data. There are a number of potential pitfalls. Some are:

- 1 edge effects;
- 2 variation in the population at risk;
- 3 first order effects can impact on tests for second order effects;
- 4 distortion of proximity if points are used to represent area features;
- 5 and using continuous data analysis techniques on discrete data.

The techniques used in this dissertation are marked with an asterisk in Tables 1, 2 and 3.

**Table 1: Spatial analysis techniques for univariate and multivariate data**

<b>Data structure</b>	<b>Univariate</b>	<b>Multivariate</b>
<b>Purely locational data</b> Events or objects at points lines or areas	Nearest neighbour* K functions Kernel density estimation Kernel regression Bayesian smoothing*	Bivariate K functions Space time interactions* Adaptive kernel density estimation Kernel regression Bayesian smoothing*
<b>Attribute data</b> Discrete or continuous values associated with points, lines, or polygons	Spatial auto-correlation Spatial correlograms Variograms Trend surface analysis Kriging	Multivariate spatial correlation Spatial regression Co kriging Spatio-temporal models Spatial general linear models Cluster analysis*

**Table 2: Spatial analysis techniques for first order effects by data type**

<b>Techniques</b>	<b>Point data</b>	<b>Area data</b>	<b>Spatially Continuous data</b>
<b>Descriptive analyses</b>	Intensity Quadrat methods Kernel estimation Kernel density estimation* Adaptive kernel density estimation	Spatial moving average* Median polish Kernel estimation	Spatial moving average Tessellation Kernel estimation using values
<b>Comparing observed distributions with hypothetical distributions</b>	Complete spatial randomness CSR Index of dispersion simple quadrat test Nearest neighbour tests* Clark Evans test K function test Heterogeneous Poisson process Cox process Poisson cluster process* Simple inhibition processes Markov point processes.	Nearest neighbour*	
<b>Bivariate or multivariate patterns, Time space interaction</b>	Chi squared test on sample quadrates Nearest neighbour methods* Knox method* Mantel method K Function method		
<b>Interpolations</b>	Bayesian smoothing* Spatial moving averages* Median polish Trend	Kernel regression Bayesian smoothing ICM	Spatial tessellation TIN and Thiessen polygons* Spatial moving average * Trend surface regression (LS)
<b>Modelling first Order effects</b>	Homogenous Poisson process Monte Carlo	Kernel Regression Spatial principle components	Trend surfaces Spatial general linear models (for counts, proportions)
<b>Accounting for Population at risk or other variables</b>	Kernel density estimate ratios heterogeneous Poisson process Cox process Poisson cluster process G statistic	Moran's I Geary's C Correlogram G statistic Lagrange multiplier	Co kriging Universal kriging Disjunctive kriging

\* Techniques applied in this dissertation

**Table 3: Spatial analysis techniques for second order effects by data type**

<b>Techniques</b>	<b>Point</b>	<b>Area</b>	<b>Continuous</b>
<b>Contiguity</b>	Nearest neighbour * Event to event Random point to event K Function K(d) Spatial Auto-correlation Moran's I*	Moran's I Geary's C Correlogram G statistic Contiguity	Covariance function Covariogram correlogram Semivariogram
<b>Tests for Clustering</b>	Homogenous Poisson CSR index of cluster size Nearest neighbour* Logistic regression and Bivariate K- function K Function* Geographical analysis machine Cuzick and Edward's Case control*	Clustering; Moran scatterplot* Spatial lag pies Spatial weights matrix Local indicator of spatial auto-correlation (Lisa)	
<b>Time space interactions</b>	Knox method* Mantel Method K function nearest neighbour		
<b>Interpolations</b>	Bayesian smoothing* Spatial moving averages* Median polish Trend surface regression (LS) Splines* Tessellation	Bayesian smoothing Trend surface regression	Kriging Splines Kernel smoothing
<b>Modelling second order effects</b>		Spatial Auto correlation Multivariate spatial correlation Spatial principal components Spatial regression	Kriging Splines Kernel smoothing
<b>Accounting for Population at risk or other variables</b>	Adaptive kernel density estimation heterogeneous Poisson process Cox process Poisson Cluster Process G Statistic	Moran's I* Geary's C Correlogram G statistic Lagrange multiplier	Co Kriging Universal Kriging Disjunctive Kriging TIN

\* Techniques applied in this dissertation

**Table 4: The Knox contingency table**

<b>Time/Space</b>	<b>Less than the Critical Distance</b>	<b>Greater than the Critical Distance</b>
<b>Less than the Critical time</b>	Close together and contemporaneous	Far apart and contemporaneous
<b>Greater than the Critical time</b>	Close together with time gap	Far apart with time gap

## **Chapter 2**

# **Evaluation of the quality of farm spatial data in New Zealand**

## Introduction

There are many sources of spatial data that are potentially useful for disease management. To facilitate farm level data analysis the location and attributes of farms or herds is necessary. There are two sources of this information in New Zealand, Agribase and the National Livestock Database (NLDB). In this chapter the quality of the spatial data in these databases is assessed.

The national livestock database (NLDB) is designed to facilitate the management and administration of Disease Control Programs. It is currently used in the tuberculosis (TB) control program. It deals with people, herds and farms. In the database, people are linked to their herds and herds are linked to farms. This flexibility allows for herds to be shifted from farm to farm or owners to change without the loss of historic herd information. When a herd is linked to a farm, the spatial references of the farm can be used. The resulting linkage allows spatial data analysis of herd information.

Agribase is a national database of rural enterprises. Agribase is based on a concept of rural enterprises and includes forestry and horticulture as well as livestock enterprises. Conceptually, each enterprise or farm is characterised by a boundary and a point location, as well as details of the enterprise, such as stock numbers or hectares of crops. Agribase is a relational database with a central farm table and four subsidiary tables linked by the key field (farm\_ID). The main table contains information about the owner, manager, as well as the trading name of the enterprise. Figure 1 shows the form used for data entry into the main table.

Figure 1: Form used for data entry into the main table of Agribase

The screenshot displays the 'Application agribase' window with a menu bar (File, Edit, Entry, Report, Security, Help) and a toolbar with icons for Add, Delete, Clear Fields, Search, Update, First, Prior, Next, Last, and Exit. The main window is titled 'Agribase Master Detail' and contains the following fields:

Farm Id:	1507391	Archived:	<input type="checkbox"/>	Surname:	Mackereth	GF
Station Name:						
Trading Name:						
Mngr Initials:	M	Surname:	Bourke	Lessee:	<input checked="" type="checkbox"/>	
Farm Type:	FRU	Fruit growing:	<input checked="" type="checkbox"/>	Farm Size:	2.2	
KDM Initials:	GF	Surname:	Mackereth			
Address:	112 Brookvale Rd			Rd Location:	Brookvale Rd	
				Town/Locality:	Havelock North	
				Brazilian Addr:	112	
Town:	Havelock North			Postal Code:		
Phone:	06 - 877 4967	Residential:	<input checked="" type="checkbox"/>	Fax:		
Map No:	V21	Farm Gate East:	2 844 300	North:	6 164 270	
Farm Dwelling East:	2 844 300	North:	6 164 270	Lat:	39S 39 47	Long: 176E 53 48
Source:			<input checked="" type="checkbox"/>	Date:	10/10/1995	
Comment:						

At the bottom of the form is a row of buttons: View Records, Enterprise, Valuation, Other DB, Flag Type, Send Mail, and Address Updated. The status bar at the very bottom shows 'Ready'.

The Valuation department keeps records of land ownership. The department of Survey and Land Information (DOSLI - Now disbanded) digitised all land parcels in New Zealand. Land ownership can be linked to the digitised land parcels by a common field containing the legal description. Unfortunately the legal descriptions used by the Valuation department differ in form from those used by DOSLI. The link between the land ownership information and the digital cadastral database is only partial. The spatial references for farms in Agribase have been obtained by matching Valuation Roll-legal descriptions with legal descriptions in the Digital Cadastral Database (DCDB). The DCDB has a legal description and a static unique feature identifier (SUFI). The Valuation Roll has ownership details, a valuation number and a legal description. Electronic processes can match the legal descriptions and produce a table of "Valuation number-SUFI" pairs. The process matches approximately 7 out of 10 rural valuation numbers with at least one SUFI (land parcel) on the DCDB. The resulting "Valuation number-SUFI" table is used to link land parcels to owners. Area data for a farm in Agribase can be obtained by entering the valuation number(s) of the farm. If the valuation number is present in the "Valuation number-SUFI" table, then the SUFI's (land parcels) are assigned to the farm. Not all farms in Agribase have a spatial reference, as the linkage is partial. The percentage of valuation numbers linked to the DCDB varies from district to district. Linked farms may only have one of many land parcels represented. For example, if the legal description in the Valuation roll reads "Lots 1 to 10 DP 2468", the matching process may only link the farm to Lot 1 and Lot 10. More sophisticated matching has been attempted to unravel the valuation roll legal description. However with concatenated legal descriptions, not all lots are necessarily linked to a farm. Considerable improvements in the matching process have occurred during the course of this dissertation. In some areas manual matching of Farms to the DCDB has been undertaken. In Hawke's Bay, farmers are sent maps of their farms each year and they are requested to mark their boundaries and homestead location on these maps. This process is now used in the South Island and much of the North Island. Figure 2 is an example of one such map. The process allows land parcels to be directly allocated to farms. The land parcels are entered directly against the farm without using the Valuation SUFI table. This approach results in a higher degree of spatial reference than the valuation matching approach.

## **Method**

The data from two regions was examined and tabulated. A comparison between the Hawke's Bay data and data from the Wairarapa was made. In the Wairarapa the spatial data is derived from the valuation matching process, whereas the Hawke's Bay data is derived from the manual matching process. Data completeness in the Hawke's Bay Region and Wairarapa Region was examined.

## **Results**

### **Spatial references for Agribase**

Table 5 shows the degree of linkage to the DCDB and how it has been achieved. The number of farms in each farm-type, and the percentage of each farm-type that has co-ordinate information, are shown. The percentage of each farm-type that have at least one SUFI (land parcel) and the percentage that have at least one valuation number are also presented. The percentage of each farm-type that has no spatial data is shown. The data



was split into the two groups. One group comprised the predominant farm-types that are of principal interest to TB or exotic disease control. The "other" group is composed of the many horticultural and miscellaneous farm-types, such as Ostrich farming.

In the farm-types of principal interest, the Hawke's Bay had 10% of farms with no spatial reference, whereas the Wairarapa had 18% with no spatial data. There were overall 11170 SUFIs in the Hawke's Bay data. Of these 8186 or 57% were obtainable using the valuation matching process. The remaining 43% were not in the Valuation-SUFI table.

Hawke's Bay farms with a valuation number had an average of 3.52 SUFIs. On average 2.58 of these were in the Valuation-SUFI table. This shows that the valuation matching process is only finding on average 73 percent of the land parcels within a farm. Over all, 16% of farms are not matched to any land parcels by the valuation matching process.

The number of land parcels per farm increases with the direct approach used in Hawke's Bay, as well as the number of farms with land parcels. Table 6 shows the percentage of farms with no spatial reference by predominant farm type. The valuation derived matching process used in the Wairarapa appears to create missing data in a random pattern. The Hawke's Bay is missing Beef, Dairy and Deer farms at about the same frequency. Table 6 shows that the median and average size of farms with no spatial reference is much smaller than the ones with spatial reference. This may reflect the focus of Agribase data collection on farms with at least 5 hectares.

Farms with or without spatial reference were compared as to when they were last edited in Agribase. Table 6 shows that farms with no spatial reference are much "older". The median time since the data was last updated was 9 months, in farms that had a spatial reference and 17 months, in farms that had no spatial reference. Farms with missing data had been in the database twice as long as farms that had spatial data. This suggests a redundancy problem. Some current farms need to be archived. The proportion of farms with spatial references may be greater if farms were archived at the appropriate time.

### **Spatial references for the NLDB**

Table 7 shows the percentage linkage for each herd-type. The completeness of spatial referencing for NLDB herds is slightly better than for Agribase. In Hawke's Bay only 8 percent of herds did not have a spatial reference, while 15 percent of Wairarapa herds had no spatial reference. The herd-type least likely to be spatially referenced is the miscellaneous herd type. Table 8 shows that the herds without spatial reference are smaller, reflecting the focus of Agribase on properties of 5 hectares or more. There is considerably less redundancy in the NLDB. The median time period since the last activity associated with a herd was under 12 months for each region. The herds without spatial reference had a longer median time since the last episode, but only by a few months.

## **Discussion**

In the Wairarapa the spatial references for farms have been derived from a valuation matching process and 16 to 19% of farms have no spatial reference. The farms that had spatial reference were estimated to be missing 27% of the land parcels that make up the whole farm. Small farms are under represented in the data set. This is a potential source of bias when analysing the data. Any process that was designed to make use of the spatial data in Agribase would have to be robust enough to cope with this level of farms that have no spatial reference and with farms that are only partially represented.

The percentage of herds with spatial references in the NLDB was higher than the percentage of farms with spatial references in Agribase. This is a reflection of the high

level of usage associated with the NLDB. The difference between the Wairarapa and Hawke's Bay was considerable. Fifteen percent of herds in the Wairarapa had no spatial reference, compared with 8% in Hawke's Bay. This represents the difference in the source of the spatial data. Small or miscellaneous herds were under represented in the data set. This is a potential source of bias when analysing animal health data

**Table 5: Table of data quality for farm types in Agribase by region**

Region	Predominant Farm type	Number of farms	% of farms with SUFI's	% of farms with a XY location	% of farms with a valuation number	% of farms with a valuation number but no SUFI	% of farms with a SUFI but no valuation number	% of farms with no spatial data
Hawke's Bay	Beef	902	78	82	72	3	9	18
	Dairy	73	89	86	86	4	7	11
	Deer	146	84	81	84	7	7	16
	Pig	5	80	100	80	0	0	0
	Sheep	841	98	99	94	1	5	1
	Sheep-Beef	510	93	92	93	5	5	7
	<b>Sub Total</b>	<b>2477</b>	<b>88</b>	<b>90</b>	<b>85</b>	<b>3</b>	<b>6</b>	<b>10</b>
	Other types	1982	72	42	62	11	21	28
	<b>Total</b>	<b>4459</b>	<b>81</b>	<b>69</b>	<b>75</b>	<b>7</b>	<b>13</b>	<b>18</b>
Wairarapa	Beef	887	76	76*	95	19	0	24
	Dairy	612	91	91	99	8	0	9
	Deer	86	66	86	99	34	1	34
	Pig	12	67	12	100	33	0	33
	Sheep	1264	88	1264	100	12	0	12
	Sheep-Beef	214	62	62	97	35	0	38
	<b>Sub Total</b>	<b>3075</b>	<b>82</b>	<b>82</b>	<b>98</b>	<b>16</b>	<b>0</b>	<b>18</b>
	Other types	1530	53	53	77	24	0	47
	<b>Total</b>	<b>4605</b>	<b>72</b>	<b>72</b>	<b>91</b>	<b>19</b>	<b>0</b>	<b>28</b>

\* The percentages in italics show that co-ordinate data can be calculated for any farm with at least 1 SUFI.

**Table 6: Table of farm size and 'months since data entry' for farms with or without a spatial reference**

Region	Farms with spatial reference	Number of farms with spatial reference	Median size of farms (Ha)	Average size of farms (Ha)	Number of farms with a 'last entry' date	Median number of months since data entered	Average number of months since data entered
Hawke's Bay	Present	3717	43	294	3178	9	16
	Missing	697	9	44	698	17	19
Wairarapa	Present	3670	62	453	3724	9	20
	Missing	593	12	85	674	17	45

**Table 7: Spatial referencing by NLDB herd types**

Region	Species	Herd Type	Number of herds	Number of herds without land parcels	Percent of herds without any land parcels	Percent of herds without XY co-ordinates	Average number of land parcels per herd
Hawke's Bay	Cattle	Breeding	1318	87	6.6	8.9	5.4
		Drystock	807	68	8.4	13.8	3.5
		Dairy Drystock	3	1	33.3	33.3	1.0
		Dairy Herds	71	3	4.2	12.7	3.9
		Miscellaneous	188	43	22.9	40.4	1.9
		All	2387	202	8.5	13.2	4.5
	Deer	Breeding	267	18	6.7	11.6	5.3
		Fattening	11	1	9.1	9.1	2.5
		Miscellaneous	30	2	6.7	20.0	5.5
		Safari Park	1	0	0.0	0.0	1.0
		Velvetting	47	1	2.1	4.3	3.3
		All	356	22	6.2	11.2	5.0
	Both		2743	224	8.2	12.9	4.5
Wairarapa	Cattle	Breeding	779	80	10.3	10.3	4.5
		Drystock	949	218	23.0	23.0	2.1
		Dairy Drystock	8	1	12.5	12.5	2.0
		Dairy Herds	270	9	3.3	3.3	4.9
		Miscellaneous	772	125	16.2	16.2	0.7
		All	2778	433	15.6	15.6	2.7
	Deer	Breeding	138	15	10.9	10.9	2.9
		Fattening	6	0	0.0	0.0	4.3
		Miscellaneous	11	0	0.0	0.0	4.9
		Safari Park	3	1	33.3	33.3	0.0
		Velvetting	12	2	16.7	16.7	2.3
		All	170	18	10.6	10.6	3.0
	Both		2948	451	15.3	15.3	2.7

**Table 8: Spatial referencing by NLDB herd types**

Region	Spatial reference	Number of herds with spatial reference	Median size of farms	Average size of farms	Number of herds with 'months since last episode' data	Median number of months since last episode	Average number of months since last episode
Hawke's Bay	Present	2466	242	366	2480	5	9
	Missing	211	0	63	224	9	12
Wairarapa	Present	1696	67	771	1765	3	4
	Missing	285	4	71	451	5	6



HS00092

Map Sheet

V19 V20 W19 W20

Location

Phone	
Fax	
Number on Gate	344
Total Hectares	511.30

I wish to donate \_\_\_\_\_ dollars

☐ to the Helicopter Trust in my area  
☐ to the following Helicopter Trust: \_\_\_\_\_  
☐ I do not wish to make a donation at this time  
(Please tick one)

My payment is by:

☐ Cheque (Please enclose)  
☐ Credit card: VISA MC Other \_\_\_\_\_

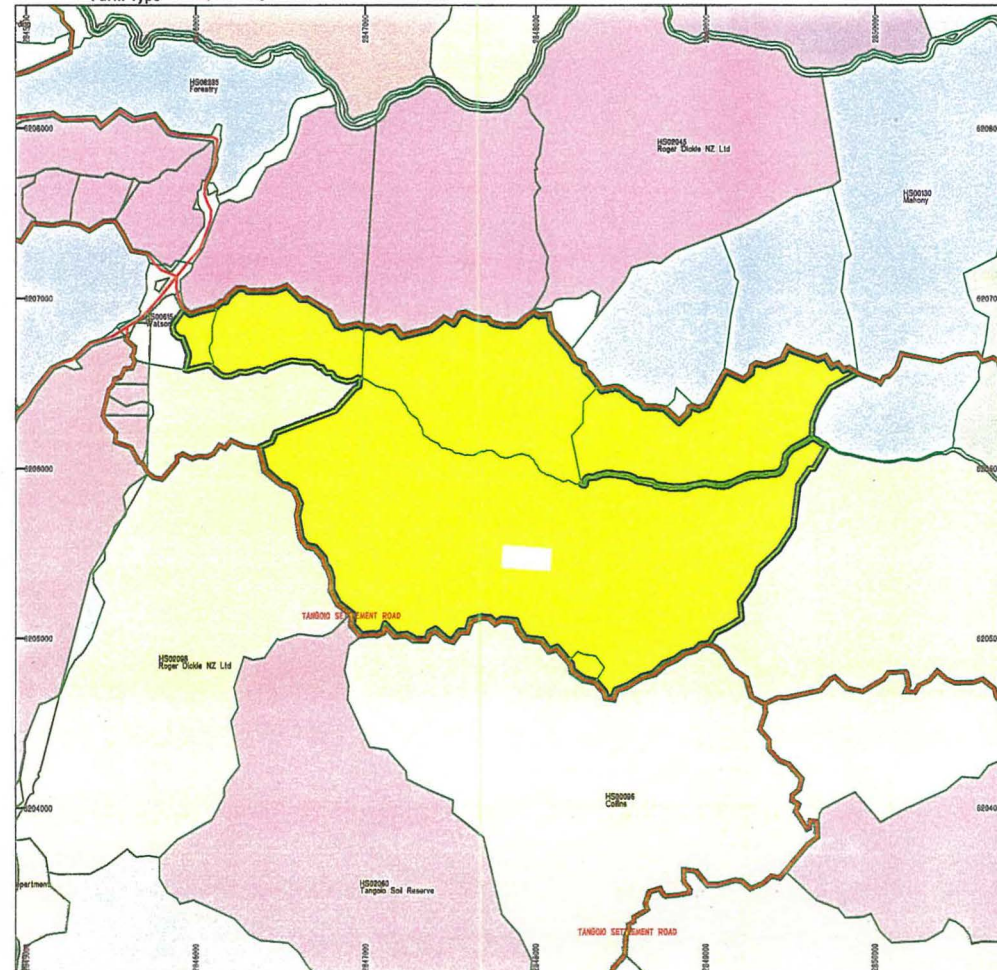
Number | | | | | | | | | | | | | | | |

Expires \_ / \_ Card Holder Name \_\_\_\_\_

(Signature) \_\_\_\_\_



**FEDERATED  
FARMERS**



<u>Animals</u>	<u>Number</u>	<u>Plants</u>	<u>Area (ha)</u>
Beef	400	Cereals	_____
Bee Hives	_____	Cropping	_____
Dairy	_____	Flowers	_____
Deer	_____	Forestry	1
Dogs	12	Fruit/Nuts	_____
Goats	300	Native Bush	_____
Horses	12	Seeds	_____
Pigs	_____	Vegetables	_____
Poultry	_____	Viticulture	_____
Sheep	3000	Other	_____

(please specify)

I understand that this information is stored in Agribase to be used by rescue services in rural emergencies and by MAF and allied organisations for managing responses to diseases, pests, residues, environmental quality issues or other problems that may limit New Zealand's productivity or ability to trade.

Signed \_\_\_\_\_

☐ Please tick this box if you do not want your farm to be located on any published farm location maps.

Published from Agribase, subject to Crown Copyright



## Chapter 3

### Spatial data and disease surveillance

## Introduction

Spatial information is useful in surveillance for diseases that involve vectors or environmental risk factors. The bovine tuberculosis (TB) control program in New Zealand deals with herds and vectors of TB that are intimately associated with geographical factors. This chapter considers in what way spatial data can enhance the management of TB surveillance. The type of spatial data (point or polygon) and level of completeness that is required is assessed.

TB testing is used to confirm that a herd is TB free and to detect TB in vectors. Wildlife in most areas of the country is free from TB. Cattle and deer are used as sentinel animals. TB testing is intensified in areas where TB is known to occur in wildlife and on the fringes of these areas. Spatial data can be used to visualise the spread and intensity of sentinel surveillance. The various types of spatial data available are shown in Figure 3. In the figure there are discrete point themes and area themes. The data layers that could be useful in TB control management include:

- scanned topographical maps;
- possum control information at farm level;
- possum control areas;
- TB testing and slaughter data at a farm level;
- livestock demography (such as herd types or the presence of sheep);
- and categorised land cover images.

From the above layers it is possible to detect areas of inadequate coverage of TB surveillance and to assess their importance.

## Method

Data was retrieved from the NLDB and from Agribase. It was manipulated in MS Access and linked to the DCDB in Arcview. Additional data on the slaughter of cattle and deer was obtained by a survey of farms in the Wairarapa and Hawke's Bay Regions. The data was explored using Arcview and (with permission) an extension of Arcview called Spatial Analyst. The statistical software Egret was used for logistic regression.

TB surveillance information was displayed and analysed using point and area data. The two data types were compared. The sensitivity of the methodology to missing spatial data was examined. TB testing data for deer and cattle in the year ending August 1997 was examined in a study area comprising four NZ 260 series maps. Each map sheet covers an area 40 kilometres wide and 30 kilometres high. The area chosen contained a vector risk area, a fringe area, and a surveillance area. A vector risk area is an area in which wildlife are thought to have TB. Such areas are surrounded by fringe areas. A *surveillance* area for the purposes of the TB control program is an area where feral animals are believed not infected with TB.

## Results

### Detecting classification errors

Area data at the farm level from Agribase was used to examine textual area classifications in and around the Waipawa vector risk area. Figure 4 shows the area classification around the Waipawa vector risk area. Classification errors can be seen, in that some herds in the fringe area are shaded red and they should be yellow. Area data may contain non-contiguous landparcels. These land parcels may appear to be classified incorrectly. This is the case with the small property shaded red in the top right hand



corner of Figure 4. If the boundaries of administrative areas are known then point data is better at disclosing classification errors. This is because the point normally relates to the home farm.

### **Representing the extent and completeness of farms**

Figure 5 shows the distribution of the different cattle herd types. This information is essential when considering surveillance issues. The light-green areas represent farms that have fattening cattle rather than breeding cattle. In these herds slaughter surveillance is relied upon and testing is not routinely carried out. Figure 6 shows the same data using points. No appreciation of the extent of different herd types can be obtained from the point data. Nor is there any appreciation of the level of missing data. Figure 7 shows a similar comparison for deer herds.

