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.

Massey University Library
New Zealand & Pacific Collection

GEOGRAPHICAL REPORTING AND ANALYSIS OF INFECTIOUS ANIMAL DISEASE OCCURRENCE IN THAILAND AND NEW ZEALAND

A thesis presented in partial fulfilment of the requirements
for the degree of Master of Veterinary Science
at Massey University

Tippawon Teekayuwat

May, 1999

Abstract

A comprehensive geographical study and reporting system is presented. Animal disease data from both Thailand and New Zealand were explored and analysed using spatial analysis methods. The particular technique used depended on the form of the data, aim of the investigation and the epidemiology of the disease of interest. Results and methods from some of these analyses were then included in the development of *a simple geographical disease reporting and analysis system for Thailand*.

A number of methods were used to investigate the presence of temporal clustering, spatial clustering and spatio-temporal clustering of foot and mouth disease (FMD) reporting data in Thailand during January 1995 to May 1997. Temporal clustering in the time series of individual districts and individual provinces was found in many districts and provinces. Some of these provinces also showed the evidence of unimodal patterns. Spatial clustering was detected both at the district and province level. Space-time clustering was found at the district level.

An exploratory analytical approach was used to investigate spatial clustering of bovine leukaemia virus (BLV) infection in New Zealand dairy herds. Two spatial clusters of BLV positive herds were detected in the Bay of Plenty area (p =0.001) and in the northern part of the South Island (p = 0.082). We recommend that further investigations be conducted to define possible reasons for the presence of these observed clusters.

The geographical patterns of FMD were described and risk maps of FMD outbreak occurrence in Thailand were developed using logistic regression and classification tree models (CART). The potential impact of spatial autocorrelation on the logistic regression models was assessed. CART models incorporating cost-sensitivity were constructed to develop sets of decision rules for the likelihood of FMD outbreak occurrence. Receiver-operating characteristic (ROC) curves were used to quantify and compare the value of the different models for production of risk maps and to provide a method for decision makers allowing them to optimise sensitivity and specificity of binary decision criteria.

A simple geographical disease reporting and analysis system for Thailand was developed using the GIS software ArcView 3.1®, the database management software Microsoft Access 97® and the spatial cluster analysis software SaTScan® version 2.1.3.

The programming language Avenue™ which is part of ArcView 3.1 was used bind the different components using a common user interface. The system allows quick and easy production of custom maps for routine reporting as the system is largely automated and requires only basic computer skills from the operator.

Acknowledgements

It's my pleasure to acknowledge all the people who have contributed to my study

Without the experience, wisdom, careful attention and patient guidance of my chief supervisor during these two years, Dr. Dirk Pfeiffer, this work could not have been completed. Dr Pfeiffer allowed, encouraged and even pushed me to be myself and in turn to discover my own potential - my heartfelt thanks go to him.

My deep gratitude also to Dr. Ron Jackson, my co-supervisor, who kindly provided help and guidance and assisted in more than proof-reading the many drafts of this thesis.

I would also like to thank Professor Roger Morris, a second co-supervisor, who invited me to study at Massey University and become a member of the EpiCentre - without the help of Professor. Morris I could not have come to New Zealand for this postgraduate opportunity.

In addition to my formal supervisors, I have would like acknowledge the assistance provided by Fiona Dickinson, Daniel Russell, Dr. David Lawton, John Lockhart and Craig Deadman in the proof reading of early drafts of this thesis.

On a personal note, thanks to my office-mate, Dr. Klim Hüttner, for the relaxed atmosphere he brought to our room, and to all the students and staff of EpiCentre for their warm hospitality and friendships during my study.

I offer my special thanks to my Thai friends, both in New Zealand and in Thailand - in particular, to Kanda Komolwatanachai, Somrutai Malaipong, U-sa Chethanond and Kaewkwan Parakamawongsa for their warm encouragement and supports. I also thank my colleagues in the Department of Livestock Development, Thailand for their assistance and encouragement.

Most of all, special thanks go to my parents in Thailand.

Table of Contents

List of Tables

List of Figures

CHAPTER 1.

General Introduction

Epidemiology is concerned with identifying patterns in the distribution of disease in populations, using these patterns to identify causes of the disease, and then to develop appropriate strategies for control or eradication of the disease. Patterns in the occurrence of a disease may be detected by categorising afflicted and un-afflicted individuals according to measurements made for a group of variables. Using this approach patterns appear when the causes of disease are distributed unevenly as reflected in these variables (Rothman, 1990). The spatial distribution of a disease can be studied if the variables being investigated have locational attributes. Recent developments in data management software and statistical methodology have made the interpretation of spatial aspects of disease information much more accessible to veterinary epidemiologists.

Spatial Data Analysis

Different methods are used in spatial data analysis depending on the type of spatial data being analysed: point patterns, geographical data, and lattice or areal data (Pfeiffer, *et* al., 1994). Spatial data analysis can be broadly categorised according to the following general categories: visualisation, exploration and modelling (Bailey and Gatrell, 1995).

Spatial data visualisation

The visual display of spatial phenomena provides a very effective descriptive analytical tool. Visualising spatial data does really mean mapping. Mapping techniques come in various forms and are useful for different kinds of spatial data. Three important characteristics of maps include: 1) Are they based on point locations or region locations? 2) Do they show regular or irregular patterns? 3) Is the variable being mapped continuous or discrete? (Hertz-Picciotto, 1998) Visual map displays can provide basic information for generating hypotheses. However, choices of map type

and scaling used for data values can lead to misleading conclusions being drawn from the display and can suggest the development of inappropriate hypotheses. The advances in GIS software development and modern data-capture technology allow the investigator to create maps and explore spatial patterns and relationships quickly and interactively.

Spatial data exploration

Exploratory analysis of spatial data is used to assist in the development of hypotheses, or search for previously unseen processes. Exploratory techniques represent an intermediate process where the distinction between visualisation and exploration is not as well defined as between spatial data exploration and modelling. The dividing line between visualisation and exploration depends on the extent of data manipulation which the analytical method requires, while the specification or restrictions of the model are used to distinguish between exploration and modelling. Simple analytical models can be used in this exploratory analytical phase.

Spatial data modelling

Specific types of hypotheses are formally tested or evaluated using statistical models. It focuses upon the development of spatial models. Since the characteristics of spatial data often express some degree of spatial correlation, modelling of spatial data has to incorporate the possibility of these correlations into the models. In turn, these models provide alternative hypotheses against the null hypothesis of absence of spatial dependence or autocorrelation.

Application of Geographical Reporting and Analysis Systems in Animal Disease Control

A list of the suggested major components of animal health and productivity information systems for veterinary services in developing countries has been compiled as part of a an FAO expert consultation on the need for information systems aimed at strengthening veterinary services in developing countries (FAO, 1993). Specific elements within each of the four major components of an animal disease information system are listed in Table 1. Various comments were made in the report particularly with respect to spatial

aspects. The report recommended that during the process of data collection and reporting special consideration should be given to the accuracy of spatial referencing of disease and production data as well as environmental characteristics (geography, rainfall etc.) of the geographical area.

Table 1 The four major components of an animal health information system.

Major components	Elements
1. Data collection	▪ Livestock demographics ▪ Disease specific monitoring ▪ Laboratory diagnostic data ▪ Animal movement data
2. Data management	▪ Processing ▪ Analysis ▪ Interpretation ▪ Synthesis
3. Information distribution	▪ Routine disease reports ▪ Specific disease reports ▪ International disease reports
4. Utilisation	▪ Feedback to data suppliers ▪ Monitoring use of information ▪ Monitoring progress of the disease ▪ Monitoring progress of the control or eradication program

Once a geographical reference has been added to the data incorporated in the information system, a geographic information system (GIS) can be used for spatial data management and presentation, and thereby the spatial aspect of animal health can be incorporated in the reporting and analysis of animal health data. Spatial analysis and GIS are often considered as integrated procedures or systems. Haining (1994) concluded that GIS provides a data management tool which may result in statistical spatial data analysis becoming more accessible to the investigator.

As part of epidemiological field studies, GIS is used for the visual display of geographical patterns and for spatial analysis. In disease surveillance, GIS can be used at the most basic level to generate up-to-date maps of the spatial patterns of disease

occurrence and at the more advanced level as part of sophisticated animal disease information systems (Pfeiffer, *et* al., 1994).

The power and flexibility of a GIS can provide a useful tool when working towards the objective of animal health improvement by providing better information to decision-makers. Knowledge of the geographical component of disease distribution has the ability to enhance most of the functions of an animal health information system and to provide entirely new areas of functionality.

The objective of this thesis was to develop a simple animal health information system for use in Thailand based on geographical and animal disease data. Specific emphasis was placed on developing a cluster detection component for this system. These analyses were conducted using disease surveillance data for Foot and Mouth Disease (FMD) occurrence recorded in Thailand between January 1995 to May 1997 and data on Enzootic Bovine Leucosis (EBL) herd infection status of dairy cattle herds for 1997-1998 from New Zealand. Using spatial data analysis and GIS, the spatial patterns of both diseases were investigated using spatial data visualisation and exploration techniques (chapters 2 and 3). Models of the spatial distribution of FMD were constructed in order to predict the occurrence of FMD outbreaks in Thailand (chapter 4). Since information distribution is one of the major objectives of animal health information systems, production of tailored reports is one of the most important activities. A simple geographical reporting system was developed using Avenue, the ArcView ® programming language (chapter 5). The system is aimed at producing reports of observed as well as predicted geographical patterns and performing spatial cluster detection.

Reference List

Bailey, T.C., Gatrell, A.C., 1995. Interactive spatial data analysis. Longman Scientific & Technical, New York, USA.

FAO, 1994. Report of the FAO expert consultation on the need for information systems to strengthen veterinary services in developing countries. Rome, Italy.

Haining, R., 1994. Designing spatial data analysis modules for geographical information systems. In: Spatial analysis and GIS. Fotheringham, A.S. and Rogerson, P.(ed). Taylor & Francis, London; Washington, DC : pp. 45-63.

Hertz-Picciotto, I., 1998. Environmental Epidemiology. In: Modern Epidemiology. Rothman, K.J., and Greenland, S. (2nd ed). Lippincott - Raven Publishers, U.S.A : pp. 555-584

Pfeiffer, D.U., Morris, R.S., Sanson, R.L.,1994. Application of GIS in Animal Disease Control -Possibilities and Limits. In: Proceedings of WHO Consultation on Development and Application of Geographical Methods in the Epidemiology of Zoonoses on 30 May 1994-2 June 1994.