Figure 8 shows the cattle TB testing coverage in a surveillance zone over a 3-year period and a one-year period. This allows the extent of the TB surveillance to be examined. Breeding cattle herds in the surveillance zone are tested every three years. It is expected that in any one year the herds tested will evenly cover the whole region. The system is called “every third farm” testing. In this case, when testing was summed for 3 years, the coverage of the program is well distributed with approximately 70% of the land area covered. However on an annual basis, the “every third farm” policy has not achieved an even coverage and testing appears to be clustered. Using the area data theme in programming testing would allow the “every third farm” concept to be more consistently applied.

Area data allows the calculation and display of stocking rates. Figure 9 presents testing density displayed using different colours to indicate intensity. Testing density was calculated as the number of tests per square kilometre per annum. The comparison using point data is shown in a variety of ways in Figure 14, 15 and 16. In Figure 14 colours have been used to indicate the number of tests. In Figure 15 different sized dots were used to display this information. These figures do not convey the information as well as Figure 9. In Figure 16 point data was converted into a surface through interpolation. Contours were created and splines were used to generate a smooth surface of testing density. Each of these methods conveyed a false impression of true spatial coverage of TB testing. The testing density surfaces do not communicate the surveillance information as well as the shaded area data in Figure 9. The enclosed figures show that surveillance information at a farm level is immediately more satisfying when displayed with area data than with point data. Gaps in the spatial coverage of testing can easily be distinguished from gaps resulting from missing data.

### **Determining the distance from or proximity to other features**

Figure 7 shows the distribution and number of deer displayed with using area features and point features. The point coverage communicates less information than the area coverage. The area coverage shows the extent of the land used by deer farmers. This may be important if you wish to know the about deer farms that bound a river or forest.

Careful examination of the area theme in Figure 7 reveals that 9 deer farms are adjacent to a river. The point data theme does not reveal this information.

The GIS can generate a buffer of a defined width around each property. This could be set to represent the expected home range of possums or other TB vectors. This approach is useful in the Vector risk area. The buffer area can represent the extent of indirect sampling of the underlying possum population through cattle TB testing. Ideally, in a vector risk area there would be no land farther than the expected home range of possums from any TB testing activity. Figure 10 shows this assessment using area data. To the

right of the figure shaded in blue are areas that are more than 1600 meters from any testing activity. This part of the figure is an area that only tests every three years. The left of the figure is on an annual testing program. The distance from 'farms that test' can be used to show gaps in the surveillance program. Point data can not be used to detect gaps in the same way.

### **Determining the cause of gaps in surveillance**

There are a number of reasons why a gap may appear in surveillance. Intense sheep farming can create significant gaps in the spatial coverage of the testing program. Figure 8 shows three-year-testing data, the gaps are predominantly caused by sheep farming. Other gaps in the spatial coverage of testing may be the result of the predominance of fattening rather than breeding herds in an area. The gap in surveillance may disappear if, as well as testing surveillance, slaughter surveillance is included. If there is still a gap, the significance can be determined by examining the amount of vector control in the area, the type of land cover, and the proximity to TB in vectors or TB in domestic animals. Figure 11 shows the Waipawa vector risk area. Gaps in the spatial coverage of TB testing are shown in red. The gaps are associated with fattening herds (shown in yellow in Figure 21). There are also some land-parcels that are not allocated to a farm and some herds that lack a spatial reference. The extent of missing spatial referencing can be assessed by comparing "unlinked" testing numbers to "linked" testing numbers. Table 10 shows details of what testing information for the area concerned was not shown on the map. In this case 2 % of the testing data was missing.

Figure 12 shows the type of vector control used in and around the vector risk area for the 1997 season. The vector control data available includes the hours worked, the number of possums killed and the control method used. Figure 12 shows the hours of vector control activity per square kilometre (100 hectares). Gaps in the vector control activity are shown in red in Figure 13. The significance of the apparent gap in surveillance in Figure 11 is greater in that it coincides with an apparent gap in pest control. The gap in vector control presented at the bottom right of Figure 13 is an artefact of the time period for which the data was retrieved. Vector control data is available at a farm level because the Hawke's Bay Regional Council uses Agribase as the basis of its' pest management recording system.

Area data allows the calculation of the number of tests per annum per unit area. Table 9 shows the testing density for herds in Hawke's Bay. The testing density is shown for deer farms, cattle farms and mixed deer and cattle farms. In areas such as Waipawa, where TB vectors are present, the testing is conducted annually and young stock are eligible for test. Deer herds in the annual testing areas (Fringe and Waipawa) have very high testing densities. They have 4 to 5 times the test intensity compared with neighbouring cattle herds. Figure 22 shows surveillance intensity in the Waipawa vector risk area. Note that the testing intensity of the deer herds is consistently high. From the point of view of using deer and cattle as sentinel animals an even testing intensity throughout the area is desirable. If the testing program was based on test intensity, deer herds could have the same testing intensity as cattle herds. As the deer in this area are grazed more intensively, and have non-specific reactivity problems, it could be argued for a reduction in the testing in these herds. In the Waipawa area only 1.6% of deer tests and 2.6% of cattle tests were not represented in the analysis. This gives confidence in data presented in Table 9 and in the figures representing testing.

TB testing represents one part of the surveillance in an area. Before adjustments to the program are made, consideration needs to be given to the amount of slaughter surveillance in an area.



**Table 9: Crosstabulation of tests per hectare per year by zone and type of farm for the Hawke's Bay**

Zone Code	Purely deer farms			Purely cattle farms		
	Number of farms	Tests per Ha year	Average size Ha	Number of farms	Tests per Ha year	Average size Ha
Surveillance	56	1.62	50	1471	0.39	382
Fringe	6	2.36	68	188	0.36	550
Waipawa	6	3.87	40	284	0.71	213
Total	68	1.88	51	1943	0.43	373
	Mixed cattle and deer farms			All farms		
Surveillance	180	0.74	711	1707	0.47	403
Fringe	47	0.77	550	241	0.50	545
Waipawa	54	1.19	376	344	0.87	237
Total	281	0.83	620	2292	0.53	393

**Table 10: Summary statistics of tests not spatially referenced by species and zone**

Zone Code	Deer tests included	Deer tests not included	Percentage deer tests not included	Cattle tests included	Cattle tests not included	Percentage cattle tests not included
Surveillance	17909	1565	8.0	79099	1928	2.4
Fringe	6588	112	1.7	30915	355	1.1
Waipawa	13291	222	1.6	28994	779	2.6
Total	37788	1899	4.8	139008	3062	2.2

### Assessing missing spatial data

The sensitivity of the above techniques to missing spatial references was examined. Figure 18 shows that the image loses its contiguity as the level of missing data increases. False gaps in testing activity showed up when less than 85% of herds had a spatial reference. In order to obtain a satisfactory depiction of testing levels in an operational area the results showed that at least 85% of herds must be linked to their land parcels. Figure 17 shows that missing spatial references are hard to detect when point data is used.

### Slaughter surveillance

Inspection at routine slaughter represents an important component of surveillance for diseases. Currently positive TB findings at slaughter are reported, but negative information is not being utilised. In some areas large numbers of cattle are slaughtered. To obtain a fair impression of the actual TB surveillance in an area, both, testing and slaughter data need to be assessed. There is no record in the NLDB of the number slaughtered from a property in a given year. Addition of this information would represent a major yet simple enhancement to the NLDB.

To obtain information on slaughter surveillance, a survey of farmers in the Hawke's Bay and Wairarapa regions was conducted. The survey is presented in Appendix 1. The aim was to develop a tool for prediction of the number slaughtered per annum from individual herds, using data already available in the NLDB. It was hoped that existing data currently recorded for individual herds could be used to estimate the number

slaughtered. Using the survey data a Poisson regression model was constructed that predicted the number slaughtered in individual herds. The model was run and the estimated number slaughtered from each farm was mapped in much the same way as the testing data. The variables from the national livestock database which were used in this analysis included: Region, herd-type, testing program, and testing data by stock class. When a herd is tested the number tested in each stock class is recorded. Stock are classified as being male or female and one of three age classifications: rising 1 year, rising 2 year or mature.

Logistic regression was used to estimate the number of animals purchased and the number of animals slaughtered from each farm in the region. The model explained only 50 percent of the variation in number slaughtered or purchased in the study herds. The goodness-of-fit of the model was assessed using the R squared estimate which was calculated based on the deviance of the intercept-only model and the deviance of the final model (Mittlbock M and Schemper M, 1996). Recording the actual number slaughtered or purchased from a farm would be preferable to using the model. The model predictions were combined with testing data and used in this chapter to display total TB surveillance.

The final Poisson regression models for prediction of the number of cattle and deer slaughtered are presented in Table 11 and Table 12. It includes the two categorical variables region (with two levels) and herd-type (with 5 levels). Stock and testing numbers are represented in units of 100. There was a statistically significant interaction between region and herd type. The expected count for each farm was calculated using NLDB data based on the following formula:  $E(Y_j) = \lambda_j = \lambda_0 \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)$ . Where  $\beta_1, \beta_2, \dots, \beta_n$  are the coefficients corresponding to the predictor variables  $x_1, x_2, \dots, x_n$ , and  $\lambda_0$  represents the baseline hazard rate when  $x_1, x_2, \dots, x_n$  are zero. The resulting statistical models for the number of animals purchased and the number slaughtered are presented in Table 11 and Table 12 respectively. Both models explain about 45% of the deviance in the data. This suggests that the predictive capacity of the regression models is not sufficient to allow reliable prediction of number of cattle purchased and slaughtered on individual properties.

The expected number slaughtered for each farm was calculated and visualised. Testing and slaughter surveillance was combined to obtain an overall impression of TB surveillance in a given area. Slaughter surveillance “fills in” most of the gaps in the testing program. The combination of slaughter surveillance and testing in some areas is such that a reduction of the amount of testing in the area could be considered.

Figure 20 shows the surveillance achieved by the testing program and the total coverage gained by tests and slaughter. Figure 11 shows the Waipawa vector risk area and a gap in the testing coverage in the centre of it. Figure 21 shows that the reasons for the gap in testing include the presence of a large number of fattening herds and one or two farms with missing spatial references. Figure 21 also shows that when slaughter surveillance is included (estimated using the above model) the gap disappears.

The inclusion of estimated slaughter data demonstrates the benefits that could be obtained from using real slaughter data. The ability to examine surveillance intensity allows the future testing to be programmed at a farm level rather than a control area level. It is recommended that slaughter data for each herd, be recorded in the NLDB each year by stock class, along with the total number of cattle and deer present, purchased and born.

**Table 11: Final Poisson regression model for prediction of the expected number of cattle slaughtered**

Regression term	Coefficient	Standard Error	P-Value	Rate ratio
Intercept	3.238	-0.0284	<.001	25.49
Region = Wairarapa / Hawke's Bay	-1.035	-0.0427	<.001	0.3551
Herd type = Beef Breeding	1.119	-0.0269	<.001	3.062
Herd type = Drystock	1.347	-0.0292	<.001	3.844
Herd type = Dairy	0.9617	-0.0403	<.001	2.616
Herd type = Miscellaneous			<.001	1
Program (annual or triennial testing)	-0.1102	-0.0063	<.001	0.8957
Number Tested 3 and 2 years ago 100s	0.0371	-0.0011	<.001	1.038
Number of sheep 100s	0.0061	-0.0001	<.001	1.006
Number of deer 100s	0.0987	-0.0023	<.001	1.104
Cows tested that year 100s	-0.0653	-0.0054	<.001	0.9368
Rising 2 year males tested that year 100s	0.4686	-0.0040	<.001	1.598
Rising 1 year cows tested that year 100s	0.3983	-0.0149	<.001	1.489
Rising 1 year males tested that year 100s	-0.136	-0.0118	<.001	0.8729
Other Cows tested 100s	-0.3921	-0.0266	<.001	0.6757
Bulls tested that year 100s	0.3068	-0.0330	<.001	1.359
Rising 2 year cows tested that year 100s	-0.3006	-0.0158	<.001	0.7404
Other Males tested that year 100s	0.131	-0.0195	<.001	1.14
Wairarapa *Herd type Beef Breeding	0.3433	-0.0442	<.001	1.41
Wairarapa * Herd type Drystock	-0.259	-0.0463	<.001	0.7719
Wairarapa * Herd type Dairy	0.5424	-0.0600	<.001	1.72
Region. Herd type = Beef Breeding			<.001	1
Deviance of intercept-only model	129129			
Deviance of current model	70670			
R squared	0.45			

**Table 12: Final Poisson regression for prediction of the number of deer slaughtered**

Poisson regression term	Coefficient	Standard Error	P-Value	Rate ratio
Intercept	2.996	-0.0709	<.001	20
Mature hinds tested that year 100's	0.3139	-0.0097	<.001	1.369
Number Tested 3 and 2 years ago 100s	0.1588	-0.0058	<.001	1.172
Rising 1 year hinds tested that year	-0.9513	-0.1140	<.001	0.3862
Rising 2 year hinds tested that year	2.092	-0.0719	<.001	8.099
Herd type = Miscellaneous	1.147	-0.0773	<.001	3.147
Herd type = Velvetting	0.0032	-0.0635	0.96	1.003
Herd type = breeding			<.001	1
Number of Sheep 100's	-0.00647	-0.0005	<.001	0.9936
Region = Wairarapa	-0.2025	-0.0419	<.001	0.8167
Region Wairarapa.Herdtype Miscellaneous	-0.7546	-0.1060	<.001	0.4702
Region Wairarapa.Herdtype Velvetting	0.3871	-0.1060	<.001	1.473
Region Wairarapa.Herdtype=beef breeding			<.001	1
Rising 2 stags tested that year 100s	-0.2685	-0.0604	<.001	0.7645
Program (annual or biennial)	0.1948	-0.0328	<.001	1.215
Rising 1 year stags tested that year 100s	-0.5092	-0.1180	<.001	0.601
Deviance of intercept-only model	9878			
Deviance of current model	5616			
R squared	0.43			

## Assessment of clustering

The type and completeness of data required to detect clustering was determined using TB as an example. A cluster of infected farms may indicate that a population of infected vectors is present in an area. Detecting clustering where a feral vector is involved is complex. Some assessment must be made of the likely size of the area occupied by the infected vector and hence the expected size of a cluster.

In 1988, a population of possums at Waipawa in Hawke's Bay became infected with TB. This resulted in TB cases in domestic deer and cattle in the area. The cases first started to occur in 1991 and by May 1992 a control operation commenced following the diagnosis of TB in possums (Mackereth, 1993). Cluster detection techniques were applied to this disease episode to assess their usefulness for detecting and locating clusters. The methods were applied at different levels of data completeness to determine the sensitivity of the cluster detection methods to missing data.

An area of approximately 30km by 40km centred on the known cluster of cases at Waipawa was selected. There has been 36 TB infected herds in that area over the last 8 years. A random selection of 36 control farms not on movement control was made and compared with the cases. This was done with area and point data and is shown in Figure 23 and Figure 24. Figure 23 shows the locations of a cluster of TB infected herds and a random selection of control farms. Figure 24 shows the farms at Waipawa (note that they are naturally clustered at townships and along roads). The cluster could be visualised using a neighbourhood technique shown in Figure 25.

Of the cluster detection methods that use area data, Moran's I (Cuzick and Edwards, 1990) was chosen. Moran's I is a test for spatial autocorrelation in disease rates. The calculation of Moran's I requires a contiguity file. This file defines the spatial proximity relationship between farms. Contiguity files were created by considering all farms within a given radius of a farm as contiguous. The contiguity matrix was also calculated based on two levels of spatial data completeness (90% and 63%). Moran's I was calculated using point data and area data.

Table 13 shows the Moran's I values and their significance for area and point data at different levels of data completeness. The results show that clustering was detected more readily with area data than point data. Clustering was detected when using area data at all levels of missing data and at buffer widths of 100m, 5000m and 7000m. Point data does not reflect contiguity as well as area data.

**Table 13: Moran's I values for area and point data in the Waipawa vector risk area at three levels of proximity**

Data type	Immediate neighbours			5 km radius			7 km radius		
	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours
<b>Area 100%</b>	<b>0.079</b>	<b>0.0003</b>	5.8						127
<b>Point 100%</b>	-	-	1.1	0.017	0.004	56	<b>0.02</b>	<b>0.00002</b>	92
<b>Area 70%</b>	-0.001	0.9	4.2	0.012	0.061	53	<b>0.018</b>	<b>0.001</b>	81
<b>Point 70%</b>	-	-	1.1	0.002	0.64	40	<b>0.012</b>	<b>0.044</b>	65

The search radius around a point encompasses fewer farms than around an area feature. To compensate for this effect, a larger search radius should be used with point data. Increasing the radius does not necessarily include the same farms in the contiguity file as using area data, but it does increase the number of farms considered to be contiguous. The extent, by which the radius should be increased, when switching from areal data to point data, is of interest to all spatial processes that use distance methods. The increase

should result in a similar number of contiguous farms as area data. Two approaches were applied to investigate this problem.

With the first approach, a hypothetical radius was calculated for each farm, based on its area. The farm was considered to be circular for this purpose. To adjust the distance criteria when switching from area to point data, the average radius was doubled. In some processes it may be appropriate to add the actual "radius" of a farm plus the average radius of all farms. For the area under study the mean radius was 708 meters with a 43-meter confidence interval. The median was 696 meters. Figure 27 shows the residual difference between the mean radius and the actual radius for each farm.

The second approach was to plot the average number of neighbours against distance for point and area data. Figure 28 contains this plot. The increase in distance required to achieve the same number of neighbours can be read off the resulting graph.

The Cuzick and Edward's test uses distances between cases and controls. The results are shown in Table 14. They indicate that clustering was detected when 90% or more of the cases were included and when considering 8 nearest neighbours. When a summarised probability was calculated for 1 to 10 nearest neighbours, clustering occurred at 90 and 80 % data completeness, but not at 100% or 70%.

**Table 14: Results from applying Cuzick and Edward's test for clustering (at 8 nearest neighbours) of cases at different levels of case data completeness**

Cases	Controls	Percent of cases used	k nearest neighbours	T(k) Sum of neighbours that are cases	E(T)	Z	p	Bonferroni P*	Simes P*
36	114	100%	8	82	67.7	1.62	0.052	0.52	0.12
32	114	90%	8	70	54.7	1.89	0.03	0.3	0.07
28	114	80%	8	59	42.9	2.2	0.139	0.12	0.02
25	114	70%	8	44	34.8	1.38	0.08	0.62	0.18

- Bonferroni P and Simes P are summarised probabilities for 1 to 10 nearest neighbours.

The Cuzick and Edward's test was applied at different levels of spatial data completeness. The results were inconclusive, probably as a consequence of the case control approach used.

### Recommended data type for the detection of clusters

Area data is superior to point data in cluster detection, as it accurately reflects contiguity. Cluster detection methods that use point data are not recommended. Point data could be used to assess clustering if a contiguity file was made based on distances that represented the cluster size, and an adjustment factor. The adjustment from area to point data was at least twice the hypothetical mean radius of the farms involved. Missing data did reduce the ability of the cluster detection methods to detect clustering. At least 80% spatial data completeness was required in this example to achieve an acceptable sensitivity at cluster detection.

## Discussion

Table 15 summarises the preferred type of spatial data for various TB management applications. Many of the functions are applicable to other diseases. The summary shows that area data is essential for most envisaged uses of the data. There were instances

where point data was useful, such as in the detection of classification errors. Point data is easily derived from area data. The results show that a surveillance system that uses spatial data should use area data. Point data has limited value on its own. Contiguous properties are shown clearly when area data is used. Point data will not represent contiguous properties to the same extent as area data. There are a number of methods for classifying farms as near or far away. It is possible to compare the actual neighbours seen in an area theme with those classified as neighbours in a point theme. If the size and shape of farms varies a lot, then the sensitivity and specificity of the classification using a point theme will be low. For disease control purposes it is sometimes desirable to examine the proximity of farms to a certain feature, such as bush or rivers. Area data was clearly an advantage in these circumstances. The assessment of the level of data completeness required to meet the needs of disease management was mostly done by visual appraisal. The effect on TB surveillance applications was tested by progressively reducing the level of data completeness until the image lost contiguity or a false impression was gained. It was clear that at 75% data completeness most of the methods were not applicable. False holes in the surveillance of TB were detected at 75% coverage, but not at 85%.

**Table 15: Table of the preferred data type for various TB control applications**

Likely Application	Preferred data type	Level of completeness required
Detecting misclassification errors	Point	85%
Representing the extent and completeness of farms	Area	85%
Determining the proximity to other features	Area	85%
Determining the cause of gaps in surveillance	Area	95%
Assessing missing spatial data	Area	NA
Assessment of clustering	Area	80%
Working with contiguous properties	Area	95%
Assessing TB surveillance	Area	85%

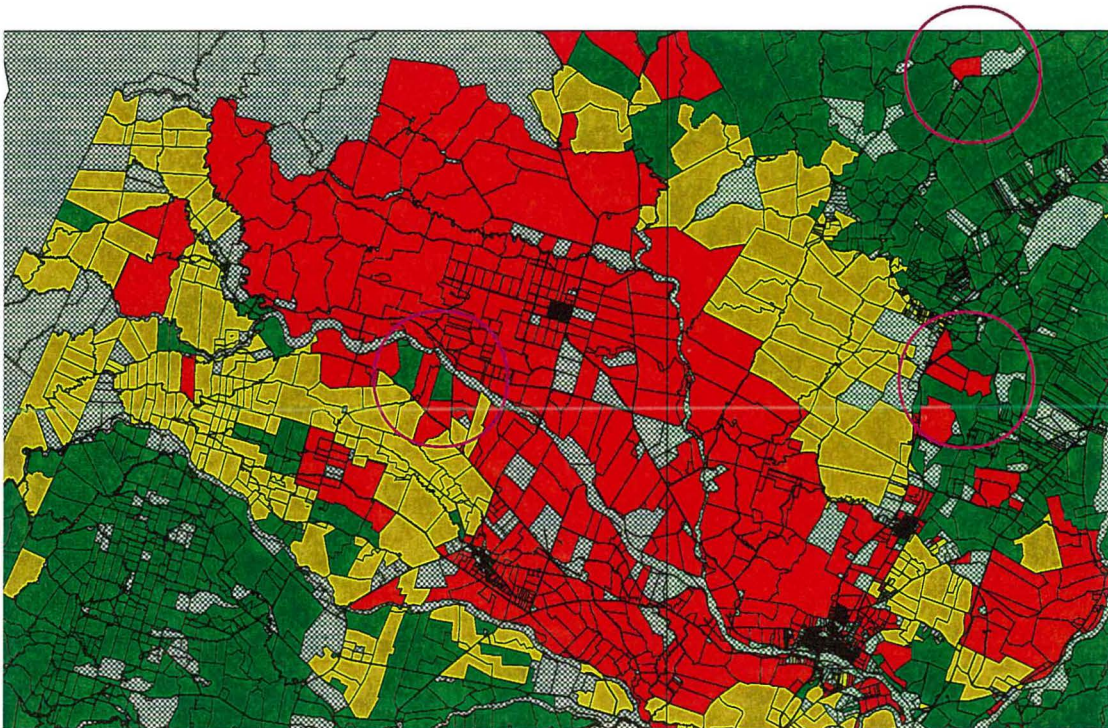




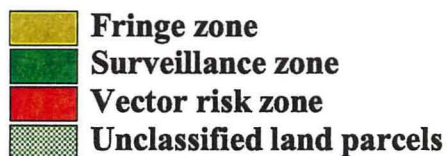


**Figure 4: Testing zones at the Waipawa vector risk area displayed as an attribute of farms.**

On the map below farms are shown as area features.  
The whole farm is shaded according to its testing zone.  
Note that some farms appear to be incorrectly classified.  
A small circle has been placed over some incorrectly classified farms.  
This is an easy way to detect classification errors in textual data.