Rothman, K.J., 1990. A sobering start for the cluster busters' conference. American Journal of Epidemiology 132 (Suppl. 1): S6-13.

CHAPTER 2.

Spatial and Temporal Analysis for Detection of Clustering of Disease Occurrence

---

Introduction

Disease clustering

A disease cluster is an aggregation of cases in space, in time or in space-time. They occur when more cases appear in a particular place and time than would be usually expected. Knox (1988) considers a cluster as 'a bounded group of occurrences related to each other through some social or biological mechanism, or having a common relationship with some other event or circumstance'. One might want to enquire whether cases of a particular disease appear at excessively high frequency in certain years or one might suspect that certain forms of illness are more common amongst animals which are raised in certain areas, perhaps because of environmental factors peculiar to these places. Another type of disease cluster is the space-time cluster which occurs when pairs of cases are relatively close together in space as well as in time. Clusters are of interest for at least two reasons. Firstly, cluster investigations can yield insights into disease epidemiology. Secondly, and perhaps more importantly, cluster investigations are one of the main tasks of public health agencies as their detection as part of disease surveillance will lead to specific investigations (Jacquez, 1996).

Types of space -time patterns

As mentioned above, disease events occurring in space and time may exhibit three types of patterns; clustering in time, clustering in space, and clustering in space-time.

These patterns occasionally arise due to chance but often are the result of a space-time process. The possible types of pattern can be used to construct a classification matrix of

eight possible types (see Table 2) which can help to focus one's thinking regarding the type of pattern thought to underlie a suspected cluster.

Table 2. Classification matrix of possible combinations of types of clustering of disease occurrence

Clustered in time	*Clustered in space*	*Space-time clustering*
no	no	no
no	no	yes
no	yes	no
no	yes	yes
yes	no	no
yes	no	yes
yes	yes	no
yes	yes	yes

Units of analysis

The method to be used in an analysis of space and/or time clustering depends of the type of data to be analysed, point or area data. Point patterns are generated from data where the location of every case is being recorded. Typically this location will already involve some level of aggregation in that the occurrence of a disease in animals kept in a particular paddock on a farm is recorded using the coordinates of the farm centroid, instead of using the actual paddock boundaries. Typically, geographical data collected as part of disease surveillance programs is recorded using certain administrative units, such as district or provinces, which can presented spatially as area data. In the current analysis, data was recorded at the district level and therefore methods for analyses of area data were used.

Clustering in time

Temporal clustering without presence of spatial aggregation indicates that the incidence of disease fluctuates over time in a similar pattern in different places. It means that individual cases of disease occur closely together in time, as is often the case with infectious diseases. Cyclical clustering suggests the association between the presence of

an environmental risk factor occurring at cyclical intervals and the aetiology of the disease. Variability in time is more revealing over shorter than over longer time intervals, since with short intervals less factors will influence variability such as seasonal swings in disease frequency which point to seasonally varying environmental factors. On the other hand, secular variation over long periods may be explained by a wide variety of factors that change with time, including such diverse possibilities as changes in diagnostic practice and changes in gene pool, and may therefore be more difficult to account for correctly (Rothman, 1987).

A rather simple approach to many problems of this sort is to divide the time period into equal intervals, to express the incidence risk in each interval as a proportion and to test the significance of the difference between these proportions by standard 2 x n contingency table methods. It may be sensible to concentrate attention on the maximum of the various case counts, on the grounds that occasional clustering may affect only one or two of the time intervals. Time clustering can be investigated in a single time series or as simultaneous clustering in several time series. Multiple time series arise when disease rates or counts are recorded for several areas at a time. The fundamental question is to determine whether time clustering occurs in most or all of the time series.

Several methods have been developed to detect clustering based on different null hypotheses. Time clustering methods include the Ederer-Myer-Mantel test (Ederer, 1964), the empty cell test (Grimson, 1993), Larsen's test (Larsen, 1973), the 0-1 matrix test (Dat, 1982), Grimson's test for time clustering (Grimson, 1989), the Scan test (Wallenstein, 1980;1987), the Chen test (Chen, 1979), the Texas test (Hardy, 1990), the Poisson test (Edmonds, 1981), the Sets test (Chen, 1986), the Cusum test (Weatherall, 1976; Hill, 1968) and the Edward's test (Edwards, 1961). Several modifications of these tests have been developed (Tango, 1984; Wallenstein 1980; Hewitt, 1971; Roger, 1977; Cave, 1975).

Clustering in space

Space clustering represents geographic variation in event occurrence. The study of space clusters amounts to comparing disease occurrence in difference places. Spatial clustering occurring over a short time span can be the result of infectiousness. An infectious disease that propagates through a population results in a contagious spatial pattern. 'Contagion' can also be applied to the spatial clustering of disease, whether or

not being infectious. Ecologists sometimes use 'over-dispersion' to refer to this type of spatial clustering in a population.

Virtually every disease varies in occurrence from one place to another, resulting from the geographic variation that exists for all causes of disease, environmental and genetic. Spatial clusters can be explained by a multitude of possible factors, their number depending on the geographic distances taken into consideration. On a local scale, geographic variability within communities or small regions can focus attention on a narrow range of possible causal explanations and seemingly facilitate identification of factors that cause diseases or other underlying causes such as inadequate farm management or public health utilities. However, there are often numerous factors that vary in a similar pattern, requiring many possibilities to be eliminated before real etiologic insight is attained.

Although clustering is usually considered in terms of local high rates, the occurrence of foci of particularly low local rates or 'negative clusters' also has etiological significance. The statistical methods to be described were developed primarily to analyse and test for positive clustering. Their application and ability to test for negative clusters has not been explored (Marshall, 1991).

Two mechanisms of spatial clustering that are of epidemiological interest are:

- a locally elevated risk so that the population in the locality is, independently, subjected to greater risk than elsewhere and
- spatial interaction, i.e. local high rates of transmission from animal to animal of an infective agent or a genetic abnormality

Statistical testing for clustering is generally aimed at describing a general tendency for clustering to occur and, if so, where. The latter issue deals with whether clusters occur in specific areas or not, e.g. near suspected environmental hazards. Testing for 'significance' of a cluster may in itself be viewed as being of little value unless accompanied by a meaningful statistic saying something about the pattern of disease occurrence. It will usually be convenient to subdivide the total population into administrative areas containing quite different numbers of individuals.

Many test statistics are of the form ' $T=\Sigma_i \Sigma_j x_{ij} y_{ij}$ ' where x_{ij} and y_{ij} are measures of similarity, or separation, of observational units i and j. Units are often geographical areas but they may also be individuals. Testing can be done by permutation tests or

point pattern analyses, without knowledge of population densities based on distance measures (Mantel, 1967; Ross, 1990; Kulldorff and Nagarwalla, 1995) or case-control status as in the nearest-neighbour-based Cuzick and Edwards' test (Cuzick and Edwards, 1990). Many adjacency statistics are also T-type statistics with a binary indicator of adjacency; x_{ij} and y_{ij}, a measure of concordance of rates or Poisson p-values in areas *i* and *j* in the Ohno test (Ohno, 1979), Grimson's test for space clustering (Grimson, 1991), Moran's I (Cliff, 1981) and Moran's I adjusted for population size (Oden, 1995).

Localised disease clustering near putative environmental hazards is a topical and sensitive issue with many methodological problems. These include the danger of making inferences from data used to generate the hypothesis, the difficulty of defining a null hypothesis, lack of statistical power, the arbitrary nature of statistical boundaries, the extent of the study area and the presence of extra-poisson variation. Although the chance occurrence of local clusters is to be expected and is sometimes used to play down the importance of observed clusters, the fact that the cluster and hazard coincide cannot be overlooked. A variety of methods based on distance of cases or areas from a source have been proposed (Marshall, 1991) as well as the increase in Poisson intensity with distance from sources (Lawson, 1988).

Space-time clustering

Space-time interaction arises when nearby cases occur at about the same time. Space-time aggregation, in which the disease rate varies with both time and place, is often extremely revealing, since explanations for such clusters are restricted to the limited set of factors that vary in the same specific pattern as the disease. Environmental agents that move from place to place or suddenly appear in specific locations, such as infectious organisms, toxic chemicals, or new drugs with local popularity cause space-time clustering. Many epidemiological investigations of clusters are of interest because the space-time pattern of cases may reflect risk factors and exposures underlying the disease.

Space-time clustering of disease outbreaks is defined as the occurrence of a pattern of cases of disease that are closer together in time and distance than could be expected if the only underlying mechanism governing the distribution of the disease were random allocation (Grimson, 1979). It is important to define appropriate units of space and time

before examining the data in detail. The resolution or level of aggregation of the units is likely to strongly influence the outcome of the analysis; for example, clusters that occur within a year may be missed if the study unit is one month. Similarly, clusters within a farm-block may be missed if the unit of study is a county. Clusters may remain undetected if the resolution chosen for the unit of study is too large or too small. In the absence of knowledge of the aetiology of the disease, it is difficult to select an appropriate statistic of clustering which will yield sufficient power. The payoff from clustering research comes from the specific hypotheses that emerge to explain the observed pattern of excess occurrence. If the research is limited to a specific cluster with only a few cases and a small relative increase in disease frequency, the prospects of useful etiologic information are dimmer.

A variety of tests for space-time clustering of disease have been developed. Some of the statistical methods require the computation of distances between each possible pair of cases (Knox, 1964; David, 1966; Mantel, 1967), or require labelling objects in space-time as adjacency areas and high-risk cells (Grimson, 1981; Symons, 1983), or use nearest neighbour measures (Jacquez, 1994; Cuzick and Edwards, 1990). The statistical power of different methods for detection of space-time clustering has been investigated by a number of researchers (Chen, 1984; Wartenberg and Greenberg, 1990; Oden, 1996). Lack of statistical sensitivity was suggested as the main cause of the low rate of space-time interaction found in cluster studies.

The purpose of this paper was to investigate whether there was clustering of Foot-and-Mouth-Disease occurrence in time, space or space-time in Thailand (Jan 1995 - May 1997) and to explore the use of different statistical methods for analysing disease clustering..

Materials and Methods

Foot-and-mouth-disease (FMD) reporting data for Thailand from January 1995 to May 1997 was analysed based on the levels of aggregation used by The Department of Livestock Development: district, province, region, and country. A total of 113 outbreaks of FMD were reported during this time period. Each outbreak was reported at the district level including the following information: month, FMD type (Type O, Asia1 and untyped), species affected, the number of new cases and the district where the outbreak occurred. Data was stored in Microsoft Access 97 and analysed using the

statistical software packages Stat!® (Biomedware, Ann Arbor, Michigan, USA) and SaTScan version 2.1.3 (National Cancer Institute, Bethesda, Maryland, USA). Geographical data was located using the GIS software ArcView® for Windows version 3.1 (Environmental Systems Research Institute, Redlands, California, USA). A p value of 0.05 or less was considered significant in all statistical analyses.

Recognising the relatively poor statistical power, one approach is to look at the similarities and differences of the methods and take into account which properties of the data they emphasise. Previous publications have presented a four-stage typology of cluster methods. The stages are 1) data type, 2) transformation type, 3) summary statistic and 4) reference methods (Wartenberg and Greenberg, 1993). According to these steps, the choice of the appropriate method still depends on the context of the study and the specific features of interest. Decision rules are necessary to determine whether perceived clusters of disease cases constitute actual clusters, or merely clumps in an overall pattern consistent with random distribution of cases. Hence, statistical techniques are used to distinguish the presence of clusters of cases from random allocation.

Methods for detection of clustering in time

Temporal cluster analysis was conducted using three methods: the scan method, Larsen's method, and the empty cells method. During the current analysis these techniques are applied to case counts for single time series and for several time series simultaneously. Cases were arranged for analysis according to FMD virus type using a total of 29 time cells (monthly from Jan 95 - May 97) aggregated at the province and district level. These methods detect clustering in time under the hypothesis that FMD cases are allocated at random amongst the time cells of each time series within a selected level of spatial aggregation (whole country, province or district). After aggregating cases by virus type into province and district time series, the data over this time period for virus type Asia1 for two provinces looks as shown below:

29 monthly time intervals																												*Province*
0	0	0	0	0	6	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	Chiengrai
0	0	0	8	0	24	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Nakhonsawan

Scan statistic

The scan statistic, a test for temporal clustering in a single time series, evaluates whether an apparent cluster of disease in time is due to chance. This method is designed to detect sudden temporal clustering of disease in a defined population. All cases of a disease in the study area are arrayed on a time line according to date of event. An interval or "window" of fixed length is then defined based on the expected duration of the epidemic. The statistic employs a 'moving window' and finds the maximum number of cases revealed through the window as it scans or slides over the entire time period. The test statistic is based on comparing the maximum number of cases in the interval, and its expectation. The test is most sensitive to clustering when the scanning window has the same width as natural clusters in the data (Wallenstein, 1987). The interval width of the scanning window for this study was set to 2 months, as this is the expected duration of FMD epidemic outbreaks and the data was limited to a resolution monthly time units.

Larsen's method

Larsen's test statistic K is sensitive to a unimodal clustering of occupied cells which occurs when occupied time cells tend to occur in a sequence. Unimodal clustering will cause K to be smaller than its expectation. Multiple clustering or uniformity (cluster avoidance) will cause K to be larger than its expectation. This module provides two tests for temporal clustering: within the individual selected level of spatial aggregation using a z-score and across all levels of spatial aggregation simultaneously using an overall z-score. Negative z-scores suggest unimodal-clustering, z-scores near zero are consistent with the null hypothesis, and positive z-scores arise under uniformity and multiple clusters. We used Larsen's method to consider the following questions:

- Within a district or province area, do time cells occupied with FMD cases tend to occur in a sequence?
- Is there an unusual pattern over time (29 months) which may not necessarily be the same for different district or province areas?

Empty cells method

The empty cells test is based on E, the number of empty cells or columns (time interval) in a sequence of consecutive time interval. Consideration of the statistic E is appropriate for rare data or if one or more of the time periods have several cases while

other time periods have none. If the observed value of E is significantly larger, the cases tend to cluster within relatively few columns (Grimson, 1993). To use this test, the number of cases must be small enough so that the expectation of the number of empty cells is greater than 1. We used the table of the maximum number of cases supported by the empty cells test provided in the manual of Stat! to determine whether this data can be analysed using the empty cells approach. The limitation of the maximum number of cases with 29 time cells is 95. All time series for provinces and districts have cases less than 95, so we proceeded to use this test for the province and district time series but not for the whole country series.

Methods for detection of clustering in space

Moran's I with adjustment for population density (Ipop)

Spatial cluster analysis for this study was performed using Moran's *I* adjusted for population size. This test was modified from Moran's *I* to adjust for spatial variation in the underlying population density. I_{pop} is large under two clustering scenarios. First, when cases cluster within areas and second, when areas with many cases are adjacent. The range of I_{pop} depends on population size, and it is useful to standardise I_{pop} (I_{pop}'). I_{pop} is more powerful than Moran's I (I) because I takes into account only the similarity in rates of neighbouring areas, while I_{pop} considers also the variance in rates across areas. Adjusting Moran's I is better suited to test the hypothesis of no spatial autocorrelation than other descriptors of spatial autocorrelation (Oden, 1995) as shown in studies of fox rabies in England and childhood leukaemia in North Humberside, England (Oden *et al*, 1996). Judging by the results derived from simulations, the statistical power of the I_{pop} test higher when compared with Cuzick-Edwards' test, Moran's I, Grimson's method using normal approximation or Poisson approximation (Oden *et al*, 1996). For the above reasons and given the availability of animal population statistics, we used the I_{pop} test for our FMD reporting data. Sixty-seven of 76 provinces and 593 of 844 districts were used to construct a connection matrix required for the calculation. The southern and eastern regions of Thailand which are declared FMD-free zones were excluded from the analysis.

The null hypothesis assumes that the probability of FMD cases of specific virus types occurring in a province or district is given by the proportion of the total animal population in that spatial unit.

Spatial scan statistic

The spatial scan statistic (Kulldorff, 1997) allows statistical inference with respect to the presence of spatial clustering as well as identification of the locations of any clusters.

This method generates for each location in the dataset a set of circles (=windows) with ever-increasing radius from zero to some upper limit throughout the study region. It performs a likelihood ratio test comparing on disease risk within and outside the circles. The null hypothesis distribution of the likelihood ratio is obtained on the basis of Monte Carlo replications. The 'most likely' cluster is defined by the highest value for a significant likelihood ratio statistic in the dataset. Other statistically significant clusters are also identified as 'secondary' clusters.

With the availability of animal population statistics for individual districts, the number of cases in each census area is assumed to be Poisson distributed. Under the null hypothesis, the expected number of cases in each area is proportional to the population size in that area. The method can scan for clusters of geographical size between zero and an upper limit defined by the investigator. This maximum is expressed as a percentage of the total population at risk, and the recommended maximum window size is 50% (Kulldorff and Nagarwalla, 1995). The test statistic was calculated based on 999 random replications.

Methods for detection of space-time clustering

Analysis of space-time clustering for this study was done using the two most widely adopted tests measuring the proximity of case pairs in space and time: the Knox test and the Mantel test. Other test used in this study were the k nearest neighbour test (k-NN) and the space-time scan statistic. Three independent methods were used because the power of techniques for space-time clustering is reported as being low. All these methods require space and time distance matrices for statistical calculations. In this study, centroids of district areas were used instead of the exact location of disease outbreaks.

Knox's test

The Knox method quantifies space-time interaction based on critical space and time distances. The test statistic, X, is a count of the number of pairs of cases that are

separated by less than the critical space and time distances. The criteria for defining closeness in space or time are judgmental and will depend upon the characteristics of the disease of interest, and the population at risk. Each of the $n\ (n\text{-}1)/2$ pairs can be classified as being close or far apart in time and close or far apart in space to form a 2 x 2 contingency table. Then the method tests the observed number of close pairs against an expected null distribution of Poisson deviates generated using permutation techniques. Pairs of cases will be near to one another when space-time interaction is present, and the test statistic will be large (Knox, 1964). The statistical significance is obtained through calculation of the proportion of the values in the upper right tail of the null distribution (i.e. probability of values equal or greater than the observed test statistic under the null hypothesis).

In the present study, 2 months were chosen as the critical temporal distance as an epidemic of FMD outbreaks would be expected to transmit from outbreak to outbreak at intervals of not more than two months. If the data would have been available at weekly intervals, smaller critical time distances more closely reflecting the biology of the infection process could have been used. The chosen critical distance in space was systematically varied to identify values maximising Knox's X. This allowed determining the spatial extent of disease clusters. Strictly, the critical distance should have been selected beforehand based on current understanding of FMD epidemiology, and not changed during the investigation.

Mantel method

The Mantel regression method uses the actual time and space distances between all possible pairs of outbreaks. It is not necessary to categorise the data as with Knox's test. For all case-case pairs, a graphical representation of the data can be produced by plotting the temporal distance against the spatial difference. The test statistic is calculated as the sum of the products of time and space distances for all case-case pairs. A null hypothesis distribution representing expected values for the test statistic assuming independence of space and time distances is calculated using a permutation approach. To generate the distribution, locations of cases are kept constant but the times are being randomly selected, and every time a value of the expected test statistic under space-time independence is recalculated. The statistical significance is estimated as was done with Knox's test through calculation of the proportion of the values in the

upper right tail of the null distribution (i.e. probability of values equal or greater than the observed test statistic under the null hypothesis).

Various transformations of the time/space distance values can be used to emphasise particular sections of the value distributions. For a contagious disease we expect short space and time distances to be correlated, but not the large distances. To emphasise the importance of the smaller distances Mantel recommends the use of the reciprocal transformation ($d'=1/(C + d)$) for the original values. To prevent an infinite reciprocal in cases of short distances in time or space it may be necessary to add a constant before taking a reciprocal. The reciprocal transformation will increase the small distances and reduce the large distances (Mantel, 1967).

One difficulty with Mantel's approach is the need to specify values for constants in the expressions for time and distance separation. Therefore, various constants for the space distance were used to identify the appropriate selection for this study but keeping the constant value for the time matrix at 1.