**Testing zones have been shaded with the following colours**





**Figure 5: Cattle herd type distribution shown with area data**

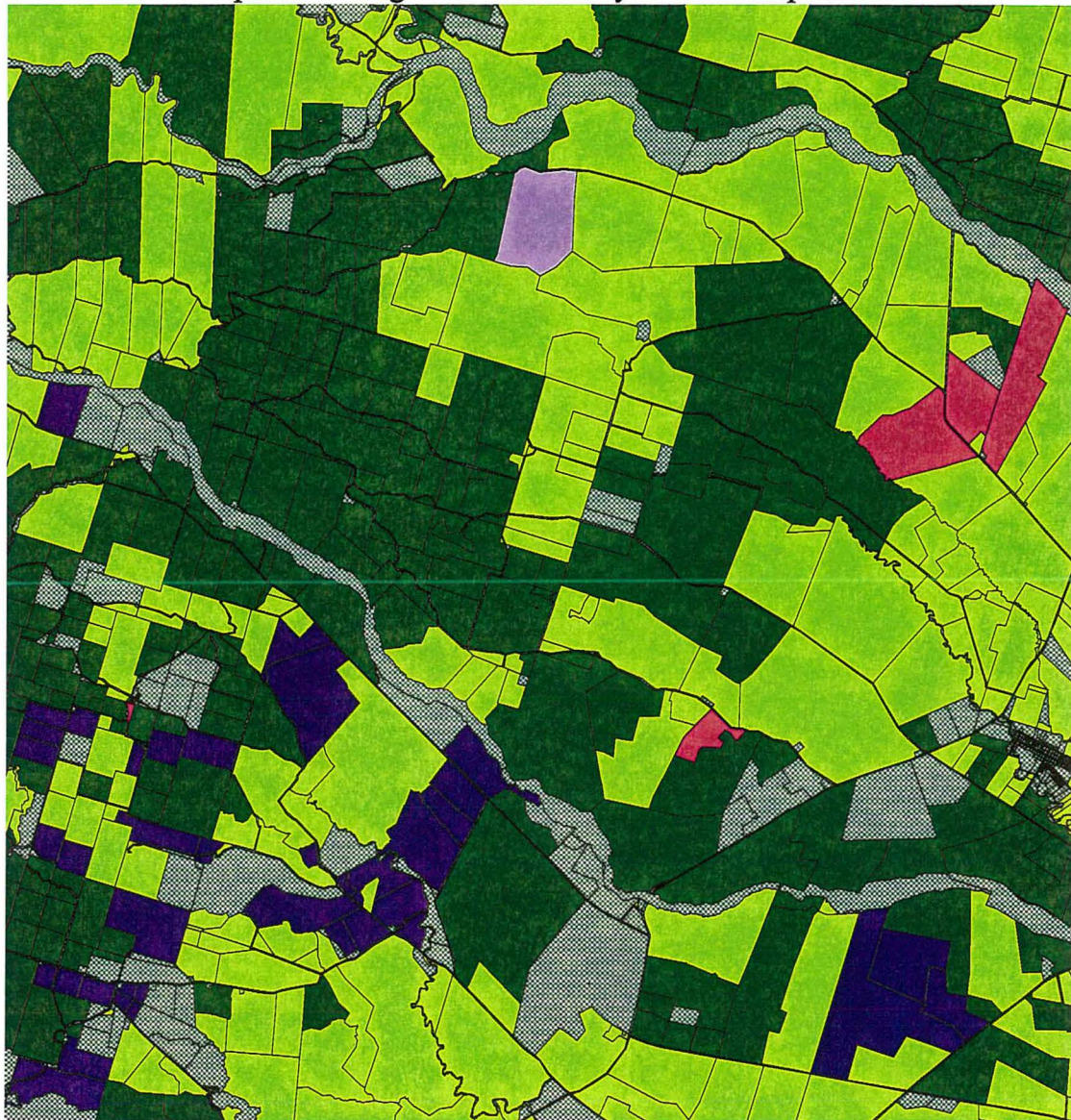
The area below is a small part of the Waipawa Vector Risk area.

Farms have been shaded according to their herd-type.

From the figure it is easy to see the extent of beef breeding herds (dark green) and beef drystock herds (light green).

The same attribute is displayed using point data to represent farms in Figure 6.

Note that the point coverage does not convey the same impression.



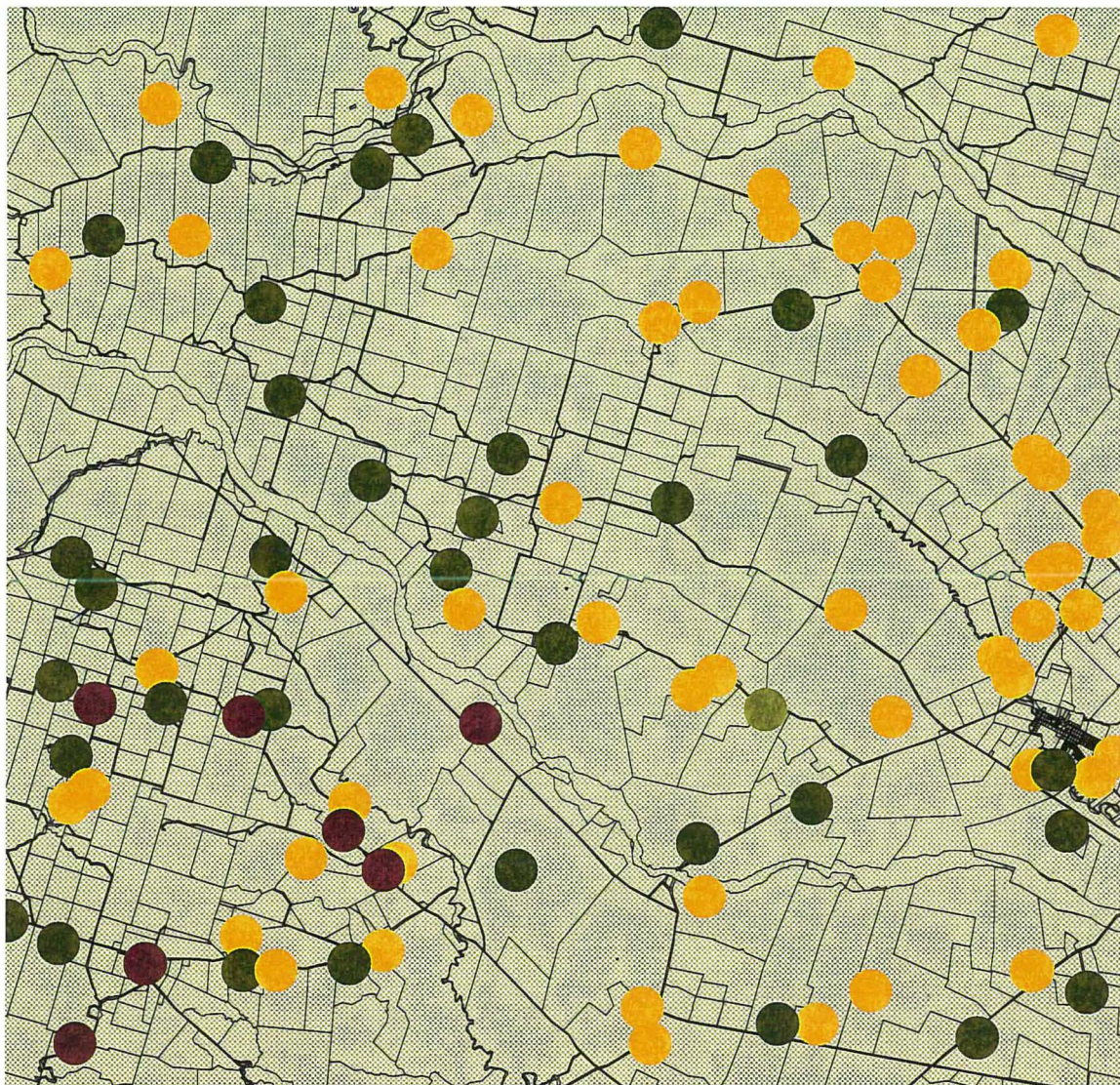
**Cattle herd type**

-  Beef breeding herds
-  Beef drystock herds
-  Dairy Dry herds
-  Dairy Herds
-  Miscellaneous herds
-  Unclassified land parcels



**Figure 6 Cattle herd type distribution shown with point data**

The area below is a small part of the Waipawa Vector Risk area. The area shown is the same as is shown in Figure 5. The farms in the area are displayed as points. The herd types on each farm are shown by the colour of the point. Compare this figure with Figure 5.



-  **Beef breeding**
-  **Beef Drystock**
-  **Dairy herd**
-  **Miscellaneous**

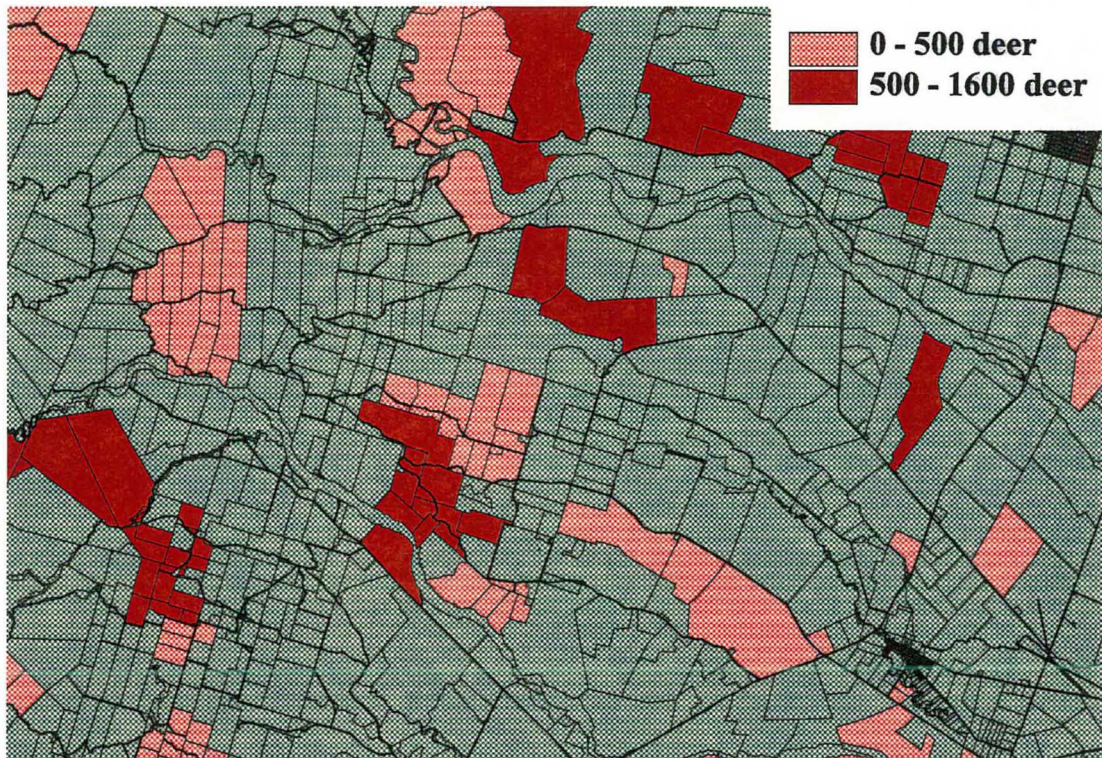


**Figure 7: The distribution and size of deer herds displayed with area and point data**

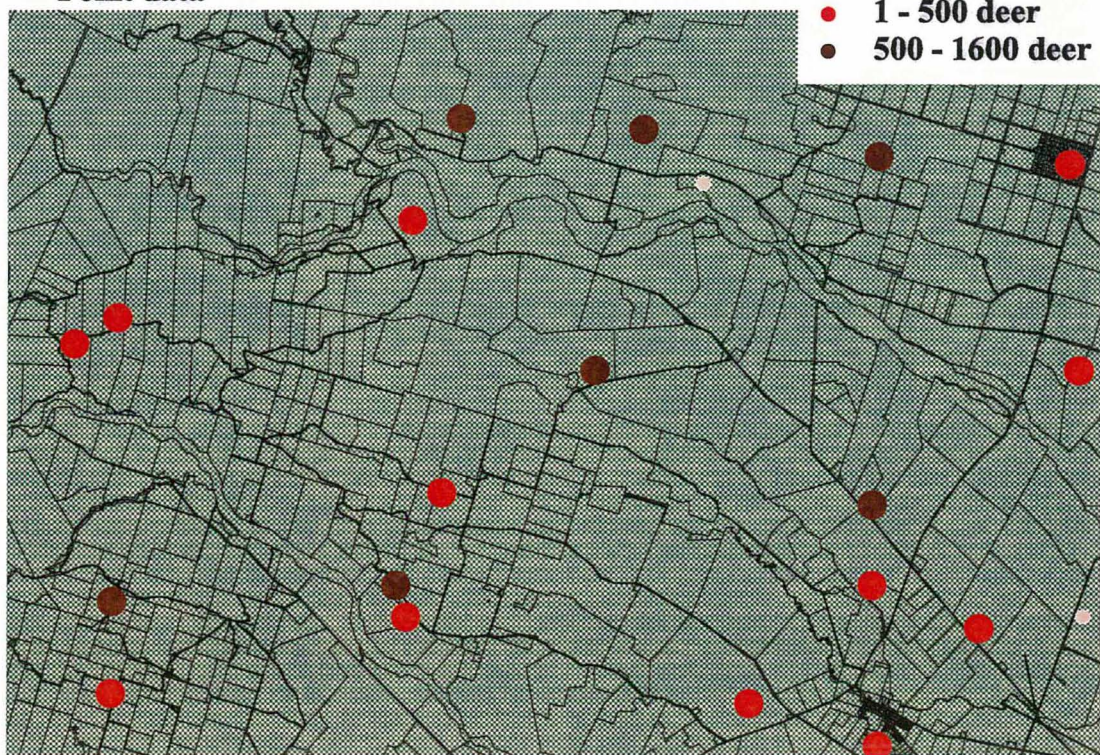
The area shown is part of the Waipawa Vector risk area.

This figure shows the difference between representing farms as area features and point features.

**Area data**



**Point data**

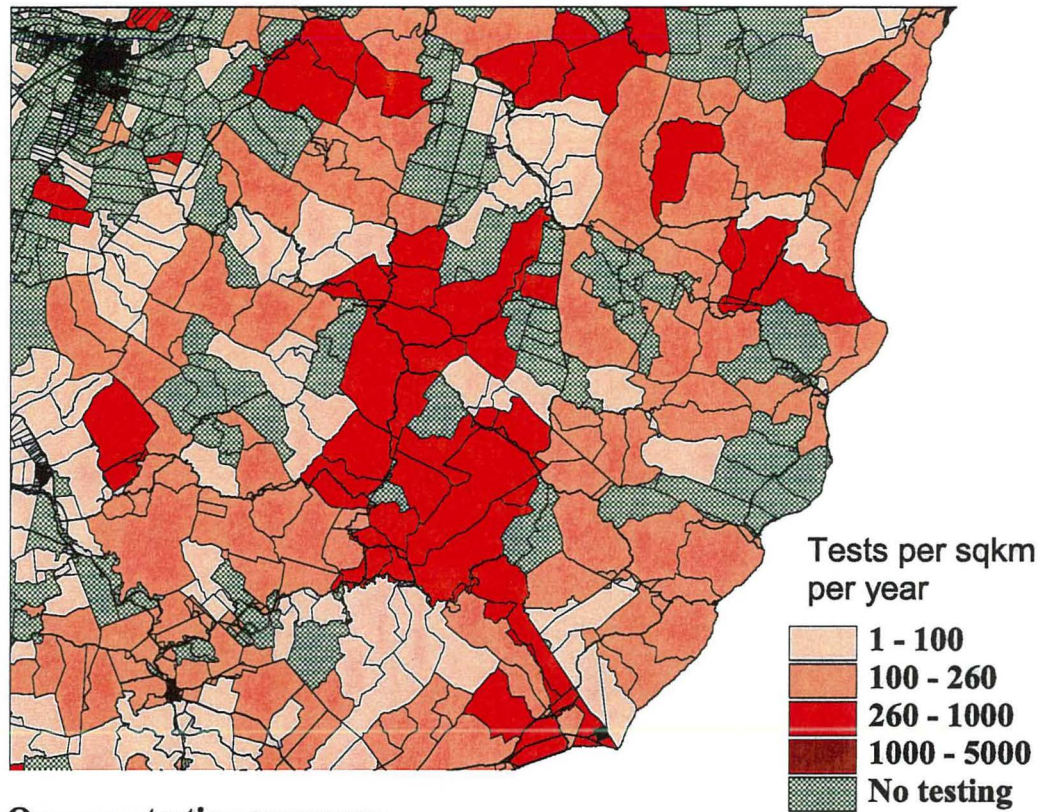




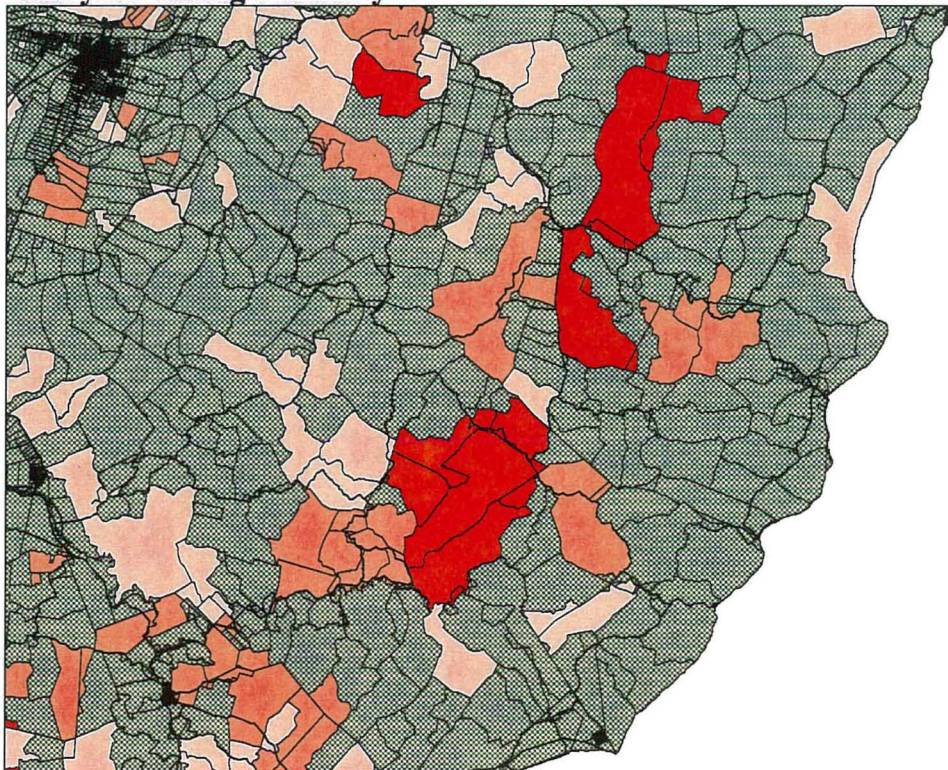
**Figure 8: Coverage of cattle TB testing in a one year and three year period in a surveillance area**

The figure shows the extent of testing and its distribution

**Three year summary**



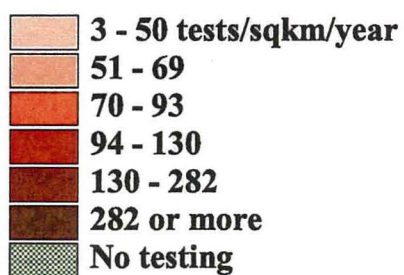
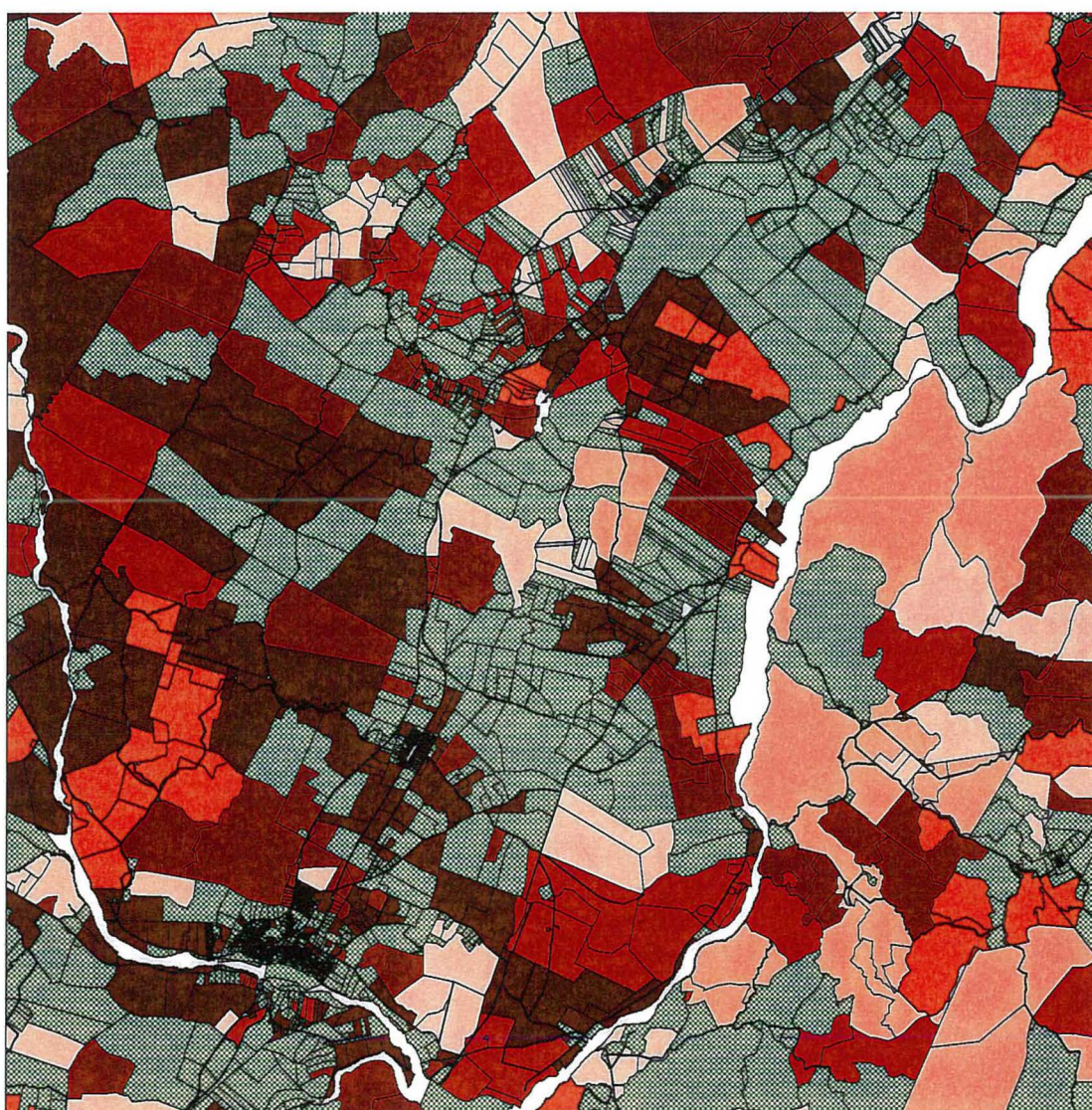
**One year testing summary**





**Figure 9: Three year summary TB testing densities for cattle shown using area data.**

The area shown is on the eastern edge of the Waipawa Vector risk area.  
The righthand side of the figure is a surveillance zone.  
The lefthand side is the vector risk area.  
TB Testing density was measured as TB tests per sqkm per annum.





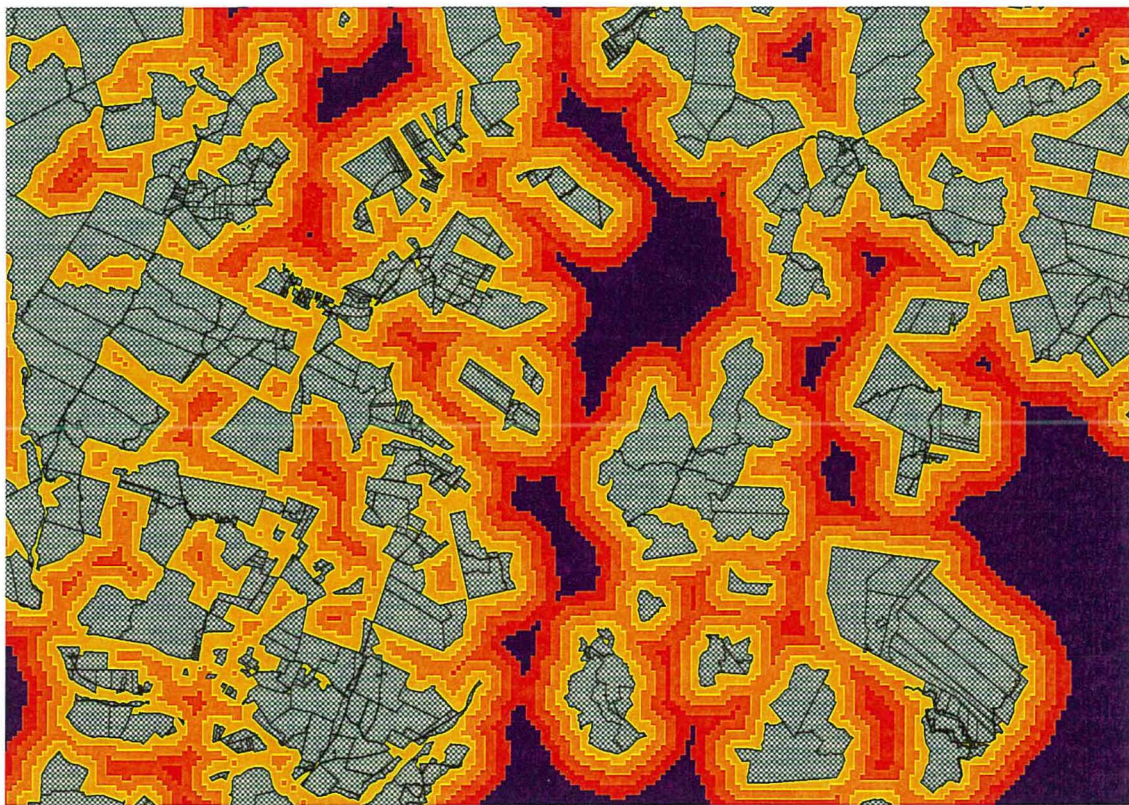
**Figure 10: Figure showing the distance to (or from) farms that test.**

**The area shown is a larger view of the area shown in Figure 9.**

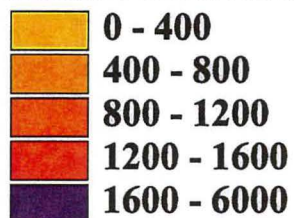
**The light green areas are farms that TB test.**

**The bands of yellow orange and blue indicate distance from farms that test.**

**Note that the blue areas are zones that are at least 1600 meters from farms that test.**



**Distance to a farm that TB test in meters**

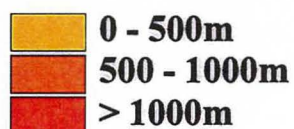
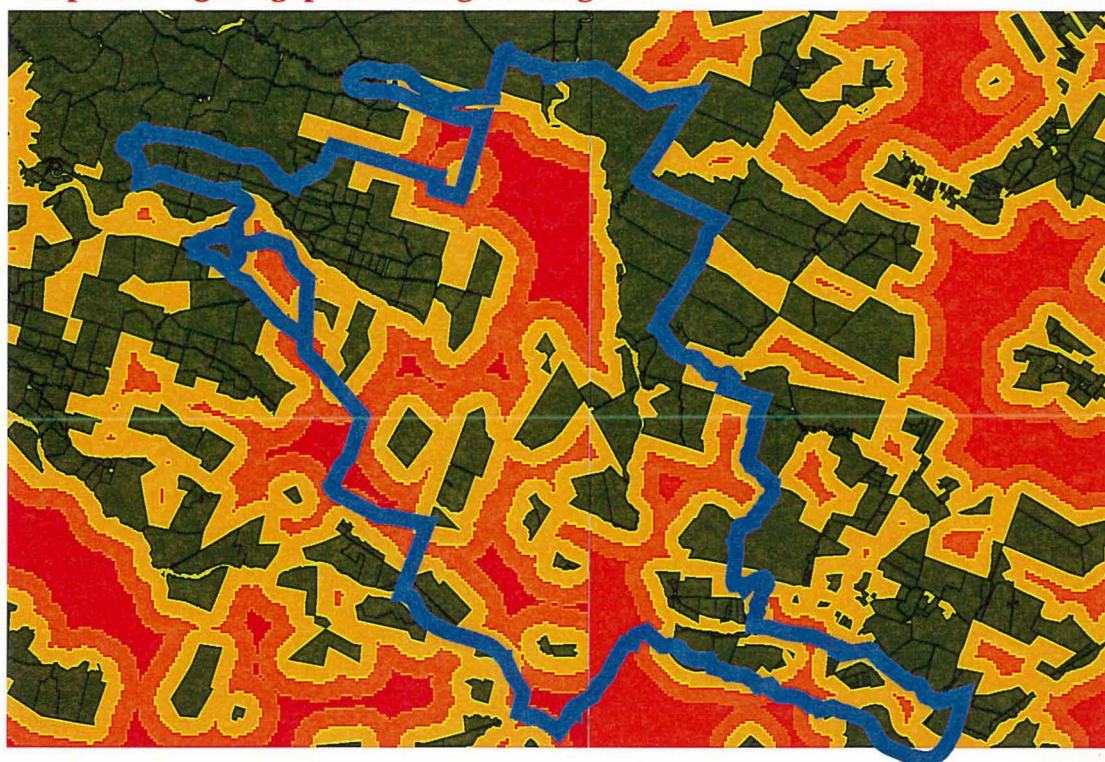




**Figure 11: Map of the Waipawa vector risk area showing gaps in the TB testing coverage.**

The blue outline indicates the vector control area.  
The green areas are farms which test cattle or deer.  
The red areas show land at least one km from testing activity.  
Note that if there are sentinel animals in the yellow orange and red areas they are not being tested.  
Ideally testing of sentinels should occur throughout the vector risk area.

**Map showing the gaps in testing coverage**

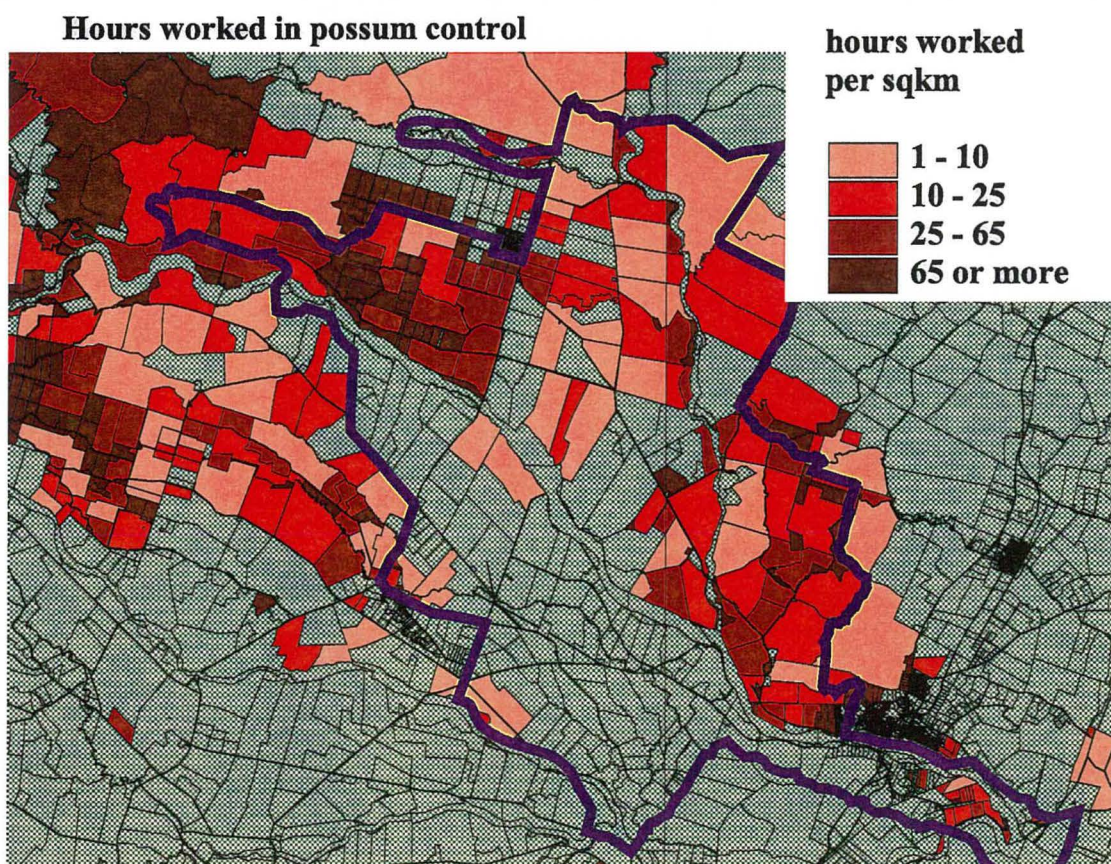
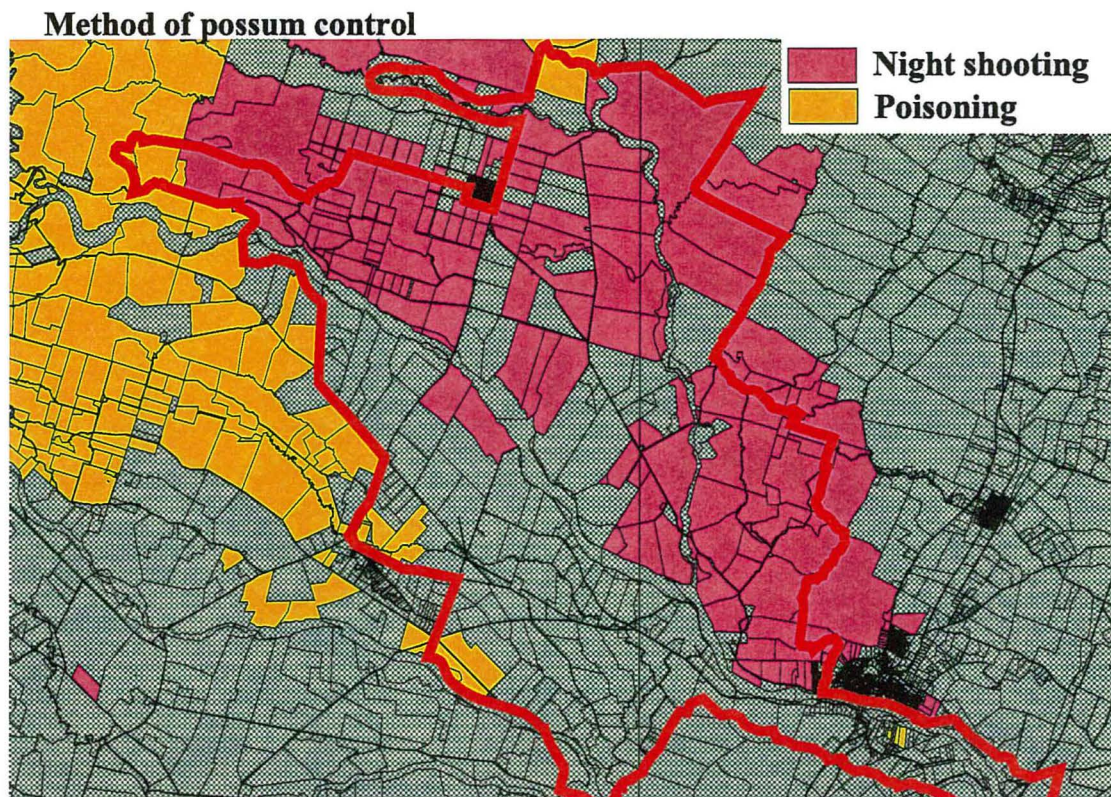


**Distance to farms that test.**

**The reasons for these apparent gaps in surveillance are explored in Figures 21.**



**Figure 12: Vector control methods and hours worked in the Waipawa vector risk area in 1997**



The green areas are areas where no vector control was undertaken!



**Figure 13: Apparent gaps in the coverage of possum control work in the 1996 1997 season**

**The areas which had no possum control activity within one km are shown in red**

**Investigation showed that the apparent lack of activity at the southern end of the vector risk area was due to the system of recording work done rather than a lack of activity**



**Distance in meters to possum control work**

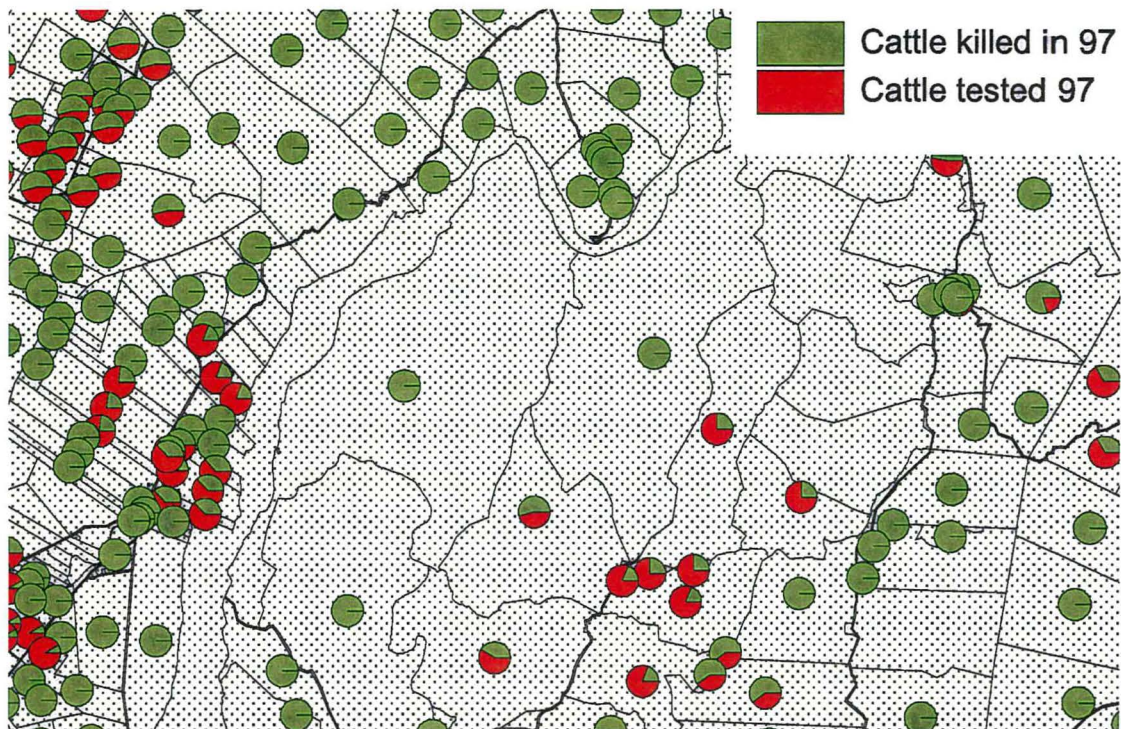
- 0 - 500
- 500 - 1000
- 1000 or more

- Vector risk area
- Possum control work  
96 -97 season

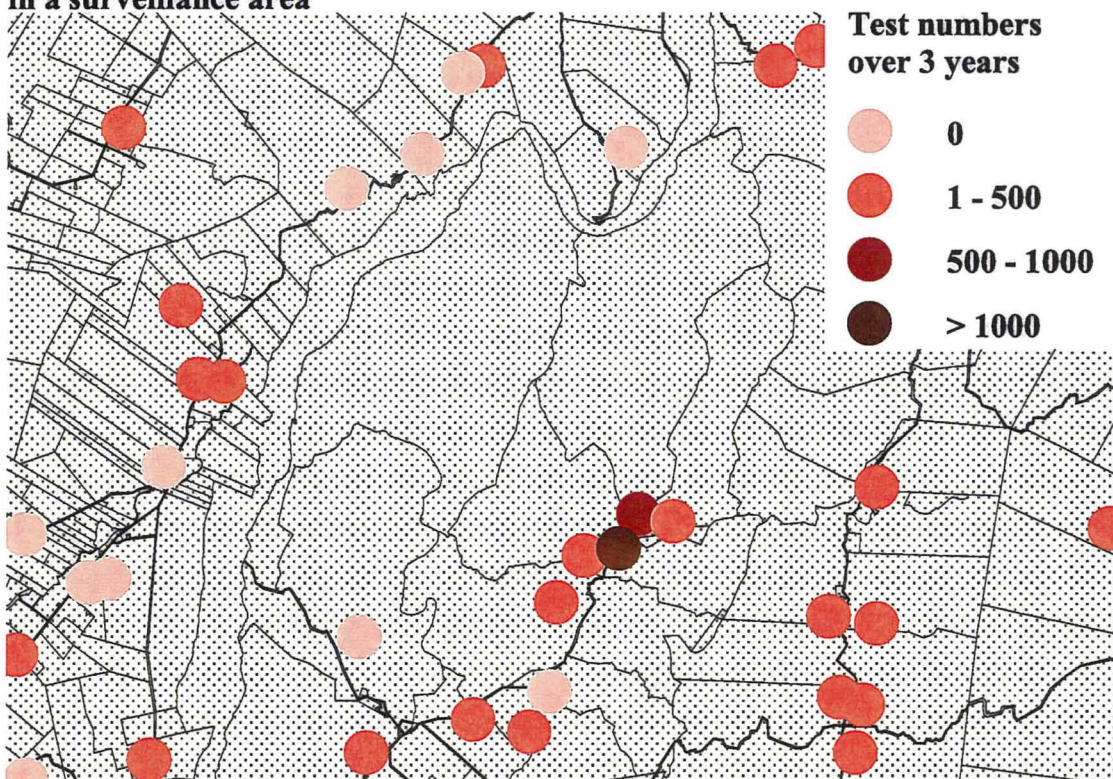


**Figure 14: Representing surveillance with point data using pie charts and colours**

Pie charts area used here to show the ratio of cattle killed to cattle tested on farms in a surveillance area



Colours are used here to show TB test numbers over a three year period in a surveillance area

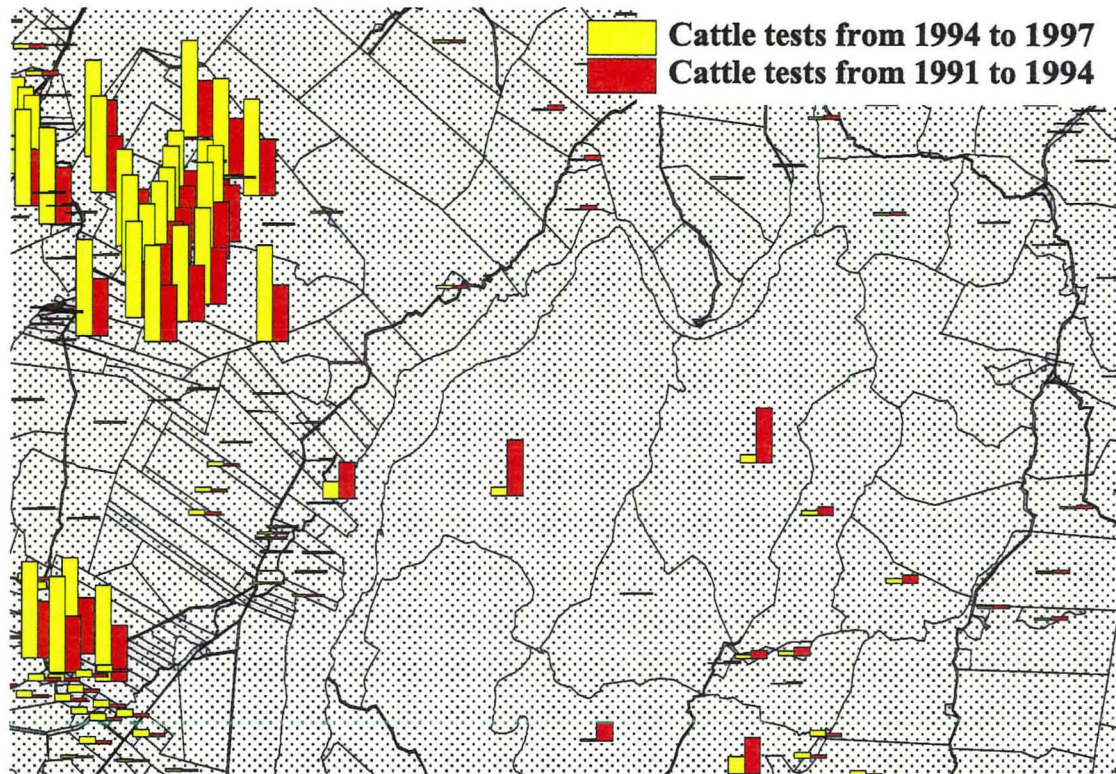


Note that it is difficult to adequately examine surveillance when points are used to represent farms

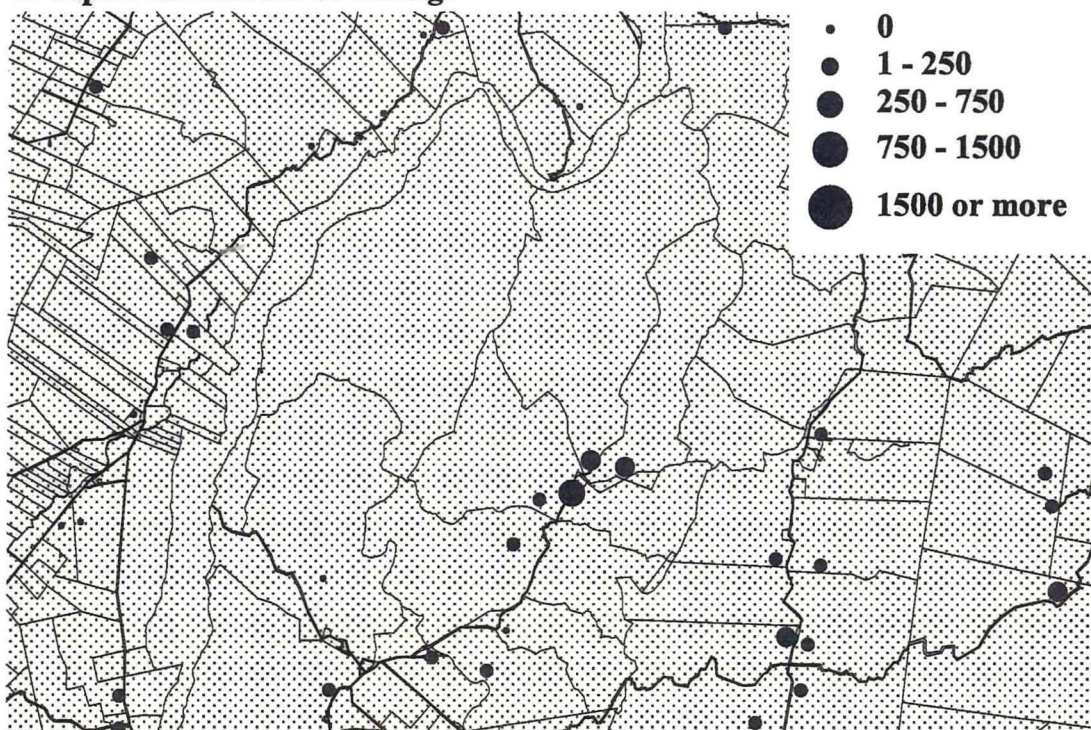


**Figure 15: Representing surveillance with point data using bar charts and different sized dots**

A bar chart is used here to show the amount of cattle testing in two time periods.



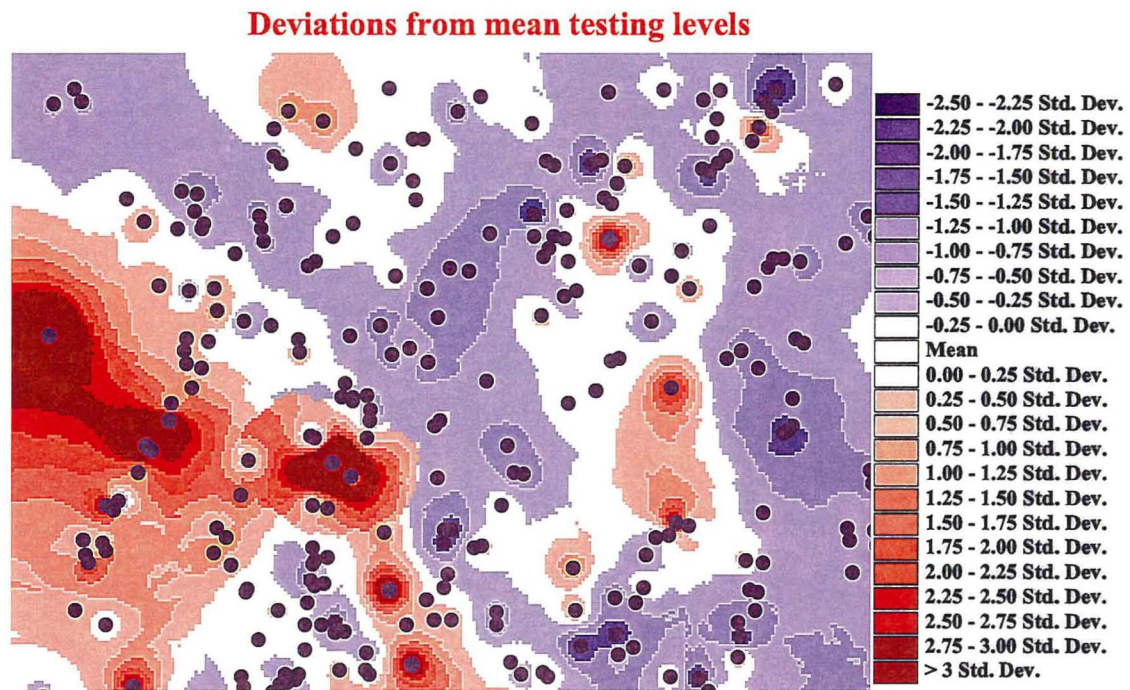
Different sized dots have been used here to depict the amount of testing.



This is not a very satisfactory way of presenting information.

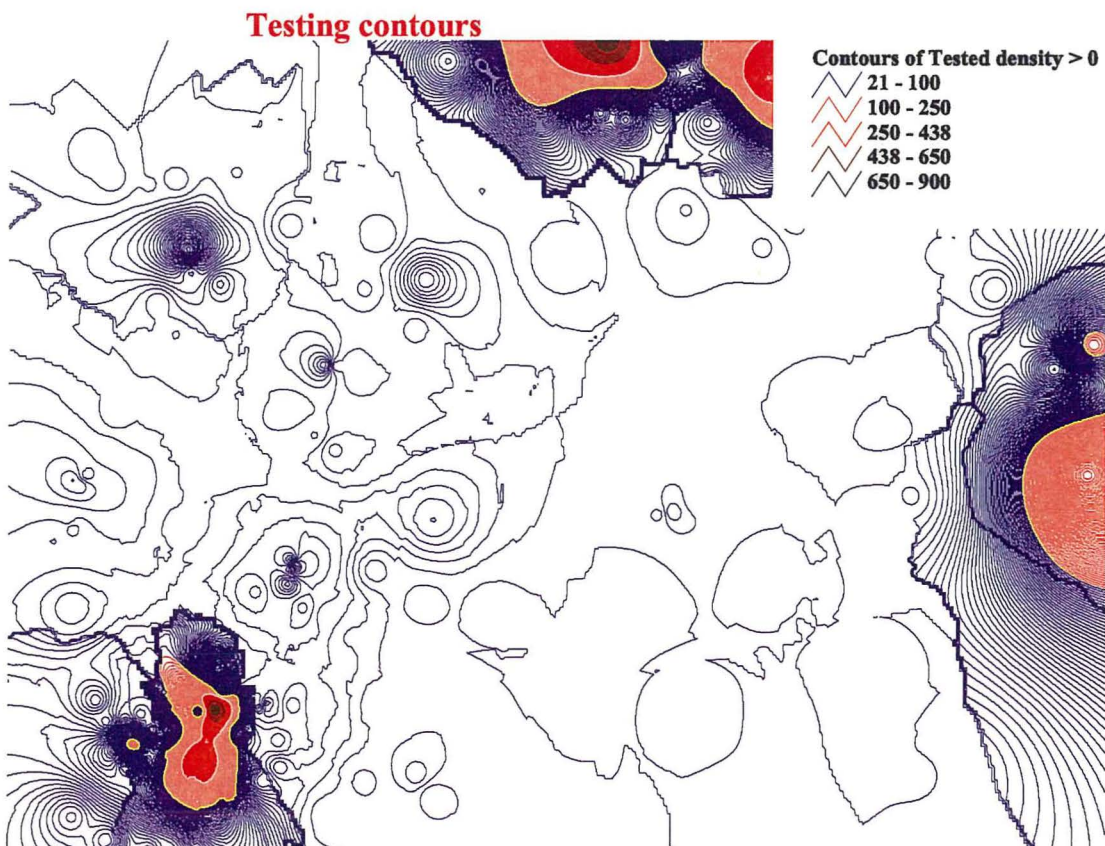


**Figure 16: Representing surveillance with point data using testing contours and deviations from mean testing levels**



These surfaces look attractive but they are misleading and should not be used.