Space-time scan statistic

The space-time scan statistic is defined exactly the same as the purely spatial scan statistic. However, this method uses a cylindrical window with a circular geographic base and with height corresponding to time. The cylindrical window is then moved for each possible geographical location and size as well as any possible time period. The maximum temporal extent of potential clusters can be as large as the length of a time interval but should not be more than 90 percent of the study period (Kulldorff, 1997).

k-nearest neighbour method (k-NN)

The *k* nearest neighbour test for space-time interaction analyses the association between space and time distances of cases. Instead of using the actual values, the data is assessed according to whether individual cases are first, second and *k* nearest neighbours in both space and time. The test statistic, J_k , is the count of the number case pairs that are *k* nearest neighbours. The statistic ΔJ_k is the difference in nearest neighbours between the statistics J_k and J_{k-1}. Probability values are calculated by comparing the observed J_k and ΔJ_k to their reference distributions obtained from permuting the space-time nearest neighbour matrices. The P_k, describes how likely the observed ΔJ_k is under the null hypothesis of no association between the spatial and temporal nearest neighbour relationships.

The ΔJ_k and their associated P_k may be used to diagnose the space-time scale of the cluster process using an experiment-wise error. Given a desired type I error of α, the experiment-wise error rate, α', is $\alpha'=1-(1-\alpha)^{1/m}$, m is the number of k levels to be inspected. Significant space-time interaction occurs at those k where p_k less than α'. If the experiment-wise error approach identifies significant interaction at any k level, it suggests that the observed events tend to reoccur in the same area at that k time scale. The Stat! software allows 10 levels for k.

The statistical significance of space-time interaction across several levels of k can be addressed by combining probabilities using three methods: Bonferroni, Simes, and centroid distance. We report results of the centroid distance approach because it is considered to have greater statistical power than the Bonferroni or Simes techniques (Jacquez, 1996).

Results

Clustering in time

A window width of 2 month intervals was selected for the scan method to determine whether the sum of FMD cases in any 2 adjacent time intervals is excessively large in the study area. With the empty cell method, we wish to determine the probability of obtaining a number of cells without any cases of FMD greater than or equal to the exact number of empty cells (no FMD cases). Finally, we used Larsen's method to answer whether the time cells occupied with FMD cases tend to occur in a sequence or have an unusual pattern over time within a district or province area. The results for these three methods applied to the time series of provinces and districts with the infection of Foot and Mouth Disease type O, Asia1 and untyped are presented in Table 3-9.

We found evidence of temporal clustering for all types of virus infections in the whole country and for some provinces. There was unimodal clustering of FMD infection with virus type O in Nakhonsawan, Ubonbratchathani and Uttaradit province, with type Asia1 in Chiengrai, Sukothai and Uttaradit province, and in Roiet province with untyped virus. Unimodal patterns were seen with the outbreaks of FMD type Asia1 and untyped virus for the time series aggregated for the whole country. (Table 1 and Figure 1). Provinces that had too few cases and were therefore classified as out of range

were excluded from the whole country analysis. There was evidence of unimodal temporal clustering for all types of virus across provinces. (Tables 4 to 6 and Figures 2 to 3)

In the district time series, evidence of clustering in time was seen for most of the districts studied, but it was not possible to identify unimodal clustering in individual districts. When looking for temporal clustering across all districts simultaneously, we found that all FMD virus types demonstrated temporal clustering. Patterns of unimodal clustering were seen for the outbreaks of virus type O and type Asia1 in the time series of 3 and 4 districts, respectively. All time series of outbreaks with untyped virus were excluded from the analysis (Tables 7 to 9).

Table 3. Results from analysis of temporal clustering using a single time series representing FMD outbreak occurrence *Type O*, *Type Asia1*, and *untyped* virus aggregated for the whole country. The results from Larsen's test for multiple time series analysis is presented by overall p-value

Virus type	*p-values for different methods*		
	Scan method	*Larsen's method*	*Clustering*
Type O	<0.001	p= 0.69 (Z= 0.51)	✓
Type Asia1	<0.001	p= 0.01 (Z=-2.26)	Unimodal
Untyped virus	<0.001	p< 0.001 (Z=-3.71)	Unimodal
Overall	-	P< 0.01 (Z= -3.57)	Unimodal

Figure 1. Temporal pattern of FMD cases aggregated for the whole country by virus type between Jan 1995 and May 1997

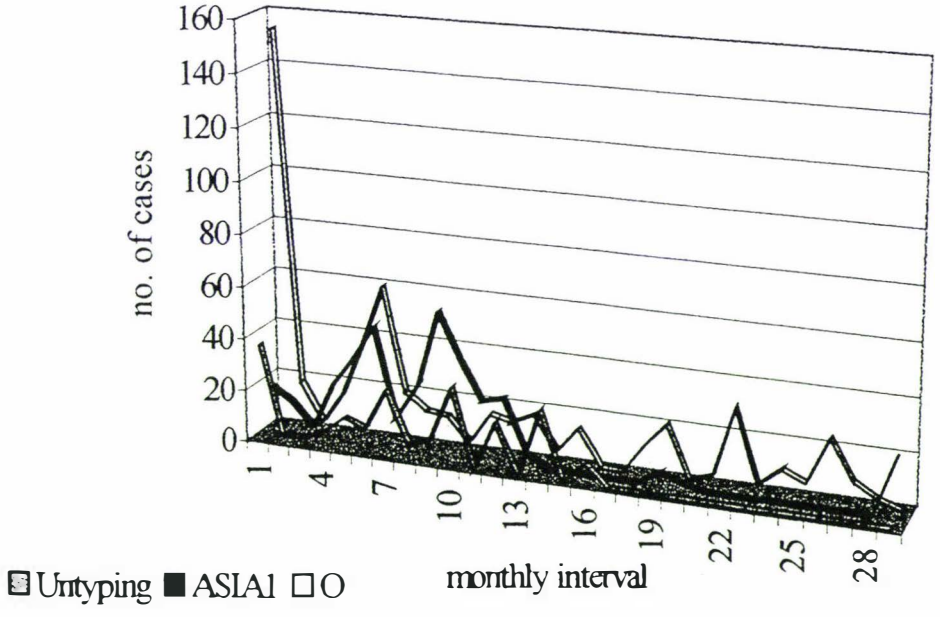


Table 4. Result of analysis of temporal clustering for the time series of individual provinces with Foot and Mouth Disease type O infection

Province	*p-values for different methods*			
	Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Chiengmai	< 0.001	Out of range	< 0.001	✓
Chiengrai	< 0.001	p = 0.55 (Z = 0.12)	< 0.001	✓
Kalasin	< 0.001	p = 0.44 (Z = -0.15)	< 0.001	✓
Kanchanaburi	< 0.001	Out of range	< 0.001	✓
Lampang	< 0.001	p = 0.06 (Z = -1.55)	< 0.001	✓
Lopburi	0.01	Out of range	< 0.01	✓
Maehongsorn	0.13	Out of range	Too few cases	✗
Mahasarakham	< 0.001	p = 0.43 (Z = -0.16)	< 0.001	✓
Nakhonratchasima	< 0.001	Out of range	< 0.001	✓
Nakhonsawan	< 0.001	p = 0.05 (Z = -1.60)	< 0.001	✓ unimodal
Nongkhai	< 0.001	p = 0.72 (Z = 0.60)	< 0.001	✓
Phitsanuloke	< 0.001	p = 0.18 (Z = -0.89)	< 0.001	✓
Phrae	< 0.001	Out of range	< 0.001	✓
Roiet	< 0.001	Out of range	< 0.001	✓
Sakonnakhon	0.01	Out of range	< 0.01	✓
Sisaket	0.13	Out of range	Too few cases	✗
Sukhothai	< 0.001	Out of range	< 0.001	✓
Tak	< 0.001	p = 0.18 (Z = -0.89)	< 0.001	✓
Ubonratchathani	< 0.001	p < 0.05 (Z = -2.78)	< 0.001	✓ unimodal
Udonthani	< 0.001	p = 0.13 (Z = -1.12)	< 0.001	✓
Uthaithani	0.64	p = 0.85 (Z = 1.04)	< 0.01	✓
Uttaradit	< 0.001	p = 0.02 (Z = -2.15)	< 0.001	✓ unimodal
Yasothon	< 0.001	Out of range	< 0.001	✓
Overall	-	P< 0.01 (Z = -3.07)	p < 0.001	✓ unimodal

Table 5. Results from analysis of temporal clustering of the time series of individual provinces with Foot and Mouth Disease type Asia1 infection.

Province	*p-values for different methods*			
	Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Burirum	< 0.001	Out of range	< 0.001	✓
Chiengmai	< 0.001	Out of range	< 0.001	✓
Chiengrai	< 0.001	p 0.03 (Z = -1.91)	< 0.001	✓ unimodal
Kampangphet	< 0.001	p 0.09 (Z = -1.34)	< 0.001	✓
Lamphun	< 0.01	Out of range	< 0.001	✓
Loei	< 0.001	Out of range	< 0.001	✓
Maehongsorn	< 0.001	Out of range	< 0.001	✓
Mukdahan	< 0.001	p 0.15 (Z = -1.04)	< 0.001	✓
Nakhonratchasima	Too few cases	Out of range	Too few cases	✗
Nakhonsawan	0.64	p 0.18 (Z = -0 89)	< 0.01	✓
Nan	0.05	p 0.15 (Z = -1.04)	< 0.01	✓
Nongkhai	< 0.001	p 0.17 (Z = -0.96)	< 0.001	✓
Phetchabun	< 0.001	Out of range	< 0.001	✓
Phrae	0.01	Out of range	< 0.01	✓
Sakonnakhon	< 0.001	Out of range	< 0.001	✓
Sukhothai	< 0.001	p < 0.01 (Z = -2.54)	< 0.001	✓ unimodal
Tak	< 0.001	Out of range	< 0.001	✓
Ubonratchathani	< 0.001	Out of range	< 0.001	✓
Udonthani	< 0.001	p 0.56 (Z = 0.16)	< 0.001	✓
Uttaradit	< 0.001	p 0.04 (Z = -1.76)	< 0.001	✓ unimodal
Overall	-	p < 0.001 (Z = -3.93)	p < 0.001	✓ unimodal

Table 6. Results from analysis of temporal clustering for the time series of individual provinces with Foot and Mouth Disease untyped virus infection