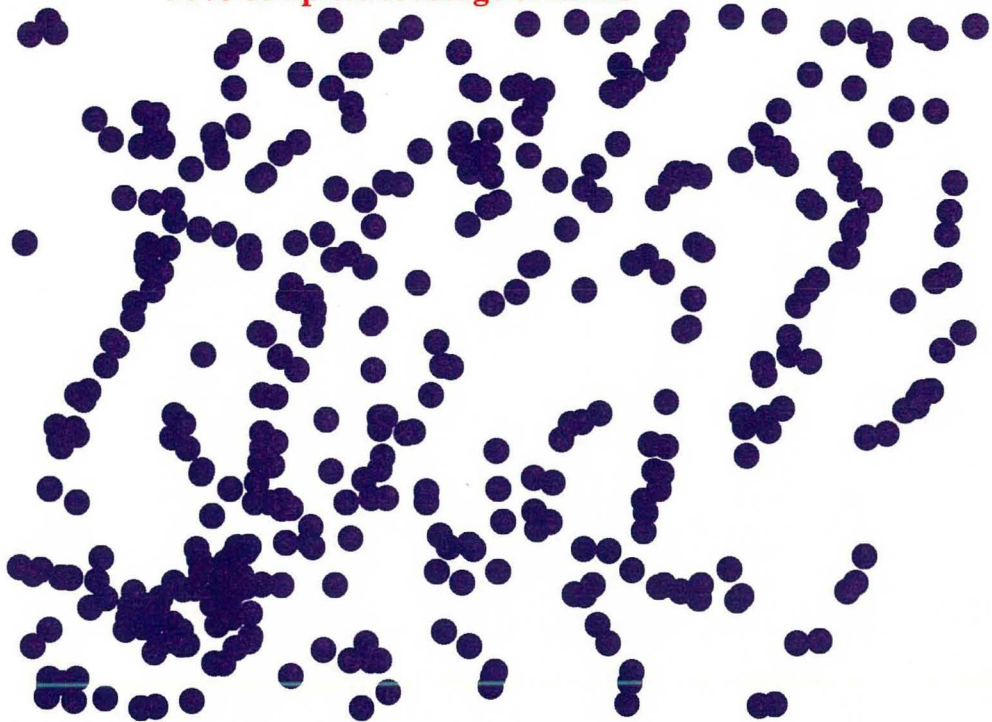
TB testing is an attribute of farms and farms are discrete objects. The tests on each farm are not a sample of a continuum of testing. The area represented is the same as in Figures 14 and 15.



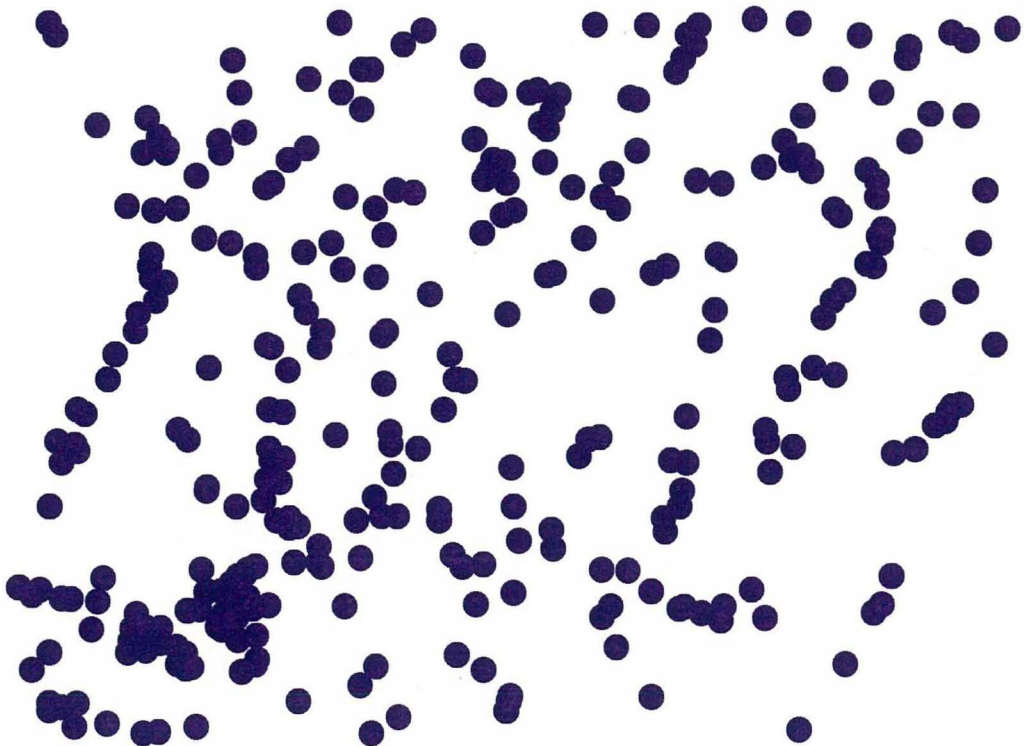
**Figure 17: Visualising missing data with point data**

Two coverages of farms are shown below. In one coverage only 75% of farms are present. By comparing these coverages with Figure 18 it can be shown that area coverages convey an understanding of what data is missing and point coverages do not.

**95% complete coverage of farms**



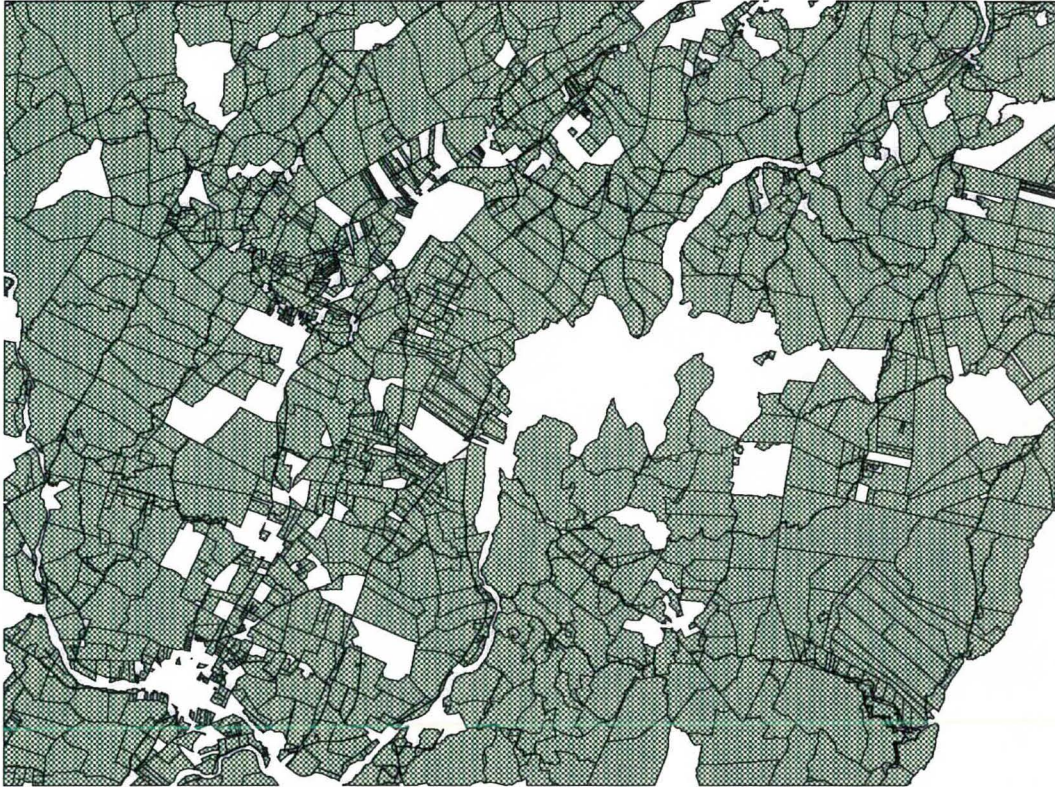
**75% complete coverage of farms**





**Figure 18: Visualising missing data using area data**

**95% data completeness**



**75% data completeness**

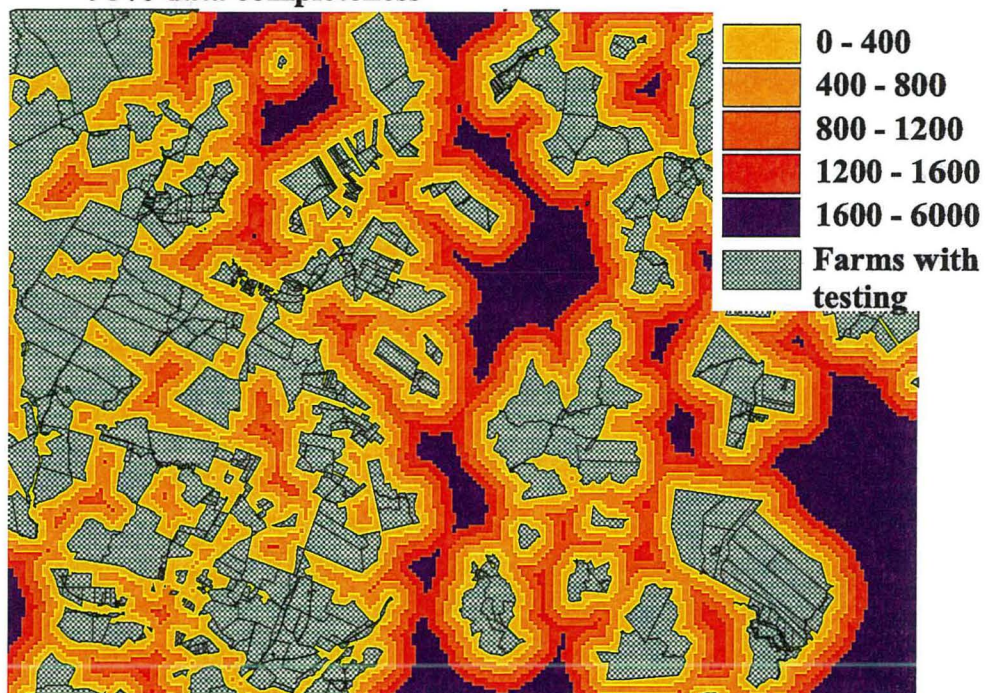




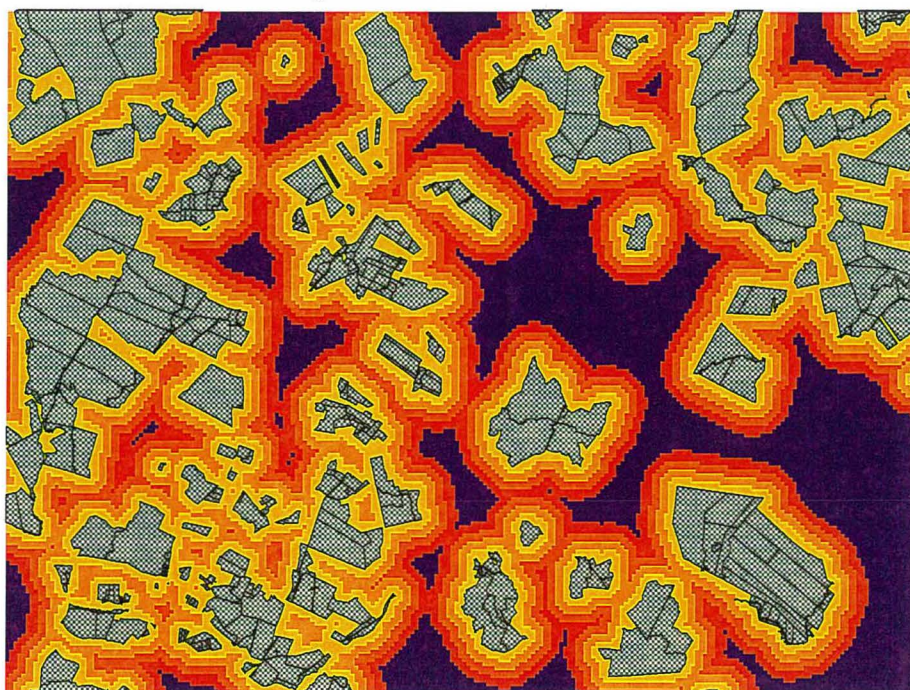
**Figure 19: The effect of missing data on detecting gaps in TB testing coverage.**

The distance (meters) from the edge of farms that tested in the 1996 -1997 testing season shown on two maps, one that includes 95% of farms and one that includes 75% of farms. Note that at 75% data completeness false gaps in surveillance appear (blue areas).

**95% data completeness**



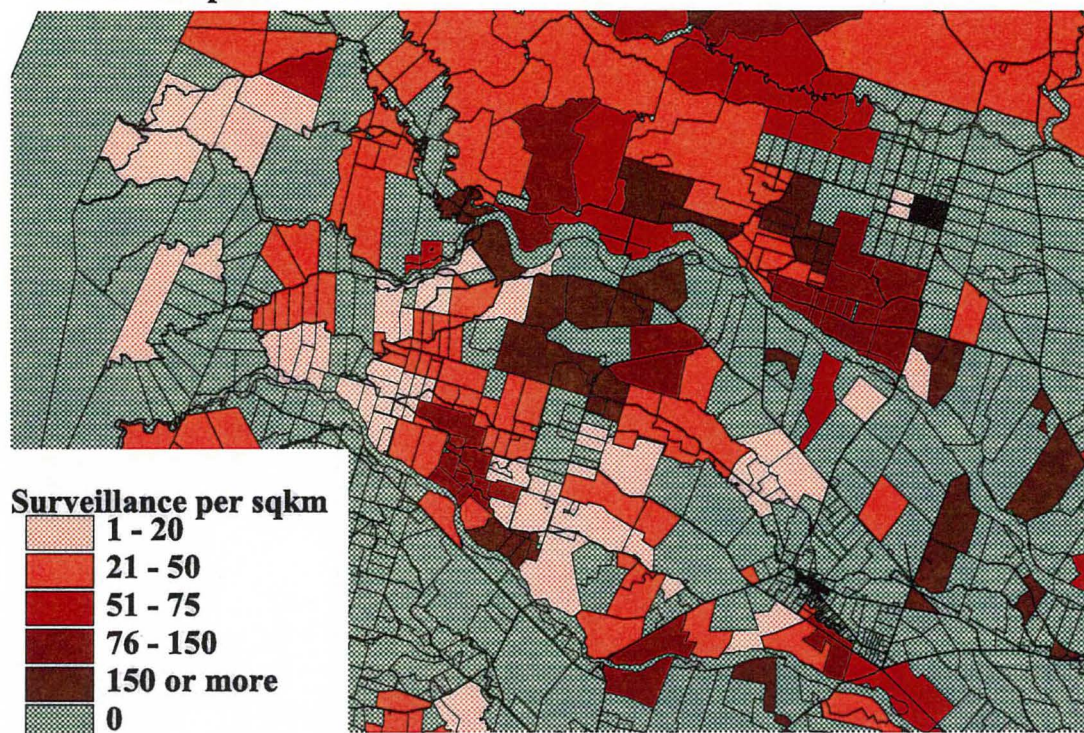
**75% data completeness**



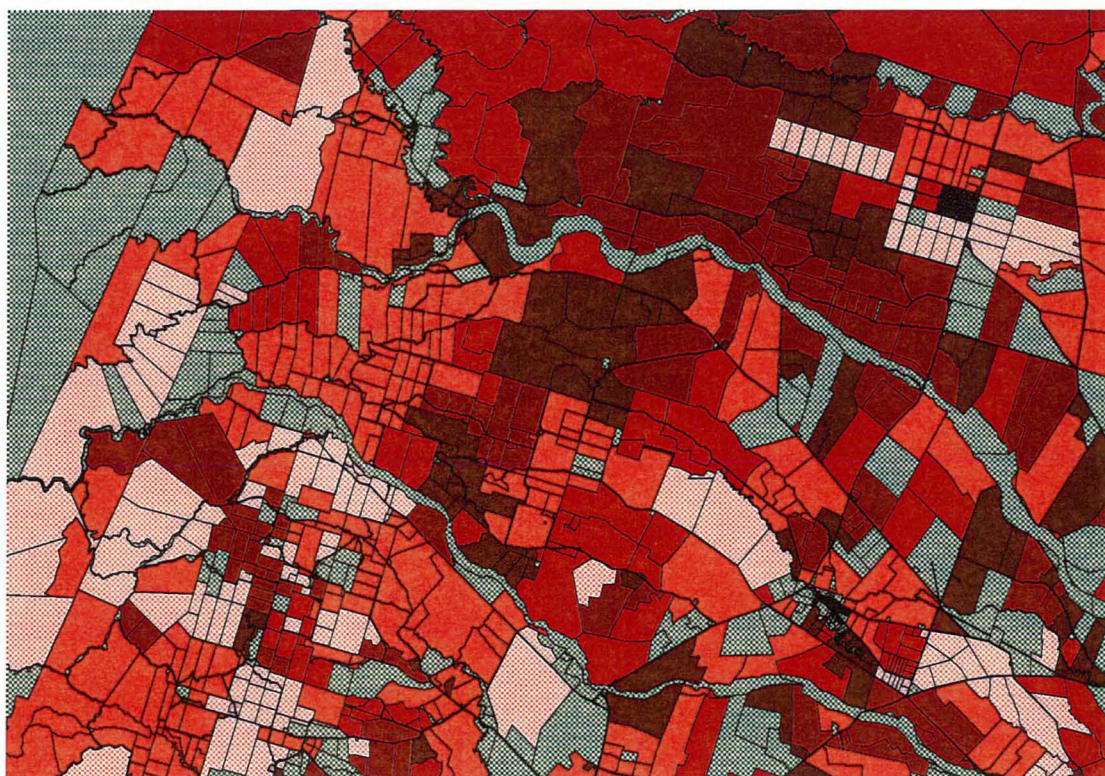


**Figure 20: Figure showing the contribution of slaughter surveillance to the TB surveillance program.**

**This map shows TB tests per square kilometer per annum in the Waipawa area.**



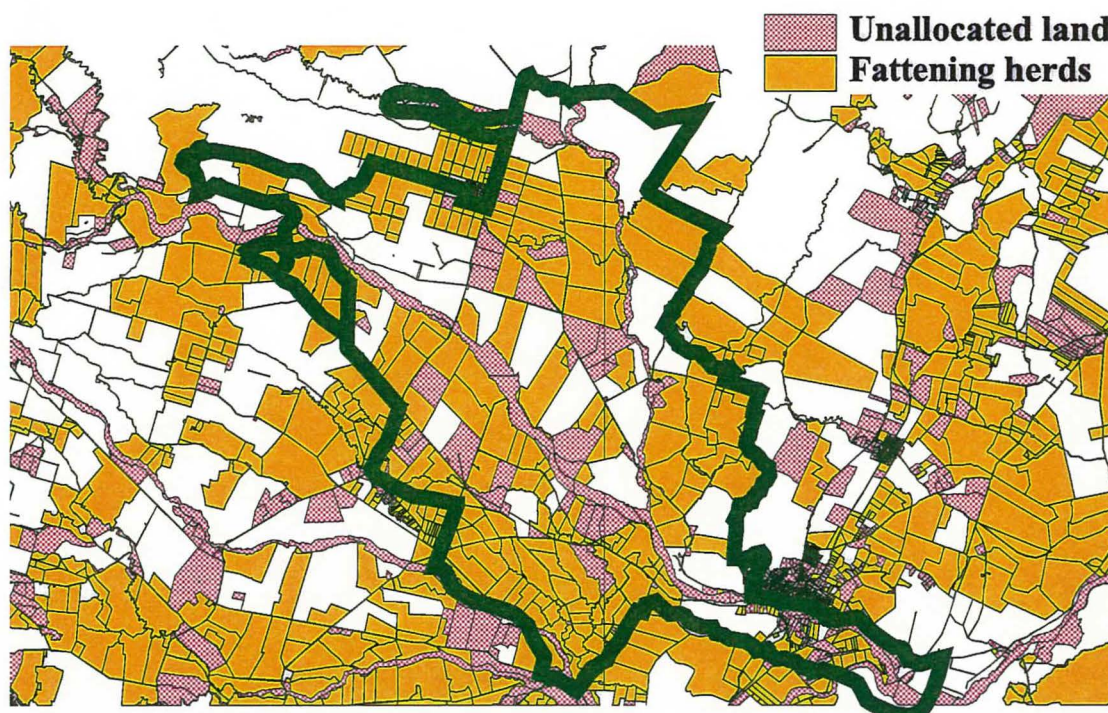
**This map shows testing and slaughter. The total resulting surveillance is quite extensive.**



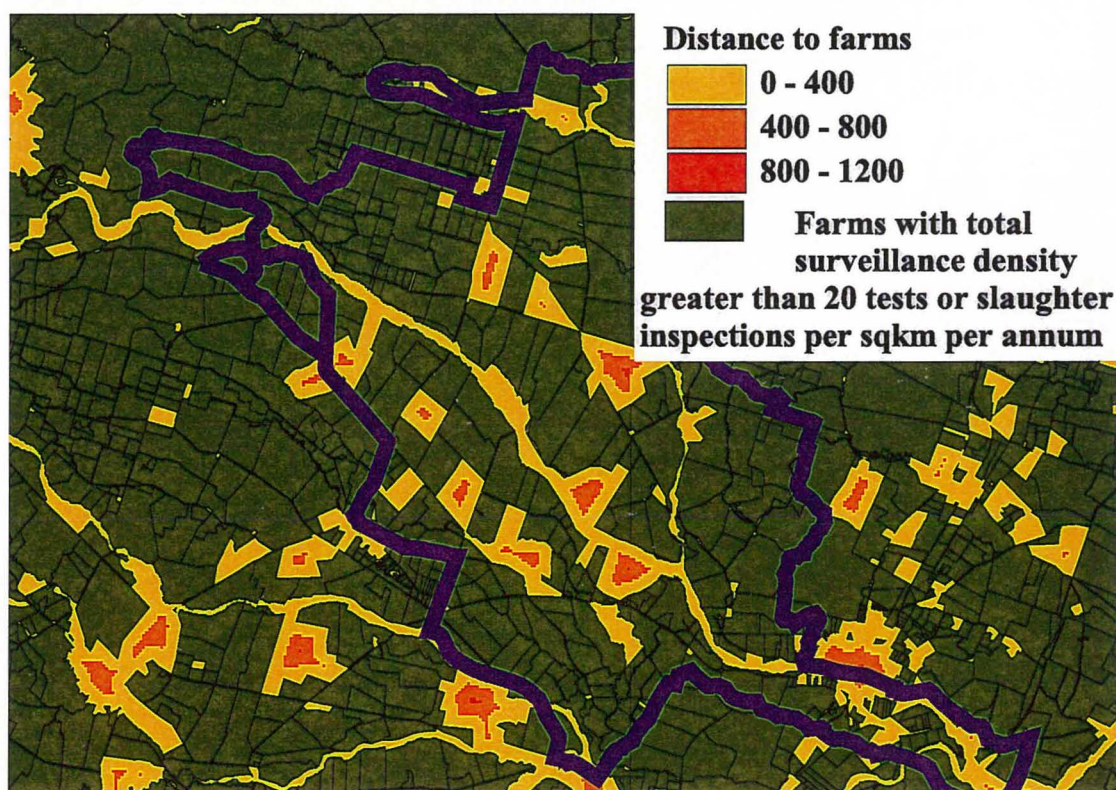


**Figure 21: The contribution of fattening herds to the total TB surveillance in the Waipawa vector risk area**

**The location and extent of fattening herds and unallocated land parcels**



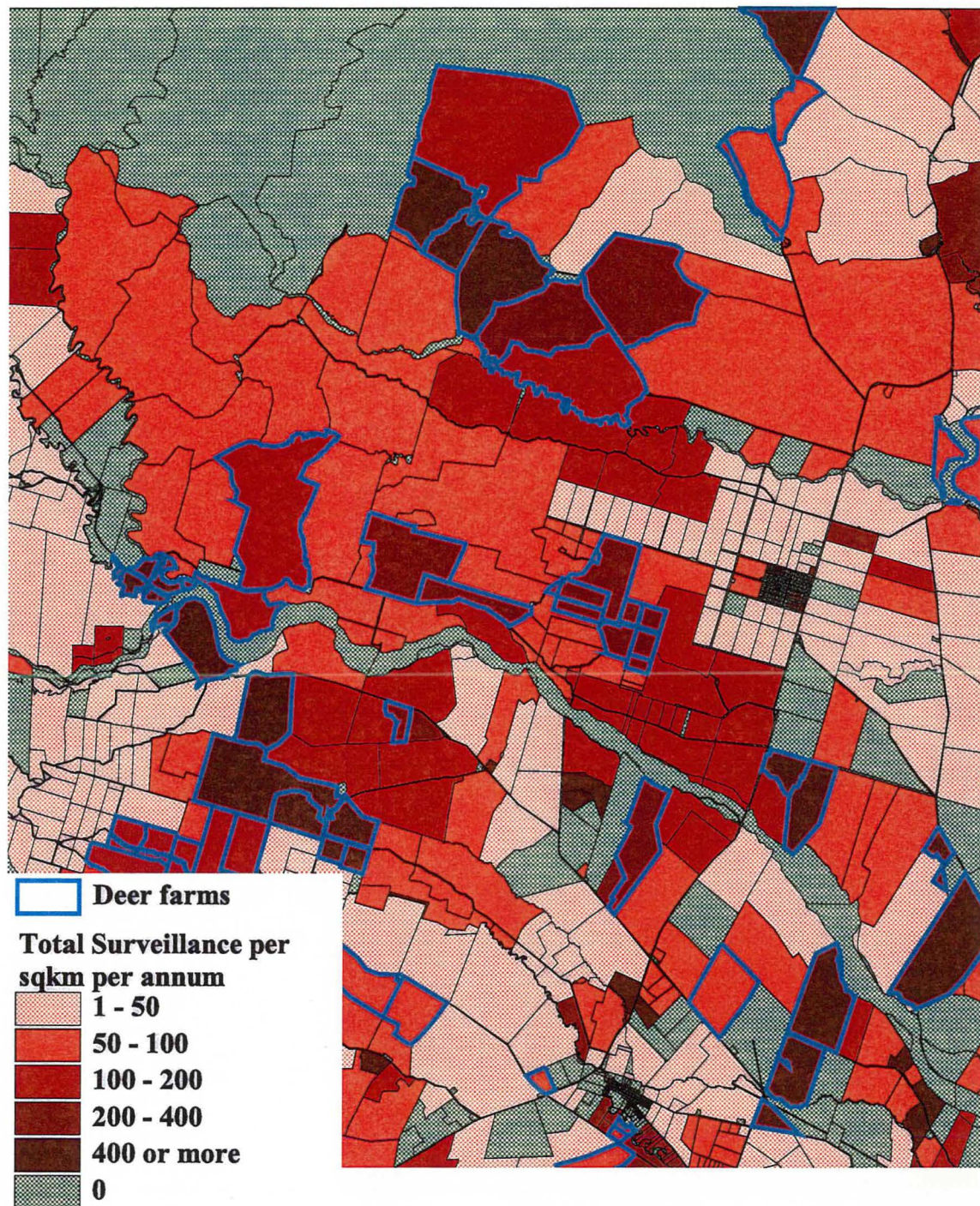
**Distance from farms with TB testing or slaughter inspection of stock.**



Note that when slaughter surveillance is included there are no gaps in surveillance coverage (compare with Figure 11).



**Figure 22: Figure showing the high levels of TB surveillance per hectare in deer herds in the Waipawa vector risk area.**



Slaughter inspections and tests from each farm in a given year were divided by the farm size to give an idea of surveillance intensity. Farms with deer herds have the highest resulting intensity of surveillance. They are shown with a blue border above. Deer herds have high stocking rates and consequently the testing intensity in these herds is high.

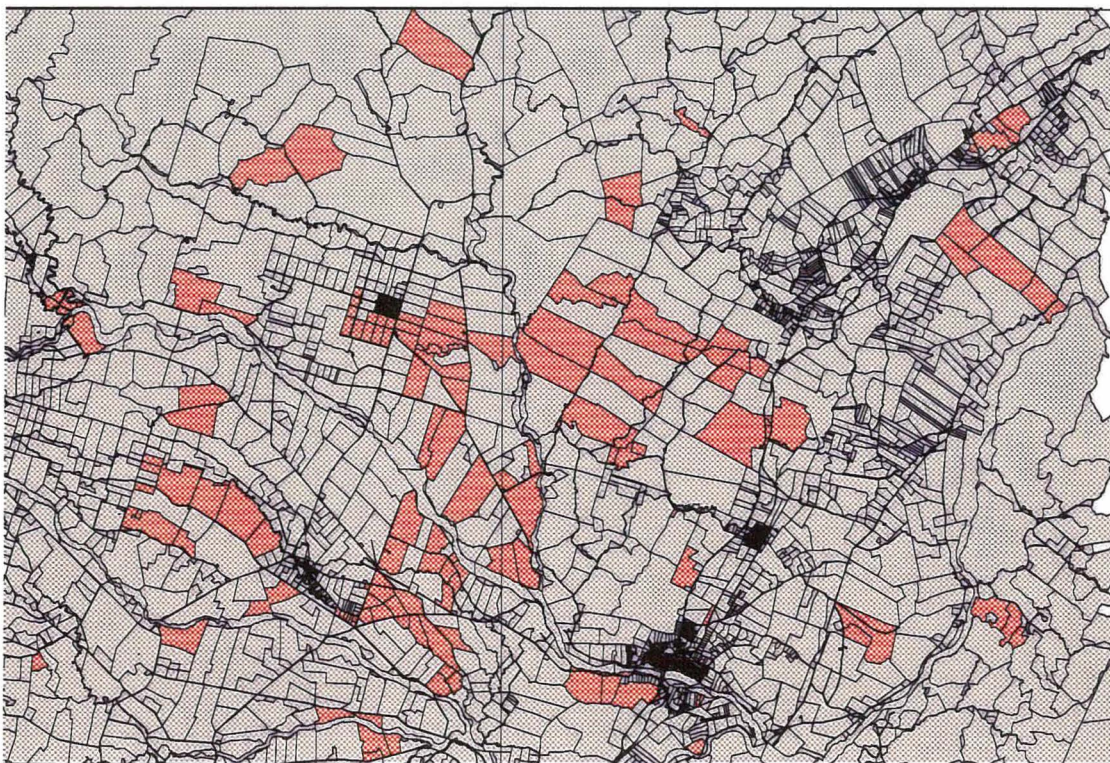


**Figure 23: TB infected farms at Waipawa since 1990  
and a random selection of farms**

**Random selection of farms**

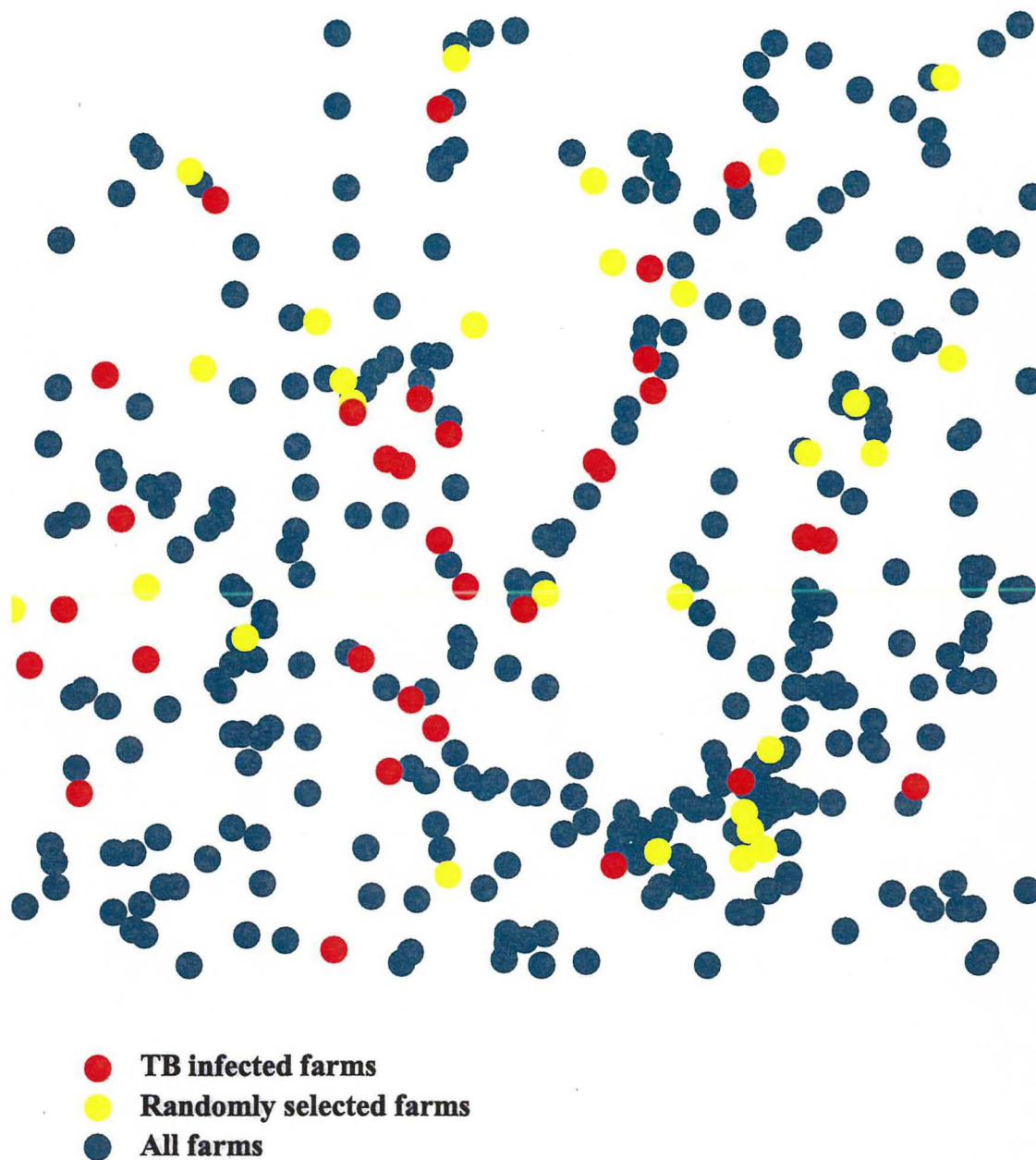


**TB infected herds**



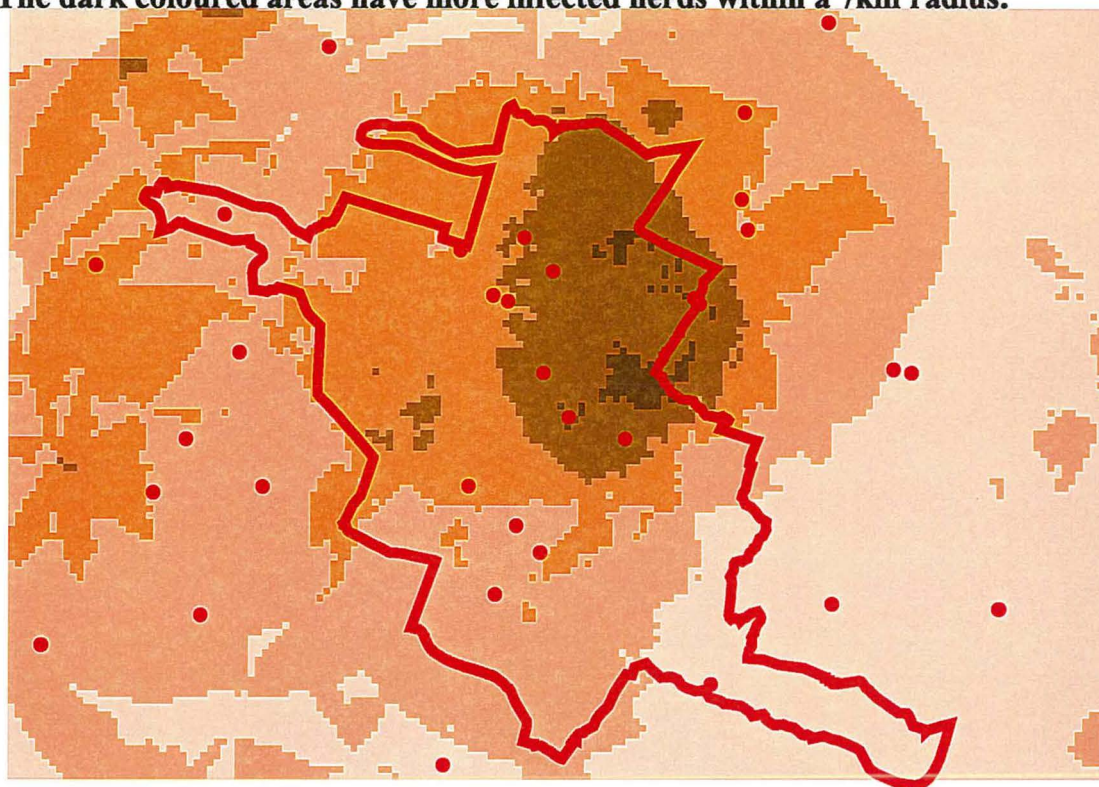


**Figure 24: Point locations of all farms, TB infected farms, and randomly selected farms at Waipawa**

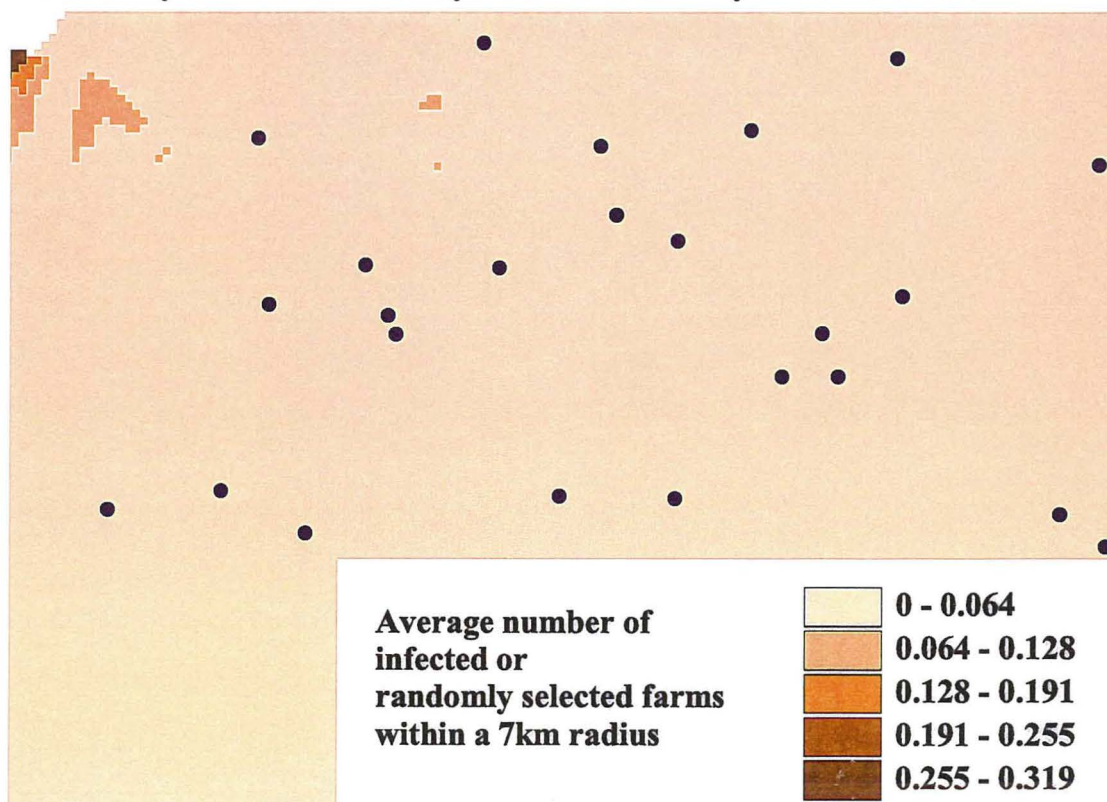


**Figure 25: Cluster analysis of TB infected herds at Waipawa compared with random controls**

**The dark coloured areas have more infected herds within a 7km radius.**



**Randomly selected herds analysed in the same way do not show a cluster**



## **Chapter 4**

### **Spatial data requirements for simulation models**

## **Abstract**

The spatial data requirements for simulation modelling were assessed by examining the effects on a simulation model of using different data types and by using incomplete data. An epidemic of foot and Mouth Disease (FMD) was simulated using the simulation model Interspread. When farms were represented as polygons the model results were markedly different from when they were represented as points. The polygon coverage resulted in 2 to 3 times as many farms becoming infected as the point coverage. This was due to the distance between area features in the polygon coverage being less than the distance between point features in the point coverage. Representing farms as points has many advantages. Point data is simple and inexpensive and requires less processing time. Adjustments were made to the spatial processes in the model that involved distance between features. The model was run using point data and these adjustments. Distance adjustments in the model resulted in similar model results for the point data coverage as the area data coverage. The distance adjustments were based on the size of the farms. The effect of missing data on a simulation model was found to be roughly proportional to the level of missing data.

## **Introduction**

Simulation models are used in disease management. EpiMAN is a decision support system for use in the event of an outbreak of foot and mouth disease (FMD). The system is based on a number of different components. Interspread is a computer model that is used for the simulation and prediction of the outcome of a foot-and-mouth disease (FMD) outbreak under different control policies. This enables a manager to conduct a series of "what if" scenarios. The program models the dynamics of disease transmitted between farms in time and space. There are four infection processes modelled: Wind borne spread, local spread, movement related spread and dairy tanker spread. The distance measurements in Interspread were derived from point data available from the FMD outbreak in England in 1967/68. The program uses enterprise details such as the presence of deer, pigs, cattle, and sheep. A farm in Hawke's Bay was selected as an infected property. The model was then run and the resulting predictions recorded. The consequences of representing farms as point features or area features for simulation modelling were assessed using the predictions generated by the simulation model Interspread. Running the model with random dilutions of these coverages assessed the effect of missing data on the model.

## **Method**

An FMD outbreak was simulated using an uncontrolled epidemic scenario and a controlled epidemic scenario. Each scenario was simulated based on area and point data. Each scenario incorporated 4 levels of data completeness: 100%, 90%, 80% and 70%. The Hawke's Bay livestock database is currently 90% complete. So the actual percentages were 90%, 81%, 72% and 63%. Table 16 shows the results for each scenario by the different data types and levels of data completeness. The results are expressed as the average of 5 iterations and represent a 5-week epidemic period. Due to the processing time required only two levels of data completeness were assessed with area data in the uncontrolled epidemic scenario.

## Results

### Data type

The effect of data type area or point on Interspread was dramatic. This is probably related to the use of distance in Interspread. The distance between two farm boundaries is not the same as the distance between two farm centroids. Table 17 shows the percentage decrease in the number of infected properties or spread mechanisms associated with the decrease in data completeness. Table 19 shows that 2 to 3 times as much spread occurs with area data. Table 19 shows that, in an uncontrolled epidemic using area data, 750 farms became infected. When points were used to represent farms only 413 became infected. In a controlled epidemic using area data 55 farms were infected. The number of infected farms decreased to 19 when point data was used. The decrease when using point rather than area data was reasonably constant when comparing between different levels of spatial data completeness. Table 19 shows that in the controlled scenario, regardless of the level of data completeness, the reduction in infected properties was about 60%. The decrease in the total spread mechanisms was also around 60%.

### Missing data

Table 19 shows that regardless of data type decreasing data completeness resulted in a decline in infected farms and a decline in spread mechanisms. The results were more consistent in the uncontrolled scenario. Area data showed a higher sensitivity to missing data. The effect of missing data, on an uncontrolled scenario using points to represent farms, was to reduce the number of infected farms by about the same amount. If 30 % of the farms were missing, then the decrease in infected farms was roughly 30%.

### Adjustment of distance criteria when using point data

The different results from the point and area coverage's stem from the use of distance in the simulation model. The distance between two farms is greater in a point coverage than an area coverage. Two approaches have been suggested to account for the difference. One method is to add twice the hypothetical mean or median radius to all distance measures related to the point coverage. This is satisfactory for a simulation model, especially where farm sizes are randomly distributed in the range. This was done the controlled epidemic scenario. The mean and mean radius for the point coverage was based on the known area of each farm. This assumed the farm had a circular shape. The mean of the hypothetical radius of each farm was 700m. The results are shown in Table 21. When using point data the addition of twice the median radius to all distance measures used in the model produced similar results to the area data coverage. The main advantage of using adjusted point data is the speed at which the model runs. The adjustment was also used at different levels of data completeness. The effect of the adjustment is to increase the mean number of properties influenced by any given farm, bringing it to a similar level as area data.



## Discussion

Incomplete data did affect the predictions resulting from different scenarios. The target level of data completeness for EpiMAN was 70%. The results show that at 72% data completeness, the percentage of infected farms is 36% lower than was calculated at a 90% level of completeness.

Data type had a large effect on the outcome of the simulation model. Area coverages provide a real measure of farm size, shape, and contiguity of neighbouring farms. Point data is simple and inexpensive and requires less processing time. Adjustment has to be made for the type of data used. Similar results can be achieved with a point data coverage as long as adjustment is made in any spatial processes that involves the distance between farms. The adjustment can be adaptive using the size of each farm to calculate the hypothetical radius for each farm. The adjustment used above used the mean of all hypothetical farm radii. A textual field containing farm size was used to calculate the hypothetical farm radii. Farms are not randomly distributed in regard to size. Small farms are clustered and so are big farms. The extent to which this occurs affects the validity of a global adjustment. The influence of large farms and small farms is reduced and increased respectively by using a mean measure. Adaptive adjustment may be required when working with heterogeneous coverages. Examining the residual difference between the adaptive radii measurements and the mean radius for the coverage can assess this. Figure 27 shows the difference between adaptive radii and the mean radius at Waipawa.

**Table 16: Cross-tabulation of number of infected farms resulting from Interspread modelling stratified by spread mechanisms and control scenario.**

Epidemic type	Data type and level of completeness		Infected farms resulting from the model	Spread mechanisms and the number of spread events				
				Local	Airborne	Movement	Dairy tanker	Total
Uncontrolled	Point	90%	413	65	524	566	129	1284
	Point	81%	342	58	410	549	106	1123
	Point	72%	318	37	328	444	104	913
	Point	63%	297	31	273	463	92	859
Uncontrolled	Area*	90%	750	109	1188	668	113	2078
	Area	63%	443	67	781	461	94	1403
Controlled	Point	90%	19	1.6	10.4	6.6	1.6	20
	Point	81%	28	1.2	13.6	11.8	4	31
	Point	72%	14	1.2	4.2	7.2	1.6	14
	Point	63%	16	0.4	5.8	6.6	3.4	16
Controlled	Area	90%	55	1.5	37.5	15.5	5.5	60
	Area	81%	51	3.5	39.5	15	0.5	59
	Area	72%	35	2	19	12.5	0.5	34
	Area	63%	42	2	24	13	3	42

\*Results for 1 iteration only.

**Table 17: Cross-tabulation of percentage drop in number of infected farms and in spread mechanisms stratified by data completeness, data type and control scenario.**

Epidemic	Data type		Percentage drop in infected farms	Percentage drop in Spread mechanisms				
				Local	Airborne	Movement	Dairy tanker	Total
Uncontrolled	Point	90%	0	0	0	0	0	0
	Point	81%	17	11	22	3	18	13
	Point	72%	23	43	37	22	19	29
	Point	63%	28	52	48	18	29	33
Uncontrolled	Area	90%	0	0	0	0	0	0
	Area	63%	41	39	34	31	17	33
Controlled	Point	90%	0	0	0	0	0	0
	Point	81%	-47	25	-31	-79	-150	-51
	Point	72%	26	25	60	-9	0	30
	Point	63%	16	75	44	0	-113	20
Controlled	Area	90%	0	0	0	0	0	0
	Area	81%	7	-133	-5	3	91	3
	Area	72%	36	-33	49	19	91	43
	Area	63%	24	-33	36	16	45	30

**Table 18: Summary statistics for farm radii in the coverage used with Interspread**

Farms	Mean radius	-95% cf	+95 % cf	Median	Sd	Skewnes	Kurtosis
3523	700m	675	725	520m	7.5	6	106

**Table 19: Cross-tabulation of number of infected farms and transmission events by adjusted point data type and spatial data completeness for controlled epidemic simulations**

Data type	Complete ness	Infected farms	Spread mechanisms				Total
			Local	Airborne	Movement	Tanker	
Area	90%	55	2	38	16	6	60
Point	90%	19	2	10	7	2	20
Point + 2r*	90%	64	4	41	22	2	69
Point+2m*	90%	57	4	32	24	3	63
Area	81%	51	4	40	15	1	59
Point	81%	28	1	14	12	4	31
Point + 2r	81%	61	3	43	20	3	69
Point+2m	81%	46	3	30	15	3	51
Area	72%	35	2	19	13	1	34
Point	72%	14	1	4	7	2	14
Point + 2r	72%	49	3	30	19	4	56
Point+2m	72%	41	3	26	14	3	46
Area	63%	42	2	24	13	3	42
Point	63%	16	0	6	7	3	16
Point + 2r	63%	37	2	26	12	4	43
Point+2m	63%	34	1	22	11	3	37

\*Point + 2r represents point data including an adjustment of all distance criteria of plus twice the hypothetical mean radius. Point +2m represents the same adjustment using the median.