Province	*p-values for different methods*			
	Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Burirum	< 0.001	p 0.12 (Z = -1.91)	< 0.001	✓
Chiengmai	0.13	Out of range	Too few cases	✗
Kalasin	< 0.001	Out of range	< 0.001	✓
Lampang	< 0.001	p 0.18 (Z = -0.89)	< 0.001	✓
Lamphun	< 0.01	p 0.15 (Z = -1.04)	< 0.001	✓
Maehongsorn	< 0.001	p 0.12 (Z = -1.19)	< 0.001	✓
Mahasarakham	Too few cases	Out of range	Too few cases	✗
Nakhonratchasima	0.13	Out of range	Too few cases	✗
Phetchabun	< 0.001	Out of range	< 0.001	✓
Phitsanuloke	< 0.001	p 0.15 (Z = -1.04)	< 0.001	✓
Phrae	< 0.001	Out of range	< 0.001	✓
Roiet	< 0.001	p 0.04 (Z = -1.76)	< 0.001	✓ unimodal
Sakonnakhon	< 0.001	p 0.33 (Z = -0.45)	< 0.001	✓
Surin	< 0.001	p 0.50 (Z = 0.00)	< 0.001	✓
Tak	< 0.001	Out of range	< 0.001	✓
Udonthani	0.01	Out of range	< 0.01	✓
Overall	-	p < 0.01 (Z = -2.66)	p < 0.001	✓ unimodal

Figure 2. Temporal pattern of FMD cases with virus type O in three provinces with unimodal clustering

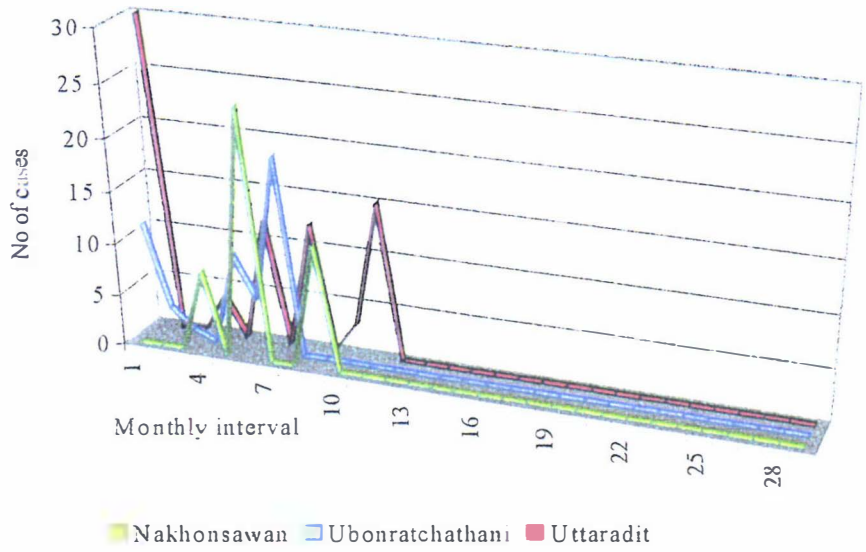


Figure 3. Temporal pattern of FMD cases with virus type Asia1 in three provinces with unimodal clustering

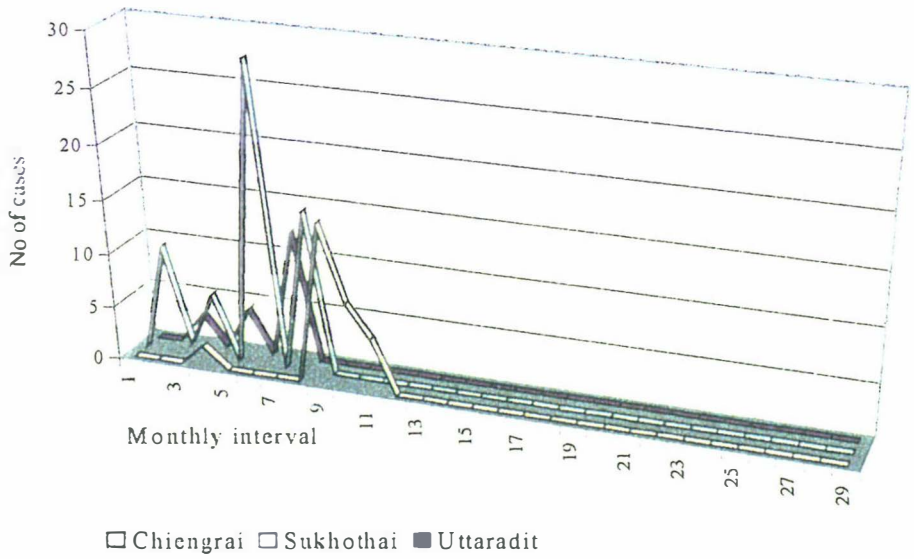


Table 7. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease type Asia1 infection

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Chiengmai	Maeaie	< 0.001	Out of range	< 0.001	✓
Chiengrai	Chiengkhong	Too few cases	Out of range	Too few cases	✗
	Maefaluang	< 0.001	Out of range	< 0.001	✓
	Maesaluea	< 0.001	Out of range	< 0.001	✓
	Padact	Too few cases	Out of range	Too few cases	✗
Kalasin	RongKham	< 0.001	Out of range	< 0.001	✓
	Yangtalat	< 0.001	Out of range	< 0.001	✓
Kanchanaburi	Nongprau	< 0.001	Out of range	< 0.001	✓
Lampang	Maetha	< 0.001	p 0.09 (Z=-1.34)	< 0.001	✓
	Sopprap	< 0.01	Out of range	< 0.001	✓
	Thoen	Too few cases	Out of range	Too few cases	✗
Lopburi	Srabot	0.01	Out of range	< 0.01	✓
Maehongsorn	Pai	0.13	Out of range	Too few cases	✗
Mahasarakham	Chiangyun	0.13	Out of range	Too few cases	✗
	Kosumpisai	< 0.001	Out of range	< 0.001	✓
	Nachuak	< 0.001	Out of range	< 0.001	✓
Nakhonratchasima	Nonthai	< 0.001	Out of range	< 0.001	✓
Nakhonsawan	Nongbua	< 0.001	Out of range	< 0.001	✓
	Phaisali	< 0.001	Out of range	< 0.001	✓
	Phayuhakhili	< 0.001	Out of range	< 0.001	✓

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Nongkhai	Phonphisai	0.13	Out of range	Too few cases	✗
	Sacka	< 0.001	Out of range	< 0.001	✓
Phitsanuloke	MuangPhitsanuloke	< 0.001	Out of range	< 0.001	✓
	Nakonthai	< 0.001	Out of range	< 0.001	✓
Phrae	Song	< 0.001	Out of range	< 0.001	✓
Roiet	Changhan	< 0.001	Out of range	< 0.001	✓
Sakonnakhon	Phangkon	0.01	Out of range	< 0.01	✓
Sisaket	Namkreng	0.13	Out of range	Too few cases	✗
Sukhothai	Bandanlanhoi	< 0.001	Out of range	< 0.001	✓
Tak	Popphra	< 0.001	Out of range	< 0.001	✓
	Thasongyang	< 0.001	Out of range	< 0.001	✓
Ubonratchathani	Detudom	< 0.001	Out of range	< 0.001	✓
	Kutkhaopun	< 0.001	Out of range	< 0.001	✓
	Laosuakok	< 0.001	p 0.09 (Z=-1.34)	< 0.001	✓
	Muangsamsib	Too few cases	Out of range	Too few cases	✗
	Tansum	< 0.001	Out of range	< 0.001	✓
Udonthani	Kudjub	< 0.001	Out of range	< 0.001	✓
	Kunpawapi	< 0.001	Out of range	< 0.001	✓
	Nongvuecosor	< 0.001	Out of range	< 0.001	✓
Uthaithani	Nongkhayang	0.13	Out of range	Too few cases	✗
	Thapthan	0.13	Out of range	Too few cases	✗
Uttaradit	Nampad	< 0.001	Out of range	< 0.001	✓

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
	Phichai	< 0.001	Out of range	< 0.001	✓
	Thongsankan	< 0.001	Out of range	< 0.001	✓
	Tron	< 0.001	p 0.07 (Z=-1.44)	< 0.001	✓
Yasothon	Patiu	Too few cases	Out of range	Too few cases	✗
	Overall	-	P < 0.01 (Z = -2.38,)	p < 0.001	✓ unimodal

Table 8. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease type Asia1 infection

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Burirum	Lamplaimat	< 0.001	Out of range	< 0.001	✓
Chiengmai	Fang	< 0.001	Out of range	< 0.001	✓
Chiengrai	Chiengkhong	0.13	Out of range	Too few cases	✗
	Thoeng	< 0.001	p 0.0899 (Z=-1.34)	< 0.001	✓
	Wiengpapao	< 0.001	Out of range	< 0.001	✓
Kampangphet	saingam	< 0.001	p 0.09 (Z=-1.34)	< 0.001	✓
Lamphun	Banhong	< 0.01	Out of range	< 0.001	✓
Loei	Nahaco	< 0.001	Out of range	< 0.001	✓
Maehongsorn	Macsariceng	< 0.001	Out of range	< 0.001	✓
Mukdahan	Dontan	Too few cases	Out of range	Too few cases	✗
	Khamcha-I	< 0.001	Out of range	< 0.001	✓
Nakhonratchasima	Chokchai	Too few cases	Out of range	Too few cases	✗
Nakhonsawan	Krokphra	0.13	Out of range	Too few cases	✗
	Phaisali	0.13	Out of range	Too few cases	✗
Nan	Namun	0.01	Out of range	< 0.01	✓
	Santisuk	Too few cases	Out of range	Too few cases	✗
Nongkhai	Phonphisai	0.13	Out of range	Too few cases	✗
	Sichiangmai	< 0.001	Out of range	< 0.001	✓
	Srakai	< 0.001	Out of range	< 0.001	✓

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Phetchabun	Nongphai	Too few cases	Out of range	Too few cases	✗
	Sithep	< 0.001	Out of range	< 0.001	✓
Phrae	Rongkwang	0.01	Out of range	< 0.01	✓
Sakonnakhon	Phangkon	< 0.001	Out of range	< 0.001	✓
Sukhothai	Khilimat	< 0.001	p 0.09 (Z=-1.34)	< 0.001	✓
	Kongkrailat	< 0.001	Out of range	< 0.001	✓
	Thungsaliam	< 0.001	p 0.12 (Z=-1.19)	< 0.001	✓
Tak	Macramad	< 0.001	Out of range	< 0.001	✓
Ubonratchathani	Sirinthon	< 0.001	Out of range	< 0.001	✓
Udonthani	Kudjub	< 0.001	Out of range	< 0.001	✓
	Phen	< 0.001	Out of range	< 0.001	✓
	Sangkom	0.13	Out of range	Too few cases	✗
Uttaradit	Phichai	< 0.01	Out of range	< 0.001	✓
	Thongsankan	< 0.001	Out of range	< 0.001	✓
	Tron	0.01	Out of range	0.0012	✓
Overall		-	p < 0.01 (Z = -2.61,)	p < 0.001	✓ unimodal

Table 9. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease untyped virus infection

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
Burirum	Nongki	Too few cases	Out of range	< 0.001	✓
	Prakhonchai	< 0.001	Out of range	< 0.001	✓
Chiengmai	Wienghang	0.13	Out of range	Too few cases	✗
Kalasin	RongKham	< 0.001	Out of range	< 0.001	✓
Lampang	Hangchat	< 0.001	Out of range	< 0.001	✓
	MuangLampang	0.13	Out of range	Too few cases	✓
Lamphun	Lac	< 0.001	Out of range	< 0.001	✓
	Macta	< 0.001	Out of range	< 0.001	✓
Machongsorn	Maclanoi	< 0.001	Out of range	< 0.001	✓
Mahasarakham	Phayakkhaphumoisai	Too few cases	Out of range	< 0.001	✓
Nakhonratchasima	Pratai	0.13	Out of range	Too few cases	✗
Phetchabun	Chondaen	< 0.001	Out of range	< 0.001	✓
Phitsanuloke	Phrompiram	Too few cases	Out of range	< 0.001	✓
	Wangthong	< 0.001	Out of range	< 0.001	✓
Phrae	Sungmen	< 0.001	Out of range	< 0.001	✓
Roiet	Chiengkwan	< 0.001	Out of range	< 0.001	✓
	Kasetwisai	< 0.001	Out of range	< 0.001	✓
Roiet	Ponsai	Too few cases	Out of range	< 0.001	✓
Sakonnakhon	Kutbak	< 0.001	Out of range	< 0.001	✓

Province	*District*	*p-values for different methods*			
		Scan method	*Larsen's method*	*Empty cell method*	*Clustering*
	Sawangdandin	< 0.001	Out of range	< 0.001	✓
Surin	Kabchocng	< 0.001	Out of range	< 0.001	✓
	Thatum	< 0.01	Out of range	< 0.001	✓
Tak	Popphra	< 0.001	Out of range	< 0.001	✓
Udonthani	Thungfon	< 0.001	Out of range	< 0.001	✓
Overall		-	-	p < 0.001	✓

Clustering in space

The Moran's *I* adjusted for population density (I_{pop}) had p-values of less than 0.001 for all virus types using data aggregated at the district as well as the province level (Tables 10 and 11). This indicates that foot-and-mouth-disease outbreaks in Thailand during the study period tended to occur clustered in space (Figures 4 and 5).

The results from analyses using the spatial scan statistic also showed evidence of spatial clustering of all FMD cases based on data aggregated at the district level (p=0.001). Table 12 shows the results of this analysis and the information for each detected cluster, including the radius of the estimated circle corresponding to the likely cluster location and the relative risk for districts within the cluster, compared with districts in the remainder of the study area.

Table 10. Results from analysis of spatial clustering using Moran's *I* adjusted for population density (I_{pop}) using province as the unit of aggregation

Virus type	I_{pop}'	*E[I]*	*% within*	*% among*	*Randomization assumption*	
					z-score	*p*
All cases	1.48	0.00	59.35	40.65	297.35	< 0.001
Types						
Type O	1.74	0.00	79.94	20.06	167.65	< 0.001
Type Asia1	2.55	0.00	81.16	18.84	169.05	< 0.001
Untyped	2.32	0.00	89.70	10.30	86.36	< 0.001

Table 11. Results from analysis of spatial clustering with Moran's *I* adjusted for population density (I_{pop}) using district as the unit of aggregation

Virus type	I_{pop}'	*E[I]*	*% within*	*% among*	*Randomization assumption*	
					z-score	*p*
All cases	5.74	0.00	84.63	15.37	417.97	< 0.001
Types						
Type O	11.44	0.00	88.20	11.80	392.54	< 0.001
Type Asia1	13.00	0.00	100.00	0.00	314.33	< 0.001
Untyped	28.15	0.00	92.13	7.87	381.63	< 0.001

Figure 4. Map of foot-and-mouth-disease cumulative incidence using province as unit of aggregation

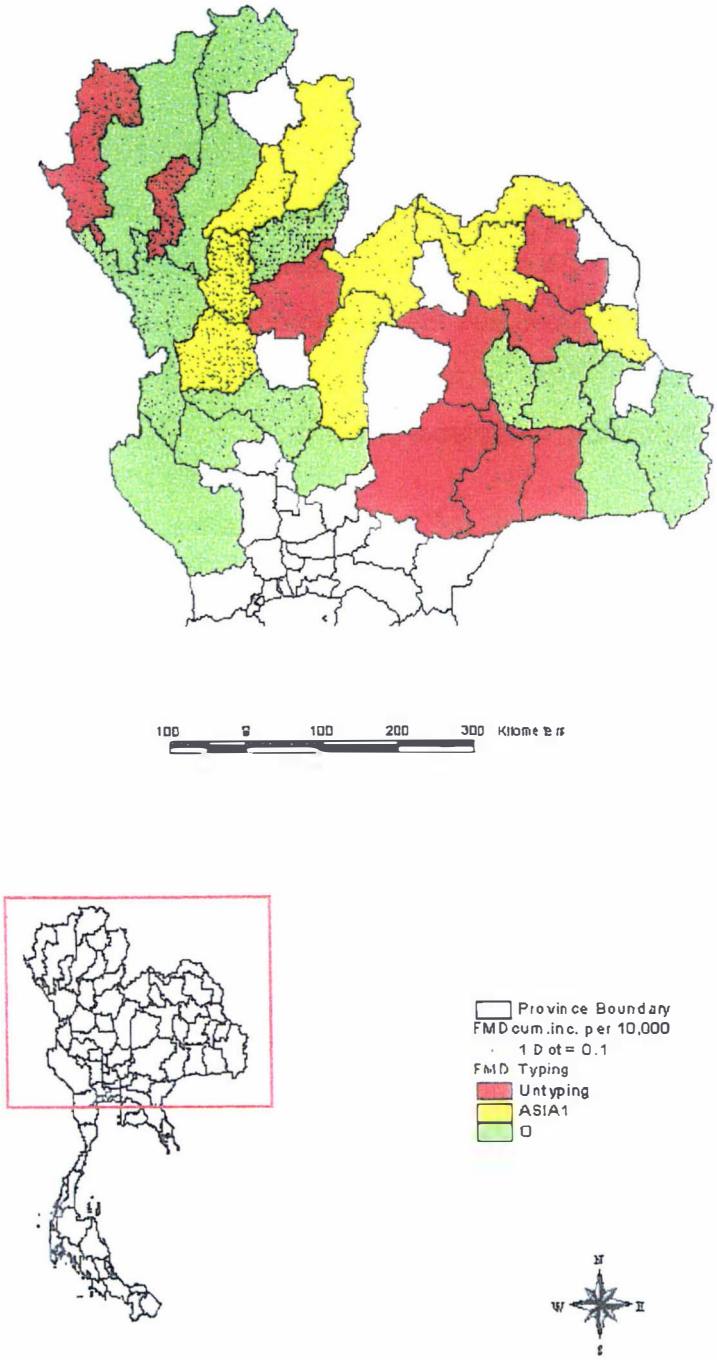


Figure 5. Map of foot-and-mouth-disease cumulative incidence using district as the unit of aggregation

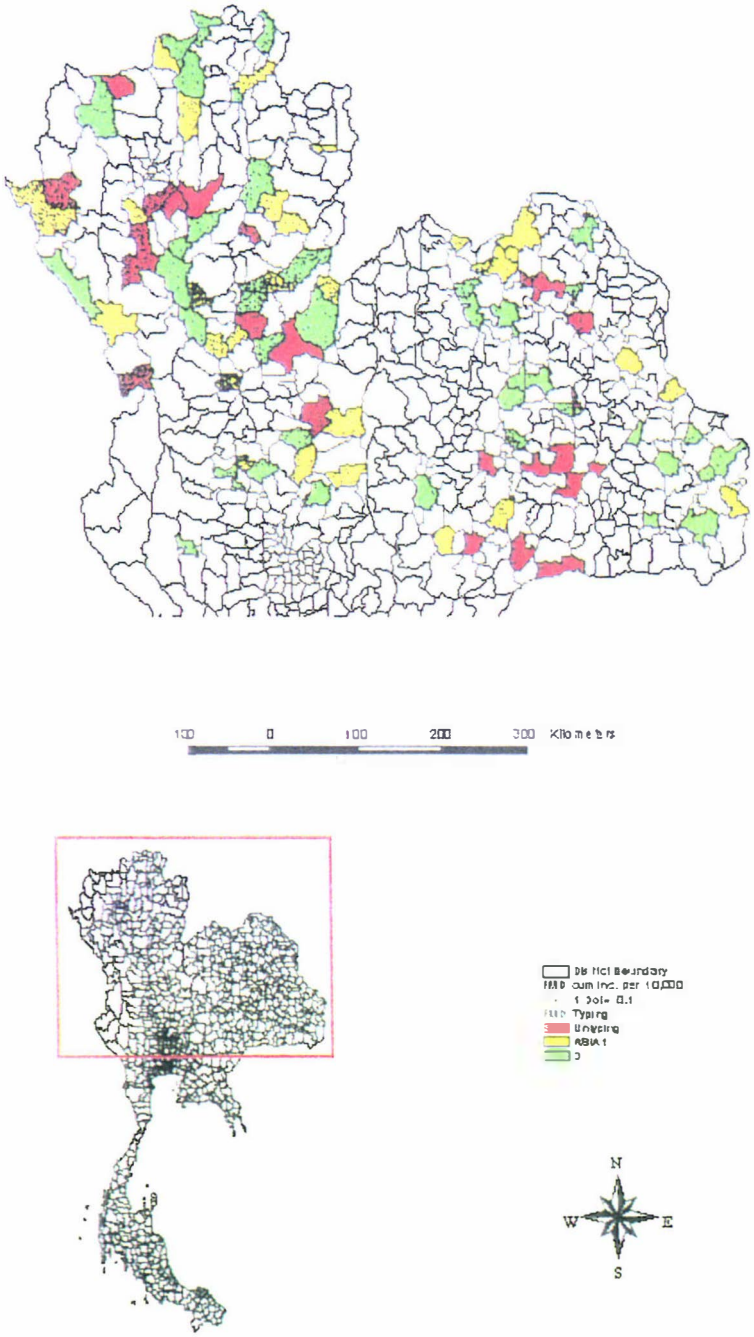


Table 12. Results from analysis of spatial clustering using the spatial scan statistic for all FMD cases using district level as the unit of aggregation and a maximum spatial cluster size of 50%