**Figure 26: Line chart of the number of infected properties resulting from a controlled epidemic under different data adjustment methods**

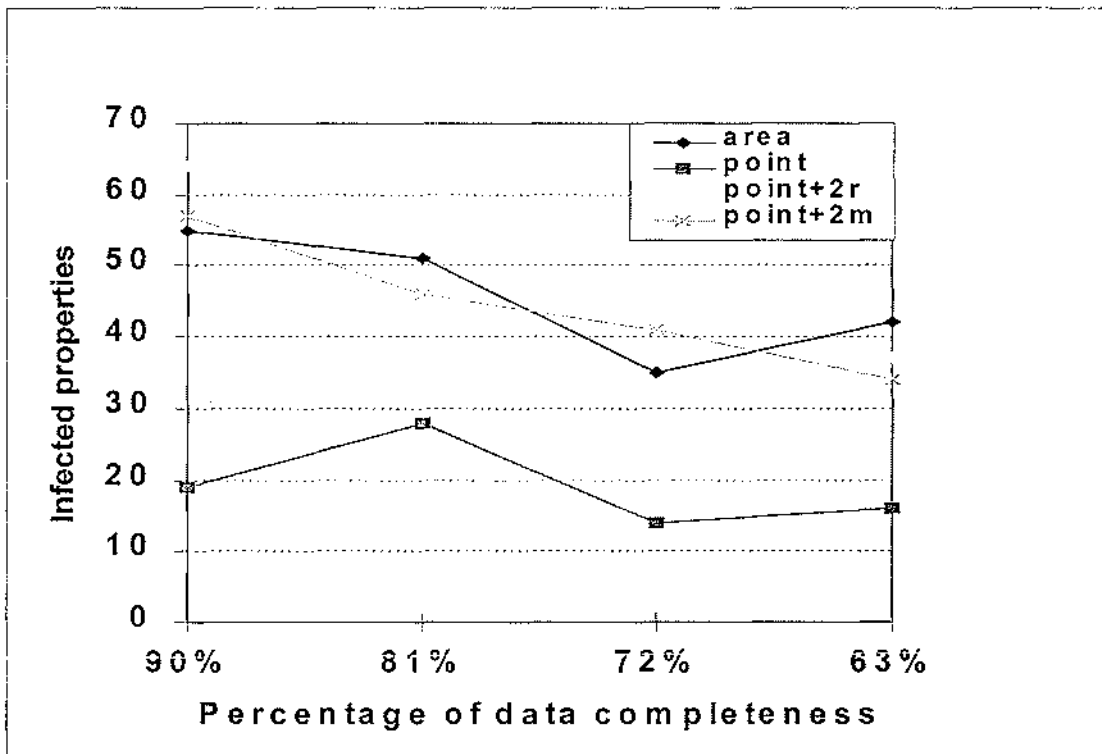
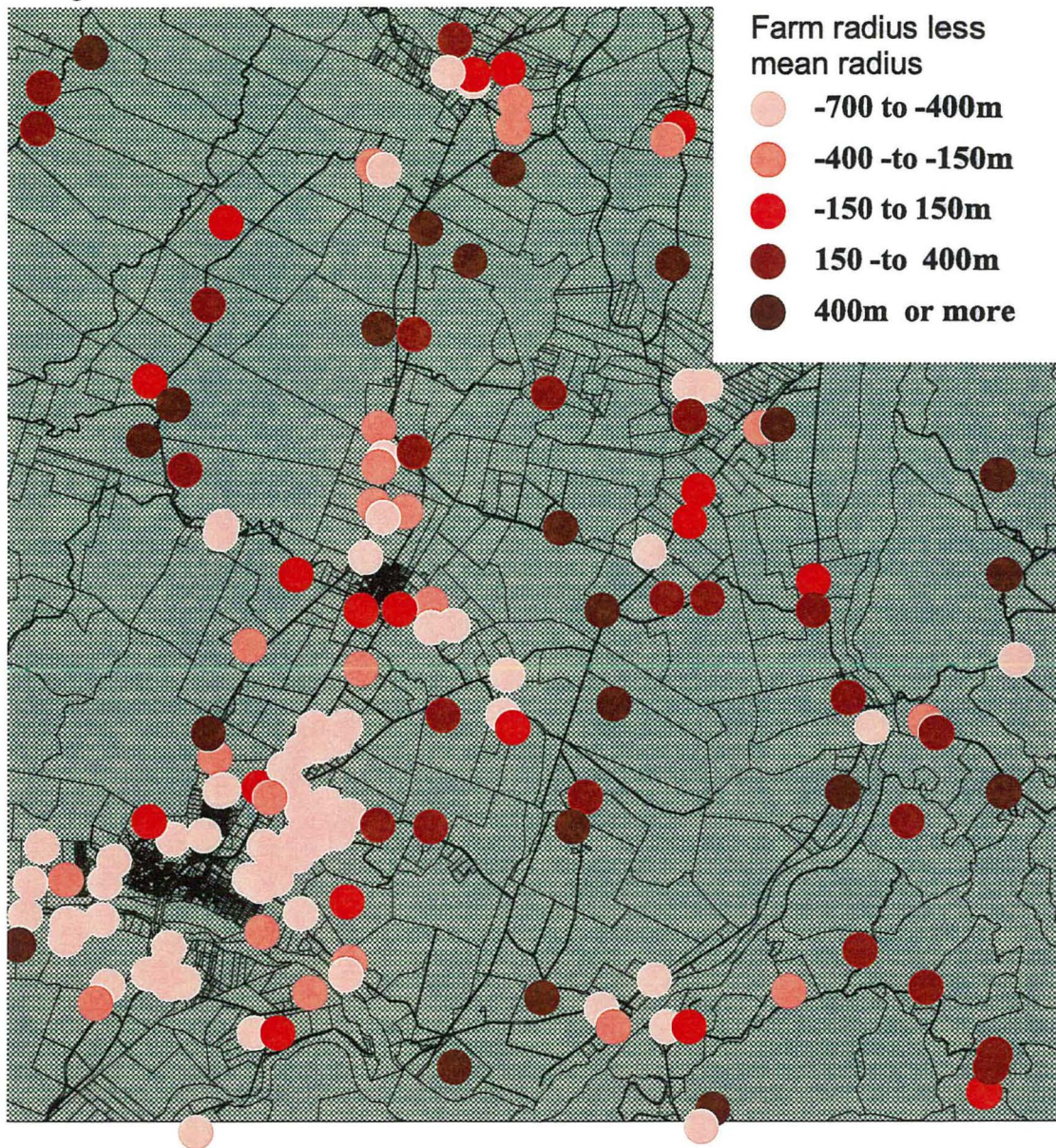




Figure 27: Figure showing the residuals between the actual radius of a farm less the mean radius of all farms

Note that the difference in actual and mean radius was clustered. This is due to clustering of farm size. Large farms are grouped together as are small farms



The "radius" a farm was calculated by assuming its area was circular. The mean radius of all farms was subtracted from the radius of each farm. The resulting number is mapped. Light coloured dots are farms that are smaller than average.

## **Chapter 5**

### **Spatial data requirements for cluster detection**



## Abstract

Points or polygon can represent farms. Polygons form a mosaic of farms. A matrix can represent the proximity between farms. The matrix can include farms that share a common boundary, or farms within a given distance from the boundary. The only measure of proximity in a point coverage is the distance between points. This can distort the relationship between farms as farms that are geographically close together may be non-contiguous. In a point coverage large farms are further from their neighbours than small farms. The consequences of representing farms as points on cluster detection were assessed. The degree to which neighbouring farms were correctly classified as neighbours was assessed.

The true relationship between farms was distorted when a point coverage was used. Adjustment was needed to account for the difference in distance between two point features as opposed to the distance between two area features. A simple adjustment was the addition of the hypothetical radius of the farm. The degree to which the true contiguity between farms is represented by a point coverage should be known before cluster detection methods are applied.

## Introduction

In the study of diseases in domestic livestock, epidemiological data is often aggregated at the farm level. Farms can be represented spatially as points or area features. Such data can be used to detect clustering of disease in time and space. The effect on cluster detection of representing farms as point features or area features was assessed using Moran's I (Cuzick and Edwards, 1990).

If farms are represented as polygons proximity can be measured using a ( $n \times n$ ) spatial proximity matrix  $W$ , where  $w_{ij}$  represents the spatial proximity of areas  $i$  and  $j$ . The matrix can be calculated using a range of methods. If two farms are considered to be in close proximity  $w_{ij} = 1$ . The element  $w_{ij}$  can be set to 1 under a variety of conditions:

- 1 where the centroid of  $i$  is one of the  $k$  nearest centroids to the centroid of  $j$ ;
- 2 where the length of the common boundary with  $j$  divided by the perimeter of  $i$  is greater than a certain proportion;
- 3 where the centroid of  $i$  is within a specified distance to the centroid of  $j$ ;
- 4 where  $j$  shares a common boundary with  $i$ .

The element  $w_{ij}$  can equal 0 where the above conditions are not met.

Points are often used to represent farms. This can distort the true proximity between farms. The aim of this analysis was to determine the validity of using points in cluster detection.

A cluster of bovine tuberculosis (TB) cases in domestic cattle and deer at Waipawa in New Zealand was used in the analysis (Mackereth 1993). The cluster of farms with TB in livestock was observed after the establishment of TB in a community of common Brushtail possums (*Trichosurus vulpecula* Kerr), a known vector of TB in New Zealand (Pfeiffer, 1993).

## Method

The data coverage used for cluster detection was a 30-km by 40-km area at Waipawa in Hawke's Bay, New Zealand. The area included a known cluster of farms with bovine tuberculosis (TB) cases in domestic livestock. Over an 8-year period TB was diagnosed on 36 farms. There are 449 farms in the coverage used. Moran's I (STAT! Biomedware 516 North State Street Ann Arbor, MI 48104) was applied as the cluster detection

method. Moran's I is a test for spatial autocorrelation in disease rates. Positive spatial autocorrelation means that nearby farms have similar rates, indicating spatial clustering. The test statistic Moran's I is the weighted product-moment correlation coefficient, where the weights reflect geographic proximity. Values of I greater than zero indicate positive autocorrelation; values smaller than zero indicate negative autocorrelation. The null hypothesis is that disease rates are spatially independent. In the calculations shown here, 99 randomised simulations were run on each scenario and the distribution of Moran's I (randomised) was plotted and compared with the actual value from the non randomised data. If the value lies outside the simulated distribution its significance can be assessed. Two files are used in the calculation of Moran's I. A data file contains the disease rates for each farm. A contiguity file containing a list of farms considered to be contiguous for each farm. Selecting all farms within a given distance created the contiguity file. The search radius around a point encompasses fewer farms than the same radius around an area feature. The search radius should be greater if point data is used than if area data is used. Adjustment of the distance criteria used in the creation of a contiguity file resolved some of the differences between point and area data in cluster detection. The degree to which point data with adjusted distances adequately represented area data was assessed. Correct and incorrect classification of point data contiguity was assessed at different distances, using contiguity derived from area data as a gold standard. Some spatial processes use k-nearest neighbours as a measure of contiguity. This was also compared with actual contiguity based on area data.

#### **Distance adjustment for point data.**

The extent to which the distance needs to be increased to give a similar result as area data is of relevance for all spatial processes that use distance methods for estimation and simulation. The complexity of computation and the processing time is significantly reduced if farms can be represented as points rather than areas. Two approaches were used to quantify the required adjustment. In the first approach a hypothetical radius was calculated for each farm, based on its area. The farm was considered to be circular for this purpose. Twice the average or median hypothetical radius was used as the adjustment. A more sophisticated adaptive approach was to use the hypothetical radius of each farm. The aim of the adjustment is to allow the spatial simulation process to more correctly classify properties as being near or far away. The gold standard is the contiguity of the polygon coverage. The sensitivity and specificity of the *representation of contiguity* can be assessed by considering the number of correct and incorrect classifications.

## **Results**

Table 21 shows the Moran's I values and their significance for area and point data. The results show clustering was detected more readily with area than with point data. Area data showed clustering when contiguity was based on distances of 100m, 1000m, 3000m, 5000m and 7000m. Analysis of the area data indicated that each property had an average of 5.7 properties to be considered contiguous (within 100 m of the boundary) to it. The same analysis of point data resulted in an average of 1.1 properties within 100m. There were insufficient farms in the point coverage within 100m of each other to calculate a Moran's I. The effect of an adjustment of twice the median hypothetical radius was to raise the average number of neighbours from 1.1 to 9.7. The hypothetical radius of each area feature in the coverage was measured. The mean radius was 708 meters with a 43-meter confidence interval. The median was 696 meters. Figure 28 shows that the adjustment was slightly more than required to give a similar number of neighbours. The

adjustment did behave consistently. While the average number of contiguous properties increased with the adjustment it was of interest to what extent the actual contiguous properties were represented by the adjusted point data. Moran's I at 100m was significant for area data but not for the adjusted point data, suggesting that the adjustment was poorly sensitive and specific. At greater distances the process was more reliable. From Figure 28 it can be concluded that an addition of 1070 meters to the point data search radius would result in a contiguity matrix with a similar average number of neighbours as a contiguity matrix derived from area data. This represents 1.5 times the mean and median hypothetical radius of the coverage used.

## **Discussion**

Point data and area data can not be used interchangeably. It is necessary to examine the consequences of data type for the spatial processes being used. The results show that point data can not be used to represent farms in cluster detection. If adjustments are made based on the size of the farm it may be possible to use point data. The validity of the adjustment depends on the heterogeneity of the coverage in regard to farm size. Points may not accurately represent the true complexity of contiguity or proximity. The classification of near or far should reflect the spread mechanism of the disease being modelled. An assessment of the sensitivity and specificity of the classification of contiguous properties in a point coverage can be made by comparison to the equivalent area coverage.



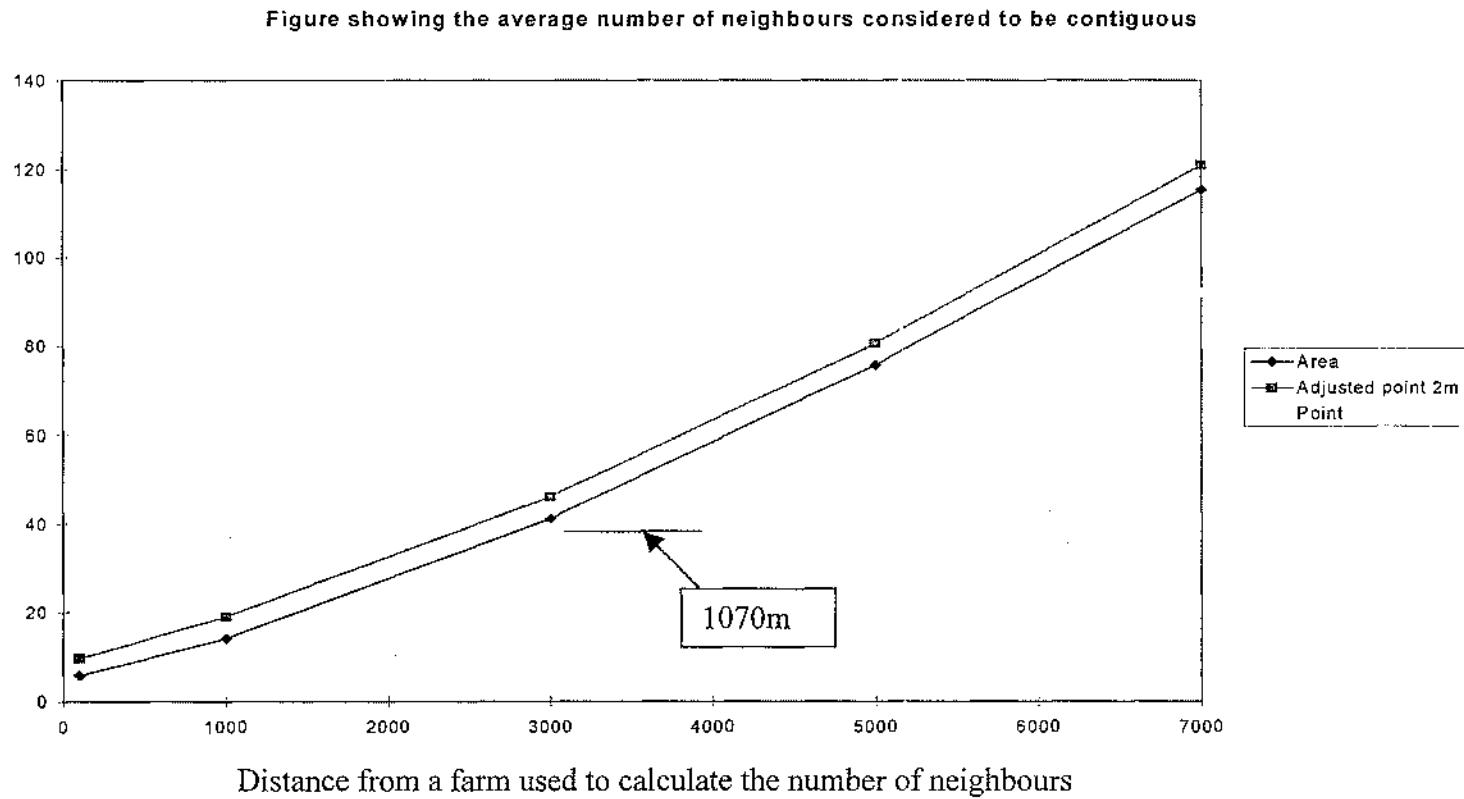
**Table 20: Summary statistics for the radius of farms at Waipawa for cluster detection**

Coverage	Farms	Mean	-95%	+95%	Median	Sd.	Skewness	Kurtosis
Interspread coverage	3523	700	675	725	520	752	6	106

**Table 21: Cross-tabulation of Moran's I values for TB infected farms in the Waipawa area using 5 different distances to define contiguity, stratified by data type.**

Data type	Immediate neighbours			1000 meters			3000 meters			5000 meters			7000 meters		
	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours	Moran's I	P value	Number of neighbours
Area	0.079	0.0003	5.8	0.052	0.0001	14.2	0.035	0.000004	41.2	0.023	0.00001	75.7	0.023	0.00000	115.4
Point	-	-	1.1	-	-	5.8	0.036	0.00019	26.6	0.017	0.004	56	0.02	0.00002	92
Point +2m	-0.019	0.67	9.7	-	-	19.1	0.021	0.002	46.2	0.018	0.0002	80.5	0.018	0.000003	120.78

**Figure 28: The average number of contiguous neighbours in a given number of meters from a farm calculated with different data types**



## Discussion

### Validity

Geographic information systems are used to create a spatial model of the *real world*. Based on this model disease managers visualise livestock production systems and farms. The model of farms and herds is used to display stock or disease data. Spatial analysis is used to reveal clusters of disease, determine sentinel surveillance or other sampling requirements and to predict disease spread. It is important the limitations of the model are not forgotten. The validity of the model must be assessed for the specific disease and analysis technique being applied to it. The primary spatial data requirement for disease management is model validity. The completeness of the data, the management of proximity, and the nature of the specific disease process influence model validity.

### Data completeness

Data completeness was examined on two New Zealand databases. Agribase and the National Livestock databases. For most disease management purposes they were found to have a sufficient level of data completeness. Incomplete data dilutes the field. The model is less populous than the real world. In this case disease spread may be reduced and the intensity of surveillance may be under represented. Missing data had an effect on disease spread and sentinel surveillance assessments. The effect was roughly proportional to the percentage of data that was missing. Some processes such as cluster detection were comparatively insensitive to missing data. Bias is a potential problem with the data available in Agribase and the NLDB. Small farms and miscellaneous farms were under represented.

### Representing proximity

Of greater concern than missing data was the validity of using points to represent farms. Farms are made up of land parcels. These land parcels can be contiguous or disjointed. The complexity of farming enterprises is such that even when all the land parcels are included, the model may not represent the actual land that the farmer grazes. Constant changes in land ownership and farming practice are difficult to represent in a spatial model. The concept of farm-neighbours is important when examining disease spread from one farm to the next. Representing farms as polygons enables the bounding neighbours to be enumerated. Using points to represent farms gives rise to a number of distortions of the relationships between farms.

The first distortion is in the classification of neighbours based on distance rather than a common boundary. Polygon data allows a variety of measures of proximity to be constructed. For a model of disease spread by direct nose to nose contact across a fence polygon data is essential. For a model of local airborne disease spread, all farms within a given distance and direction from a boundary can be considered neighbours. The relationship between disease events on farms and the spatial distribution of a vector can only be represented with polygon data. The proximity matrix that can be constructed from polygon data is far superior to any measures of proximity based on point data. Cluster detection techniques based on point data may not be valid for this reason. If all farms were square and formed a uniform grid it may be valid to use points to calculate contiguity. In this case the classification of neighbours and the inference about the

influence between neighbours would be the same for the point coverage as the polygon coverage. In reality farms are all shapes and sizes.

Another distortion that occurs with point data has to do with farm size. The distance between a farm and its neighbour is zero in a polygon coverage. In a point coverage it depends on the size of the farm. The position of the point is relevant. Is it in the centre of the farm? It can be argued that for spatial processes that involve larger distances these effects would become minor. This would be true if in the *real world* farms were randomly distributed with respect to size.

A variation on this distortion arises because farms are not distributed randomly with respect to farm size. Large farms are often clustered together as are small farms. In cluster analysis the variation in the underlying population at risk is taken into account. No procedure is available to take into account variation in the distribution of farm size. Cluster detection methods based on point data may not be valid for this reason.

In the *real world* a farm has a fixed number of immediate neighbours and this varies from farm to farm. Analysis on point data can be based on nearest neighbour or *k*-nearest neighbours. This approach fixes the influence of each successive nearest neighbour for a farm. The process is not real and distorts the relationship between farms.

### **Sensitivity and specificity of the representation of contiguity or proximity**

The proximity matrix from a point coverage can be compared to that from the corresponding polygon coverage and the number of farms correctly or incorrectly classified as neighbours can be enumerated. The polygon coverage could be considered the *gold standard*. From this comparison the sensitivity and specificity of the point coverage representation of neighbours can be calculated.

### **Improving the representation of contiguity or proximity**

In this dissertation adjustments were made to point data methods to improve the classification of neighbours (and hence the proximity matrix). The simplest adjustment is to add the *radius* of each farm to any distance processes. The radius reflects the size of the farm it represents. Textual data on farm size can be used to calculate the *radius* of each farm. Hypothetically the distance between the centre of two adjacent circles that touch each other is the sum of the two radii. Treating farms as circles or varying size may be an improvement of treating them as points. This would help resolve difficulties arising out of differences in farm size and clustering of farms of similar sizes. The effect on contiguity can be assessed by calculating the sensitivity and specificity of the representation.

### **Summary**

Spatial data was found to be useful for the management of sentinel surveillance, detection of disease clusters, modelling disease spread, and visualising textual data. For these purposes farms need to be represented as polygons. Area data inherently contains the spatial characteristics of interest. The maps comparing polygon and point data speak for themselves.

## References

- Alexander, F.E. and J. Cuzick. 1992. Methods for the assessment of disease clusters. *Geographical and environmental Epidemiology: Methods for small-area Studies*. Editors P.Elliott, J. Cuzick, D. English, and R. Stern, 238-50. 382. Oxford: Oxford University Press.
- Animal Health Board (1996) *Annual Report for the year ending 30 June 1996*. Animal Health Board PO Box 3412 Wellington. Phone 04 4722858.
- Animal Health Board (1997) *Annual Report for the year ending 30 June 1997*. Animal Health Board. PO Box 3412 Wellington. Phone 04 4722858
- Anselin, L. 1995. Local indicators of spatial association - LISA. *Geographical Analysis* 27, no.2:93-115.
- Anselin, L., R.F. Dodson, and S. Hudak. 1993. Linking GIS and spatial data analysis in practice. *Geographical Analysis* 1: 3-23.
- Besag, J.E., 1986. On the statistical analysis of Dirty Pictures, *J. Royal. Stat. Soc. (B)*, 48, 259-279.
- Bailey, T.C. and A.C. Gatrell, 1995. *Interactive spatial data analysis*. Harlow, Essex, England: Longman Group. 413pp
- Carter, CE. (1994) *Use of TB tests within the NZ deer TB control programme*. Proceedings of a deer course for Veterinarians. No 11. Deer Branch or the NZ Veterinary Association.
- Carter CB, (1984) Corrin KC, de Lisle GW. *Evaluation of the tuberculin test in deer*. Proceedings of a deer course for Veterinarians. pp 1-8, Palmerston North, Deer branch of the NZ Veterinary association.
- Cressie, N.A.C. 1984. Towards Resistant Geostatistics, *Geostatistics for Natural Resources Characterisation, Part I*, Verly, G., David, M., Journel, A.G. and Marechal, A. (Eds.), Dordrecht, Reidel, pp. 21-44.
- Cressie, J.A.C. and Read, T.R.C., 1989. Spatial data analysis of regional counts, *Biometrical Journal*, 31, 699-719.
- Cuzick, J., and R. Edwards. 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society B* 52, no. 1: 73-104.
- Diggle, P.J. 1985. A Kernel Method for smoothing point process data, *J.R. Stat. Soc. (C)*, 34, 138-147.
- Dorling, D. 1994. Cartograms for visualising human geography, in Hearnshaw, H.J. and Unwin, D.J. (Eds) *Visualising in geographical Information Systems*, John Wiley, Chichester.
- Getis, A. and J.K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24 (3): 189-206.
- Glick, B., 1979. Tests for space-time clustering used in cancer research, *Geographical Analysis*, 11(2), 202-207.
- Griffin JFT, Chinn DN, Mackintosh CG, Rogers CR, Buchan GS. 1993. *Serial Elisa and skin testing in deer for diagnosis of Mycobacterium bovis infection*. Proceedings of a deer course for Veterinarians. pp 288 - 296, Palmerston North, Deer Branch or the NZ Veterinary Association.
- Isaaks, E.H. and Srivastava, R.M. 1989. *An Introduction to Applied Geostatistics*, Oxford: Oxford University Press.
- Kingham, S.P., A.C. Gatrell, and B. Rowlingson, 1995. Testing for clustering of health events within a geographical information system framework. *Environment and Planning A* 27: 809-21.
- Knox, E.G. 1964. The detection of space-time interaction. *Applied Statistics*. 13: 25-29.
- Mackereth G, 1993. The Waipawa endemic area: The epidemiological picture. Deer branch course No 10, *Proceedings for a deer course for veterinarians*. pp 222- 228, July 6 1993
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*. 27(2): 209-20.
- Mittlbock M and Schemper M (1996) *Explained variation for logistic regression*. Statistics in Medicine, Vol 15, pp 1987 - 1997.
- National TB Strategy (1995) *Proposed National Pest Management Strategy for Bovine Tuberculosis*. Animal Health Board. PO Box 3412 Wellington. Phone 04 4722858.
- O'Kelly, M.E., 1992. A clustering approach to the planar hub location problem, *Annals of Operations Research*, 40, 339-53.
- Oliver, M.A., and R. Webster, 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems* 4(3): 313-32.