Cluster group	Districts included*	Radius* (km.)	Population *n*	Cases *n*	Expected *n*	RR*	p value
Most likely cluster							
1	2	17	26,105	73	2.31	31.62	0.001
Secondary clusters							
2	3	27.1	41,712	64	4.1	15.61	0.001
3	1	0	8,690	35	0.71	49.07	0.001
4	1	0	8,273	36	0.99	36.38	0.001
5	2	28.25	13,205	35	1.15	30.34	0.001
6	1	0	10,921	30	1.10	27.17	0.001
7	42	251.8	541,703	147	49.09	2.99	0.001
8	3	19.6	64,255	60	8.32	7.21	0.001
9	1	0	12,204	26	1.01	25.72	0.001
10	7	18.08	82,867	55	8.77	6.27	0.001
11	4	19.26	60,011	50	7.82	6.40	0.001
12	4	32.83	42,049	41	5.03	8.16	0.001
13	1	0	14,520	19	0.81	23.36	0.001
14	1	0	16,445	24	2.00	12.02	0.001
15	1	0	5,594	15	0.55	27.32	.0001
16	1	0	19,354	24	2.49	9.65	0.001
17	1	0	38,632	25	2.87	8.70	0.001
18	5	38.63	95,615	39	11.48	3.40	0.001
19	1	0	21,080	17	2.80	6.08	0.001
20	3	17.51	37,040	12	2.60	4.61	0.045

* District included = number of districts included in that potential cluster;

* Radius size = radius of the circle corresponding to that potential cluster;

* RR = Relative risk for district(s) within the cluster compared with districts in the remainder of that pre-defined area.

Clustering in space-time

We found no indication of space-time clustering when using Knox's method for any of the FMD virus types using the critical distances of 20, 30 and 50 km in space and 2 months in time. The results of Knox's test including observed and expected frequencies are shown in Table 13.

The Mantel method also did not show clustering in space - time for any of the virus types (see Table 14).

Evidence of significant space-time clustering for all FMD virus types was found using the space-time scan statistic (see Table 15). The units of spatial aggregation used in this analysis were district levels.

The results of the *k*-nearest neighbour method shown in Table 16 indicate significant space-time interaction for all FMD virus types with a combined p-value of the centroid distance of less than 0.01. On examination of the ΔJ_k and the P_k using an experiment-wise error of 0.05 to determine the space-time scale of the FMD cluster process, no significant space-time interaction was indicated at any of the *k* levels (p_k not less than α'). The result agrees with the Knox and Mantel tests suggesting that none of the FMD virus types was clustered in space-time.

Table 13. Results from analysis of time-space clustering of FMD cases using the Knox method with critical time of 2 months and spatial aggregation at the district level

FMD cases	*Critical space distance (Km.)*	*Observed pairs*	*Expected pairs*	*p value*
All cases	20	6	4.45	0.28
	30	14	9.22	0.10
	50	31	26.71	0.22
Type O	20	1	0.72	0.54
	30	2	2.02	0.63
	50	4	5.92	0.86
Type Asia1	20	0	0.78	1.00
	30	2	1.55	0.48
	50	3	3.57	0.72
Untyped	20	0	0.54	1.00
	30	0	0.72	1.00
	50	1	2.16	0.93

Table 14. Result from analysis of time-space clustering of FMD cases using the Mantel method after reciprocal transformation of the distances and addition of various constants to the space distance with spatial aggregation at the district level.

FMD cases	*Constant*	*Mantel R*	*p value*
All cases	1	0.01	0.14
	1,000	0.01	0.14
	10,000	0.01	0.28
Type O	1	-0.01	0.53
	1,000	-0.01	0.57
	10,000	-0.03	0.88
Type Asia1	1	-0.03	0.78
	1,000	-0.03	0.79
	10,000	-0.02	0.68
Untyped	1	-0.04	0.83
	1,000	-0.04	0.86
	10,000	-0.03	0.66

Table 15. Results from analysis of time-space clustering of all FMD cases with the space-time scan statistic based on a maximum spatial and temporal cluster size of 50% of the total population using data aggregated at the district level

Cluster group	*Districts included**	*Radius** *(km.)*	*Time frame** *(month)*	*Population n*	*Cases n*	*Expected n*	*RR**	*p value*
Most likely cluster	7	38 80	01/95 - 09/95	127,996	111	3.30	33.64	0.001
Secondary clusters	29	84.10	01/95 - 11/95	391,656	150	13.45	11.16	0.001
2								
3	1	0	10/96 - 10/96	8,190	30	0.03	1139.81	0.001
4	1	0	01/95 - 01/95	8,273	30	0.03	862.55	0.001
5	1	0.00	10/95 - 10/95	10,921	30	0 04	773.103	0 001
6	3	21.80	02/95 - 08/95	64,255	60	2.00	30.01	0.001
7	42	240.2	09/95 - 12/95	521,982	76	6.50	11.69	0.001
8	1	0.00	04/95 - 05/95	7,001	24	0 06	37.040	0.001
9	1	0.00	06/95 - 06/95	19,3544	24	0.08	283 82	0.001
10	1	0.00	05/95 - 05/95	5,594	15	0.02	777 19	0.001

* District included = number of districts included in that potential cluster;

* Radius size = radius of the circle corresponding to that potential cluster;

* Time frame = time interval corresponding to that potential cluster;

* RR = Relative Risk for district(s) within the cluster compared with districts in the remainder of that pre-defined area.

Table 16. Results from analysis of time-space clustering of FMD cases in individual districts using K nearest neighbour analysis spatial aggregation at the district level

FMD cases		*k*										*Overall*
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*p*
All cases	J_k	11	49	90	169	237	318	434	554	707	863	< 0.01
	PJ_k	0.07	0.14	0.58	0.48	0.98	1.00	1.00	1.00	0.99	0.89	
	ΔJ_k	11	38	41	79	68	81	116	120	153	156	< 0.01
	$P(\Delta J_k)$	0.07	0.33	0.92	0.41	1.00	0.98	0.56	0.96	0.02	0.06	
Type O	J_k	3	9	30	54	89	129	178	219	274	330	0.08
	PJ_k	0.41	0.98	0.89	0.93	0.93	0.81	0.82	0.98	0.96	0.99	
	ΔJ_k	3	6	21	24	35	40	49	41	55	56	0.52
	$P(\Delta J_k)$	0.41	0.99	0.55	0.76	0.69	0.22	0.60	0.97	0.43	0.84	
Type Asia1	J_k	4	13	29	53	81	108	151	191	246	293	0.32
	PJ_k	0.10	0.53	0.68	0.72	0.89	0.98	0.94	0.95	0.68	0.59	
	ΔJ_k	4	9	16	24	28	27	43	40	55	47	0.54
	$P(\Delta J_k)$	0.10	0.79	0.73	0.66	0.85	0.92	0.34	0.68	0.06	0.36	
Untyped virus	J_k	0	9	26	47	74	105	144	175	206	244	0.38
	PJ_k	1.00	0.60	0.27	0.28	0.17	0.18	0.10	0.17	0.28	0.13	
	ΔJ_k	0	9	17	21	27	31	39	31	31	38	0.95
	$P(\Delta J_k)$	1.00	0.39	0.19	0.51	0.31	0.56	0.34	0.84	0.84	0.48	

Discussion

Disease clustering can be broadly defined as any excess of disease occurrence. Research into disease clustering, especially with regard to space and time, has often been accompanied by specialised methods. However, the main methodologic tools are based on the comparison of disease rates, and the utility of a particular comparison of disease rates depends on the range of the possible explanations for each investigation (Rothman, 1987).

In this study, foot-and-mouth-disease reporting data for Thailand from January 1995 to May 1997 was investigated for detection of temporal or spatial or spatio-temporal clustering.

Temporal clustering

The analysis of the temporal pattern using the whole country, as well as individual provinces and districts as the level of aggregation, based on the scan method and the empty cells test, indicates that foot and mouth disease in Thailand did not occur randomly during the 29 months under consideration. Statistical significance based on the scan method supports clustering of FMD within the 2-month scanning window fitted for the studied distance unit. The p-value of the scan method is a measure of the probability of observing equal or more than the maximum number of FMD cases in the specified time window. However, p-values for individual time series cannot be combined to yield an overall p-value since the method only allows testing single time series. Changes in population size during the time period included in the analysis can result in bias and may produce type I errors. If the total number of cases is small, as would be the case if fewer than 10 outbreaks were observed in one province or district, the scan statistic may have insufficient statistical power to detect temporal clustering of disease in a population (Wallenstein, 1980).

These problems can be avoided by using the empty cells test. The empty cells method is sensitive to clustering in the presence of time intervals containing sequences of zeros. Under the null hypothesis of randomness of cases among the time cells, the significance of the test is determined by the probability of obtaining a number of empty cells (in this case corresponding to no reports of FMD) greater than or equal to the expected value for the number of empty cells. The overall p value is combined as a continuity-corrected chi-square with one degree of freedom. This analysis provided evidence for temporal clustering of FMD cases (based on $p < 0.05$) at district, province and country levels of aggregation, with

the exception of provinces and districts where there were too few cases to allow any inference to be made (Tables 3 to 8).

Larsen's method reported 'out of range' analysis results for all but 3 districts for virus type O and 4 districts for virus type Asia1 (Tables 6 to 8) where there was evidence of unimodal clustering (based on significant z scores). 'Out of range' was due to having cases fewer than 2 occupying intervals in that district. Although the test had limited value for detecting temporal clustering at the district level, it had worthwhile application at the province level and for the whole country (Table 2-5 and Fig. 1-3). The total z score is not biased by differences in population size across time series, but it can be influenced by population shifts over the time period as in the case of highly transient animal populations. Since the *K* (Larsen's test) statistic specifically addresses unimodal temporal clustering only, it may not be used to detect multiple clusters or to distinguish true clustering from population shift effects (Larsen, 1973).