- Openshaw, S. and Taylor, P.J. 1979. A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In *Statistical Applications in Spatial Sciences*, N Wrigley, ed , Pion , London, 127-144.
- Pfeiffer DU 1994. *The role of a wildlife reservoir in the epidemiology of bovine tuberculosis*, unpublished Ph.D. Thesis, Massey University, New Zealand.
- Ripley, B.D. 1977. Modelling spatial patterns (with discussion), *J. Royal Stat. Soc. (B)*, 39, 172-212.
- Ryan TJ 1993. *The New Zealand National Livestock Database*. Epidemiology proceedings of the Australian Veterinary Association Conference.
- Sanson R 1993. *The development of a decision support system for an animal disease emergency*. Unpublished Ph.D. thesis 1993 Massey University.
- Simes PJ 1986 An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73:751-754.
- Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

## **Appendix 1**

**A cross sectional survey of cattle and deer movements in  
the Hawke's Bay and Wairarapa Regions of New Zealand**



## **Abstract**

Movements of stock into and out of cattle and deer herds were investigated using a mail survey of cattle and deer herds in the Hawke's Bay and Wairarapa Regions of New Zealand. The survey quantified the number of deer or cattle, purchased, sold, slaughtered, born and present during the year commencing 1 July 1995. The information gained was combined with bovine tuberculosis (TB) testing data and used to assess the contribution that slaughter surveillance makes to TB surveillance. Slaughter surveillance was shown to represent a significant and unique component of TB surveillance.

Slaughter surveillance represented 71 % of total TB surveillance in cattle in the triennial testing areas of the Hawke's Bay and 25% of surveillance in the annual testing areas in the Wairarapa. In the study deer herds TB was mainly detected through slaughter, in the cattle herds it was mainly detected at test.

The herd classification system used in the TB control program was compared to the herd composition found at survey. The TB control program makes assumptions about the behaviour of different herd types. The survey enabled these assumptions to be evaluated. Inspection at slaughter for TB is thought to have a lower sensitivity than the TB test.

In the Wairarapa study herds the rate of deer with TB at routine slaughter was 14 times the rate at test. This raised concern about the performance of the deer TB test in field conditions.

## **Introduction**

This paper presents the results of a cross sectional survey of herds in the Wairarapa and Hawke's Bay regions of New Zealand. The study was undertaken to determine the contribution of slaughter surveillance to the TB control program and to examine the accuracy of the herd classification system currently in use.

The TB control program uses two methods of surveillance. The TB testing program and the examination of stock during routine slaughter. The contribution of slaughter surveillance is not known for any given area. National statistics show that the percentage of routinely slaughtered deer with TB is greater than the percentage detected through testing (1995-1996 and 1996-1997 Animal Health Board annual reports). To examine the contribution of slaughter surveillance to the TB control program the number of deer and cattle routinely slaughtered from each herd is required. This information is not available from the NLDB. This information was collected by survey and used to compare the TB testing program with the inspection of stock at slaughter.

Each herd is classified in the NLDB as either a breeding or fattening herd. Different TB control policies apply to each herd type. These policies are based on assumptions made about the type of animals present and assumptions about the amount of movements associated with the herd. The relationship between the herd classification system and the movement of stock was examined. The relationship between the herd classification and the types of animal present in the herd was examined.

The Animal Health board is planning to introduce a permanent identification system for cattle and deer. Information on the movement of stock was required to assess the logistics of a permanent identification system.

## **Materials and Methods**

The study was conducted as a postal survey. A random sample of 2020 cattle and deer herds stratified by region and herd type was selected. Each farmer included in the sample was sent a personalised letter and a questionnaire. Of 1002 returned questionnaires, 997

were used in the data analysis. There were 50 herds that did not have any cattle or deer at the time of the survey. Eighty-seven questionnaires were returned with an incorrect address by the post office. Table 16 shows the sample frame, the sample, and the number returned. The questionnaire asked for information about the number of deer or cattle purchased, sold, slaughtered, and born during the year commencing 1 July 1995. The number of deer or cattle present at the end of that year was also requested. Farmers were asked to differentiate between fattening and breeding stock. For the purpose of the study breeding cattle included all potential replacements, breeding cows, breeding heifers, cull cows, empty cows and breeding bulls. Fattening cattle were all cattle bred or purchased and fattened for sale or slaughter.

## **Results**

### **Animal movements**

The mean and median number of deer and cattle purchased, born, slaughtered, or sold to market in the year ending 30 June 1996 are shown in Table 22 with respect to region and category of herd type. The herd types are derived from the National Livestock Database. The herd type "Beef Drystock" refers to herds that fatten cattle for slaughter or sale. The number of cattle or deer present at the end of the study period is also shown. For most herd types the stock movements were in balance. Approximately the same number of stock were sold or slaughtered as were born or purchased. The exception was dairy herds, where the practice of grazing replacement stock off farm produced an apparent imbalance. Overall there was a slight increase (4%) in stock during the year of the study. The movements of animals on and off relative to the number present at June 30 was similar in both regions even though the number kept over winter in Hawke's Bay was considerably higher.

In herds classified under the TB control program as beef breeding herds, the equivalent of 40% to 45% of the number present at June 30 left the herd during the previous 12 months and the loss was compensated by purchases (12%) and new births (33%). Farmers were asked to differentiate between stock kept primarily for breeding and stock kept primarily for fattening. The percentage of the total herd size during the winter period that was kept for fattening purposes is shown in Table 22. The results suggest that the classification of herds into breeding and fattening categories does not reflect the true composition of herds, with some fattening herds having significant numbers of breeding animals, and some breeding herds having significant numbers of fattening animals. In both regions 8% of herds classified as beef drystock had breeding animals. Just over half of the beef breeding herds could be considered closed herds, with 58 % purchasing less than 10 animals during the study year. Figure 27 shows the purchasing pattern for these herd types. Dairy herds were similar to deer herds with regard to their purchasing behaviour. Approximately 80% of deer breeding and dairy herds purchased less than 10 animals during the year of the study.

### **Relative contribution of slaughter and TB testing to total surveillance**

Table 23 shows the relative contribution of slaughter and TB testing to total TB surveillance. The table presents the data stratified by region and by test frequency. All study herds in the Wairarapa and some herds in the Hawke's Bay were in an annual testing zone. Table 23 shows the percentage of total surveillance (testing and routine slaughter) classified as slaughter surveillance. Slaughter surveillance represents 71 % of total TB surveillance in cattle in the triennial testing areas of the Hawke's Bay and 25% of surveillance in the Wairarapa. Table 23 also shows the percentage of cattle herds in

the study with no testing during the 12-month period. In the Hawke's Bay triennial testing area this was 75% and in the Wairarapa it was 13%. Forty five percent of Hawke's Bay deer herds in surveillance areas did not test that year. The percentage of herds that did not slaughter any livestock is shown in Table 23. The percentage without any slaughter or testing is also shown. Slaughter does represent one component of total surveillance on most of the farms where testing is not undertaken. Only 7% of herds in the annual testing zone in Hawke's Bay had no surveillance at all in the year of the study. In the Wairarapa, 5% of cattle herds and 14% of deer herds had no surveillance at all in the year of the study.

### **Comparing disease detection rates at test and at slaughter**

The number of TB test positives and TB cases detected through routine meat inspection for the herds in the study are shown in Table 24. A TB test positive is an animal that reacts to a diagnostic test for *Mycobacterium bovis* infection such as the tuberculin skin test. Such an animal may undergo additional tests. Test positives that are slaughtered are examined for the presence of tuberculous lesions. Those that do not have any visible signs of TB may be in the early stages of infection, may be recovered cases or may be false-positive reactors.

In the Hawke's Bay study herds, only 3 animals had TB lesions during the year of the study. In the Wairarapa study, there were 185 with TB lesions. The difference in disease levels between the regions is due to the presence of infected TB vectors in the Wairarapa (Pfeiffer 1994).

Table 24 shows the percentage of tested animals and the percentage of routinely slaughtered animals that have lesions at post mortem. It is reasonable to expect that the test will lead to the detection of more animals with TB lesions than the routine slaughter of animals for market. In areas under annual testing such as the Wairarapa the test was less effective at revealing lesioned TB cases than routine slaughter. The test detected lesioned cattle in the Wairarapa at 65% of the frequency that routine slaughter detected lesions in cattle. Testing found 68 to 74% of TB cattle in the Wairarapa.

For deer herds, the detection probability through testing was 14 times lower than through routine slaughter. This finding prompted examination of national figures presented in the 1995-1996 and 1996-1997 Animal Health Board annual reports. The detection probabilities at a national level are shown in Table 24. The national statistics show that in deer the rate of TB at slaughter is twice the rate at testing. The diagnostic test sensitivity has been estimated to be 80 to 85% (Carter et al, 1994, Griffin et al, 1993). The sensitivity of meat inspection has been reported to be between 53 and 84% (Carter CE, 1994). On properties under annual testing it is expected that the testing program would find and remove TB deer. For the 41 deer breeding herds included in this study the testing program detected only 16% of lesioned animals. If all test positives that were slaughtered are considered tuberculous then the test detected 23% of all TB cases. If every test positive was considered tuberculous, then the test detected 34% of all TB cases.

Dairy herds were the only herds where the number of stock born, purchased, sold and slaughtered showed a major imbalance. The imbalance was consistent between regions. This was probably related to the practice of grazing young stock off the property. The dairy farms were smaller and more intensive than the beef farms. Analysis of all dairy farms in the Wairarapa showed that unlike other herd types most TB cases occurred in young animals. This suggests a different pattern of disease infection for these herds. It may be that young stock grazed off the property are more at risk than the adult stock grazed on the relatively intensive home farm. Dairy farms are essentially closed herds

grazing on intensive improved pasture. Figure 29 shows that with respect to purchasing behaviour dairy herds behave more like deer breeding herds than beef breeding herds.

## Discussion

About 4% of all questionnaires mailed were returned by the post office as not deliverable suggesting some redundancy in the National Livestock Database. The percentage of questionnaires returned was similar for the different herd types.

The presence of breeding stock in a herd results in a breeding or miscellaneous herd classification. The 8% of fattening herds that had breeding stock should be reclassified.

Annual checking of all drystock herds is undertaken to determine if the herd type classification should be changed. Many breeding herds were involved to a considerable extent in fattening and there was some breeding activity in fattening herds.

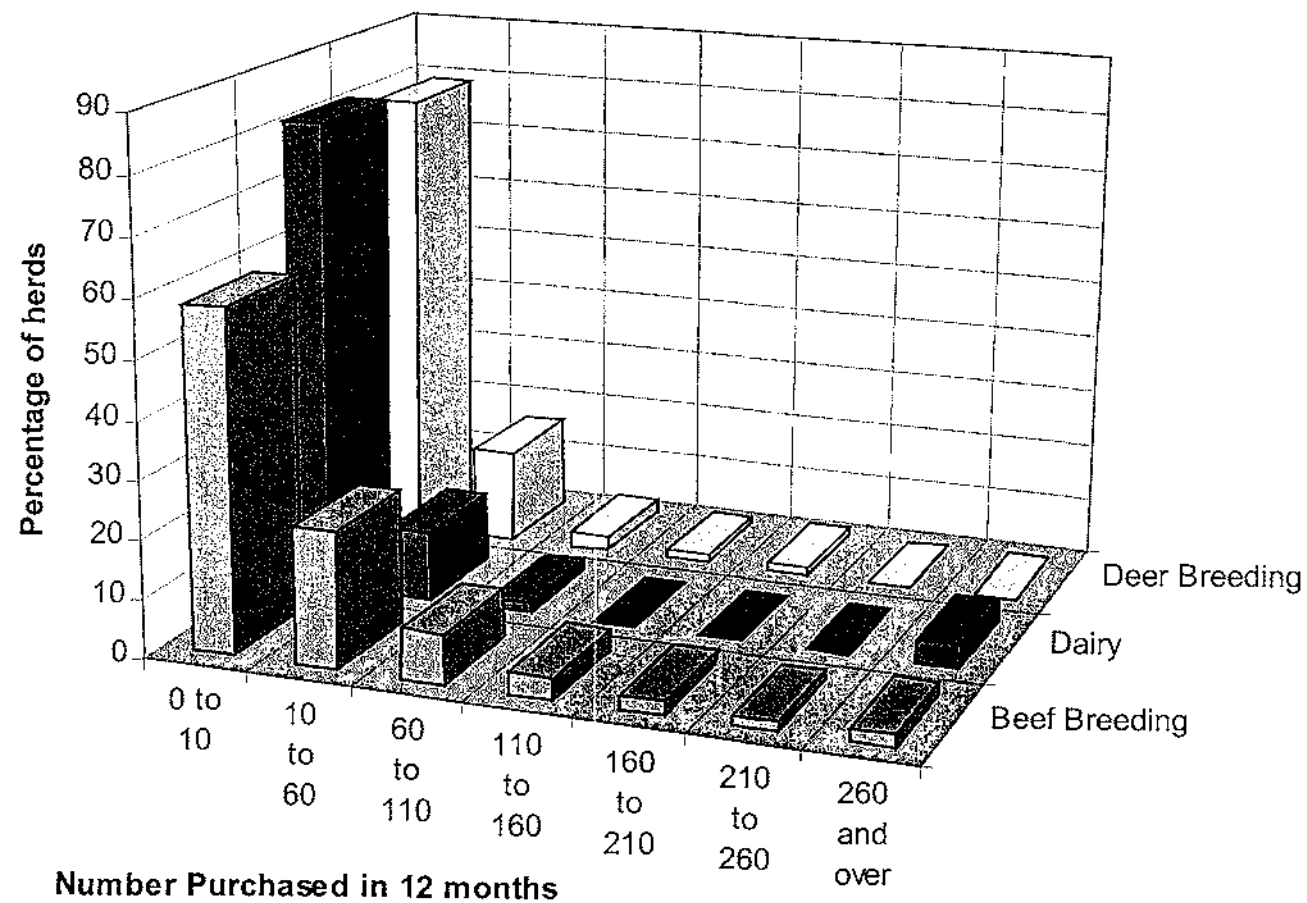
The turnover of animals from breeding herds has implications for the TB testing program. In clear areas breeding cattle herds are tested every three years and only 250 animals are tested. One of the assumptions of this testing strategy is that breeding herds have little trading activity. The assumption is valid for 60% of breeding herds. Figure 29 shows that 60% of beef breeding herds purchase very little. The herd classification system could be modified to take into account purchase and sale information. The number of deer or cattle purchased and sold each year would have to be recorded in the NLDB.

The different pattern of infection shown in dairy herds, their different purchasing behaviour, and the smaller more intensive farm size, does suggest that the epidemiology of TB in dairy herds is different to that in beef herds. At present, dairy herds are subject to the same control procedures as beef breeding herds. If dairy herds could prevent young stock from becoming infected while grazing elsewhere, then progress may be made in the control of TB in the dairy industry. If age and sex specific data at a herd level was collected for all dairy animals slaughtered, a more accurate analysis of the patterns of TB infection in dairy herds would be possible.

The results show that routine inspection of cattle and deer at slaughter represented a significant proportion of TB surveillance and that a significant proportion of TB cases were detected at slaughter. Some of the routine examinations at slaughter will be on animals that have tested negative. These animals will therefore have been doubly counted in the calculation of total TB surveillance in this paper. As the number of animals slaughtered is not recorded in the NLDB it has not been possible to compare the performance of these two detection methods. The study indicated that the two methods were detecting disease in different stock classes. As most cases found through routine slaughter inspection were over 2 years old and nearly half were female, it can be assumed that they have passed through the testing program.

Slaughter surveillance forms a significant part of TB surveillance. Each herd in the NLDB is inspected in some form every year. Consideration should be given to the collection and recording of slaughter and purchase information on an annual basis. The information is potentially useful in the design of the TB testing program for a herd and area and in improving our understanding of the epidemiology of TB and thereby the development of more effective control strategies.

Figure 29: Bar chart of cattle purchase patterns by herd type





**Table 22: Cross-tabulation of number of deer or cattle purchased, born, slaughtered, or sold to market in the year ending 30 June 1996 stratified by herd type and region**

Region	Herd Type*	Sample details			Stock Purchased that year			Stock Born on the property that year			Stock Slaughtered that year			Stock Sold to market that year			Stock Present on the 30/06/1996			
		Frame	Sample	Replies	Media	Mean	se~	Median	Mean	se	Median	Mean	se	Media	Mean	se	Median	Mean	se	Percent^ Fattening
Hawke's Bay	Beef Miscellaneous	241	154	74	0	31	12	1	6	2	1	21	9	0	13	5	9	52	15	42
	Beef Breeding	1243	336	174	3	37	5	81	115	10	48	86	8	20	53	7	233	349	29	18
	Beef Drystock	895	262	139	39	109	16	0	11	4	39	117	16	0	27	7	86	215	34	56
	Dairy Herds	54	47	19	0	54	34	270	320	75	32	55	10	4	39	14	359	451	92	3
	Deer Breeding	310	112	40	1	20	6	76	110	17	55	68	12	0	28	9	292	322	41	21
	Deer Miscellaneous	62	18	6	27	47	31	0	12	12	46	64	30	0	0	0	35	41	13	39
	Deer Velvetting	48	20	9	0	29	16	0	7	7	5	38	27	0	8	5	174	269	85	0
	All Herds	2853	949	461	3	57	6	6	71	6	32	81	6	0	35	4	148	257	17	
Wairarapa	Beef Miscellaneous	851	183	106	0	4	1	1	3	1	1	8	3	0	2	1	6	12	2	84
	Beef Breeding	813	373	202	2	33	5	54	85	7	40	77	8	9	35	5	154	243	20	24
	Beef Drystock	963	336	148	4	47	16	0	2	1	6	52	21	0	11	3	14	66	20	84
	Dairy Herds	264	59	30	0	2	1	175	175	18	35	41	6	4	16	5	247	266	22	1
	Deer Breeding	141	96	41	0	5	3	20	76	19	13	61	14	0	23	11	60	175	38	21
	Deer Miscellaneous	15	12	5	45	35	15	0	13	13	0	30	19	0	31	31	45	67	39	67
	Deer Velvetting	13	12	4	0	63	63	0	14	14	2	42	41	0	0	0	30	163	140	38
	All Herds	3060	1071	536	1	28	5	3	49	4	10	53	7	0	20	2	41	142	11	
Total	All Herds	5913	2020	997	2	41	4	4	59	4	20	66	5	0	27	2	77	195	10	

\* Herd Type refers to the classification of the Herd in the National Disease Database.

^ Percent fattening refers to the percentage of animals in the herd at the 30th of June that the farmer considered to be farmed for fattening rather than breeding purposes.

~ se is the standard error for the mean

**Table 23: Cross-tabulation of the contribution of slaughter surveillance to the total TB surveillance in the year ending 30 June 1996, stratified by herd type and region based on herds included in this study.**

Region	Herd Type	Number of herds in the study		Number of animals tested in the 1995-1996 year		Number of animals slaughtered in the 1995-1996 year		Percentage of total surveillance* that is slaughter surveillance		Percentage of herds that did not test for TB in the 1995-1996 year		Percentage of herds that did not slaughter in the 1995-1996 year		Percentage of herds with no surveillance at all in the 1995-1996 year	
		Annual test zone	Triennial test zone	Annual test zone	Triennial test zone	Annual test zone	Triennial test zone	Annual test zone	Triennial test zone	Annual test zone	Triennial Test Zone	Annual test zone	Triennial test zone	Annual test zone	Triennial test zone
Hawke's Bay	Beef Miscellaneous	21	53	229	126	714	858	76	87	5	72	33	49	0	36
	Beef Breeding	51	123	10150	7751	4295	10651	30	58	2	63	16	8	2	8
	Beef Drystock	45	94	646	156	6496	9784	91	98	69	95	24	21	16	19
	Dairy Herds	5	14	1519	1079	205	845	12	44	0	57	0	0	0	0
	Cattle Herds	122	284	12544	9112	11710	22138	48	71	27	75	21	20	7	17
Wairarapa	Deer Breeding	9	31	1734	2494	402	2026	19	45	27	45	33	28	7	13
	Beef Miscellaneous	107	0	1350	0	887	0	40	-	20	-	37	-	9	-
	Beef Breeding	199	0	56761	0	15330	0	21	-	5	-	13	-	1	-
	Beef Drystock	148	0	6440	0	7686	0	54	-	24	-	30	-	9	-
	Dairy Herds	30	0	11623	0	1236	0	10	-	0	-	3	-	0	-
	Cattle Herds	484	0	76174	0	25139	0	25	-	13	-	23	-	5	-
	Deer Breeding	41	0	8622	0	2776	0	24	-	20	-	30	-	14	-

\* Total Surveillance is simply the sum of the number of tests and the number slaughtered.

**Table 24: Cross-tabulation of risks of TB detection based on testing and slaughter surveillance stratified by region and herd type**

Region	Herd Type	Number Tested that year	Number Slaughtered that year	Test positives that are slaughtered	Test positives with TB lesions at post mortem	TB cases found at routine meat inspection	% of animals tested that are slaughtered	% Test positives with TB lesions at post mortem	TB cases found at routine meat inspection per 100 killed	Percentage of TB cases* found by testing	Relative Risk of lesion detection by testing versus routine slaughter
Hawke's Bay	Miscellaneous	276	1572	1	0	0	0.362	0.000	0.000		
	Beef Breeding	18072	14948	3	1	0	0.017	0.006	0.000		
	Beef Drystock	821	16280	0	0	1	0.000	0.000	0.006		
	Dairy Herds	2781	1014	0	0	0	0.000	0.000	0.000		
	All Cattle	21950	33814	4	1	1	0.018	0.005	0.003	50 to 80	
	Deer Breeding	4655	2457	6	0	1	0.129	0.000	0.041		
	Miscellaneous Velvetting	112	382	0	0	0					
		1744	341	4	0	0	0.229	0.000	0.000		
	All deer	6511	3180	10	0	1	0.154	0.000	0.031	0 to 91	
Wairarapa	Miscellaneous	1245	887	8	7	2	0.643	0.562	0.225	77 to 80	2.494
	Beef Breeding	56802	15619	113	82	32	0.199	0.144	0.205	72 to 78	0.705
	Beef Drystock	8839	7715	9	5	14	0.102	0.057	0.181	26 to 39	0.312
	Dairy Herds	11702	1236	10	6	2	0.085	0.051	0.162	75 to 83	0.317
	All Cattle	78588	25457	140	100	50	0.178	0.127	0.196	68 to 74	0.648
	Deer Breeding	8738	2776	9	6	29	0.103	0.069	1.045	17 to 23	0.066
	Miscellaneous Velvetting	155	261	0	0	0					
		637	166	0	0	0					
	All deer	9530	3203	9	6	29	0.094	0.063	0.905	17 to 23	0.070
NZ 1995 - 1996	Cattle	4150000	2500000	4727	2220	994	0.114	0.053	0.040	69 to 83	1.35
	Deer	451892	323215	1315	171	239	0.291	0.038	0.074	42 to 85	0.51
NZ 1996 - 1997	Cattle	4560000	2390000	4488	1968	889	0.098	0.043	0.037	69 to 83	1.16
	Deer	527643	281550	1387	224	224	0.263	0.042	0.080	50 to 86	0.53

The percentage of cases found by the testing program is presented as a range corresponding to two case definitions. If only lesioned animals are considered to be cases the lower percentage applies. If all reactors are considered cases the upper percentage applies.