Spatial clustering

The value of the test statistic (I^*_{pop}) of the adjusted Moran's I test was used as a measure of the significance of the spatial clustering of outbreaks, both within and between provinces and districts, respectively. In general, the test has good power although geographic aggregation causes loss of resolution when applied to area-based data. The test examines spatial patterns of regional outbreaks (not exact locations) and also takes high variance of those disease occurrences into account. The adjusted Moran's *I* is powerful and robust and can be relied on to detect any pattern of disease clustering (Oden *et al*, 1996).

The space scan statistic produced the same result with evidence of statistically significant spatial clustering of all FMD cases using districts as the aggregation unit. This method also detected locations of 'most likely' and 'secondary' cluster groups. The spatial scan statistic is based on a likelihood ratio. In our study, the objective was the detection of clusters of any size and anywhere in the study area. Therefore, the maximum spatial size was set to 50 percent of the total population at risk, which allowed the system to define clusters of both small and large sizes without any pre-selection bias in terms of cluster size (Kulldorff and Nagarwalla, 1995)

Space-time clustering

We were unable to detect any space-time interaction in the FMD outbreaks in Thailand between January 1995 and July 1997 with the Knox and Mantel methods. However, the space-time scan statistic and the *k* nearest neighbour methods did detect space-time clustering when all virus types were considered together. The implications of these various test results are considered later in this discussion.

The level of aggregation represented by the units of time and space are an important consideration when investigating space-time clustering. Clusters may go undetected if the unit is too large or too small. For this study, the data limited us to one-month time units, which is generally unsuitable for most infectious diseases. The most appropriate time unit for FMD is daily intervals since the incubation period of Foot and Mouth disease is generally within a range of 2-14 days (Donaldson, 1990). Furthermore, the location of outbreaks was not precisely defined. This forces us to use central points of districts as locations of outbreaks. These data limitations are likely to have adversely affected our ability to detect space-time interaction in this study.

Inherent limitations of statistical procedures and statistical power are acknowledged as the most important issues affecting the outcome of space-time cluster studies. The Knox test needs a critical space and time distance to define cluster units, and for infectious disease, the geographic critical distance should reflect the average distance between 2 outbreaks where there are common factors involved in initiating outbreaks. In general, critical distances are selected with due regard to the particular disease being investigated.

The Mantel test uses transformed data in the space and time distance matrices but there are some inherent difficulties in that process that need to be taken into account. Although addition of a constant to the distance prior to transformation helps to minimise the effects of zeros or extremely short distances, the results of any analysis are unavoidably affected to some extent by the particular transformations and constants chosen for addition to the time and space distances. A particular problem for selecting a constant for addition arises when the logarithmic transformation is applied to data in which zero values occur. If the constants chosen are too small, the region near zero will be unduly expanded with resulting loss in power for detecting clustering. The constant should, in some way, be commensurate with the anticipated possible or probable distance in time or space between related cases (Mantel, 1967)

Statistical power in the context of this discussion is essentially the ability of a method to detect or identify true clusters (true positives). The paired distance approach (Knox and Mantel methods) is sensitive primarily to time-space clustering and insensitive to clustering which is purely spatial or purely temporal (Mantel, 1967). If we consider a district where a high frequency of outbreaks (within spatial clustering) regularly occurs either due to high disease incidence rates or a high density of population, then outbreaks within that district will tend to be close together in space but no closer to one another in time than they are to outbreaks outside the district.

Alternatively, let us consider a sub-interval in time in which disease incidence rates rise everywhere throughout the area of interest (temporal clustering). Outbreaks in the district will be close together in time, but as their spatial distribution is unaffected, they will be no closer to each other in space than they are to cases arising outside the time sub-interval. In such circumstances the *k*-nearest neighbour method which uses the interaction of space and time nearest neighbour relationship is more appropriate for detecting that type of chain of infection. Chen *et al.* (1984) compared the Mantel method and the Knox method for detecting simultaneous clustering in space and time of Hodgkin's disease and found similarity in their insensitivity to simulated patterns. Wartenberg and Greenberg (1990) also reported low power with the Mantel method for small numbers of cases increasing with greater number of cases for his hot-spot model. McAuliffe and Afifi (1984) compared an approach using nearest neighbour distance with the Knox and Mantel methods in their investigation of space-time cluster detection procedures. They argued that the nearest neighbour approach is superior because the user is not required to specify either critical distance or a constant for the inverse distance transformation. Later, Jacquez (1993) described a *k* nearest neighbour statistic that was sensitive to the pattern of cases expected in the presence of space-time clustering and compared it's power with the Knox and Mantel tests. He suggested the *k*-NN has superior statistical power relative to the Knox and Mantel tests for the following reasons.

1. The Mantel and Knox tests are subjective because they require the selection of data transformations, constants, and critical distances prior to the analysis, whereas the *k*-NN test does not require prior parameter specification and the levels of *k* inspected do not alter the value of the test statistic.

2. Unlike the Knox test, the *k*-NN test does not use critical distance or other parameters for assessing geographic proximity.

3. Unlike Mantel's test, the *k*-NN test does not rely on a linear model. It uses instead the intersection of a space and time nearest neighbour relationship, which is intuitively appropriate for detecting a contagious process.

4. Because it is the sum of the products of the space and time distances, large space and time distances unduly influence Mantel's statistic. The k-NN test is sensitive only to an excess of cases that are near in both space and time as it uses a nearest neighbour relationship. Paired cases separated by large space and time distance do not enter into the calculation of the test statistic.

5. The *k*-NN test has greater statistical power under realistic and simulated heterogeneous population density distributions while the Knox and Mantel tests have limited statistical power when population density varies.

The above reasons provide a number of possible explanations why space-time clustering was not detected the Mantel and Knox tests, but with the k nearest neighbour test in the current investigation.

The space-time scan statistic uses the same method of detection of cluster in the study area as does the spatial scan statistic. It is an extension of the latter, with time incorporated in the scanning process as another dimension. Without any pre-selection of size of spatial and temporal clusters the system was allowed to detect any space-time clusters of any size. The locations of likely space-time clusters were identified. However, the results for these detected clusters should be interpreted taking into account epidemiological knowledge about FMD as well as the quality of the underlying data.

In comparison, both the *k*-NN method and the space-time scan statistic were the most powerful methods for detecting time-space clustering in this dataset. The space-time scan statistic has the additional advantage over the *k*-NN method that it also identifies the actual locations of the clusters, and generates relative risks as parameters which can be easily interpreted by epidemiologists. The concepts behind the space-time scan statistic are also much more easily to understand than they are in the case of the *k*-NN statistic.

Miscellaneous issues

With any of these statistics, confounding should be considered when analysing and interpreting the results. Confounding can arise from influences as simple as a change in population density, or it can involve patterns of other factors such as age and breed. Since

risk may vary dramatically at different levels of these variables, corresponding changes in true incidence rates may be expected that are not always attributable to external factors. Amongst the spatial clustering methods, clearly Moran's *I* adjusted for population density and the spatial scan statistic are most appropriate if dealing with area type data as in the current case, because both can take into account any spatial variation in the underlying populations at risk. The spatial and the space-time scan statistic have the additional advantage over the other methods that they allow incorporating further confounding factors through stratified analyses.

The level of aggregation of locational data can bring biases to the analysis associated with the modifiable areal unit problem (MAUP). The MAUP is in reality composed of two closely related problems: the scale and the aggregation problem. The scale problem is the variation in the results that can be obtained when data from one set of areal units are aggregated into fewer or larger units for analysis purposes. The scale problem arises from the uncertainty about the number of areas needed for a particular study. The latter problem results from the use of alternative unit analyses when the number of units is held constant. The aggregation problem arises because of uncertainty about how the data are to be aggregated to form a given number of zones or clusters in a study. The problems relate to the difficult statistical problems of identifying the nature of the underlying relationships implicit in an aggregate level study (Openshaw, 1984). As in our study, the nature of the administrative areas can also create difficulties. Their boundaries may be drawn according to demographic features that are directly or indirectly related to disease, thereby introducing selection effect confounding. Further difficulties arise if variables in the analysis do not share the same administrative zones. Boundaries may cut through genuine clusters of disease and thereby mask them. It is therefore preferable if the data is always collected at the highest spatial and temporal resolution possible. This means cases of disease should be recorded using farm or village locations and the date of occurrence. The investigator then has more control over the selection of appropriate levels of aggregation.

Conclusion

It is important to choose the most appropriate method of analysis for any disease cluster investigation, and that usually involves consideration of a range of cluster scenarios and an evaluation of various methods. This places considerable responsibility on the investigator who needs to have a thorough understanding of the data and the epidemiology of the disease

of interest. A range of analytical methods should be applied with due consideration of their strengths and weakness in relation to the disease process before deciding on the final method of choice.

Considerable attention must be paid to confounding. Since its presence may lead to false positive and false negative results. Although it is a complex and difficult issue to address, it is an important component of the interpretation of cluster studies.

There are some issues for the study of foot-and-mouth-disease clustering in Thailand which we were unable to take into account in our studies such as animal movement across the borders between Thailand and neighbouring countries, as well as movement within Thailand, or the effect that livestock markets have on the epidemiology of this disease. Despite these difficulties we are confident in our conclusions that FMD outbreaks in Thailand during January 1995 to May 1997 demonstrated a unimodal pattern of temporal and spatial clustering.

Reference List

Cave, D.R., Freedman, L.S., 1975. Seasonal variation in the clinical presentation of Crohn's disease and Ulcerative Colitis. International Journal of Epidemiology, 4: 317-320.

Chen, R., 1979. Statistical techniques in birth defects surveillance systems. Control Epidemiology Biostatics 1: 184-89.

Chen, R., 1986. Revised values for the parameters of the sets technique for monitoring the incidence rate of a rare disease. Methods of Information in Medicine, 25: 47-9.

Chen, R., Mantel, N., Klingberg, M.A., 1984. A study of three techniques for time-space clustering in Hodgkin's disease. Statistics in Medicine, 3: 173-84.

Cliff, A.D., Ord, J.K., 1981. Spatial processes: models & applications. Pion Limited, London.

Cuzick, J., Edwards, R., 1990. Spatial clustering for inhomogenous populations. Journal of the Royal Statistical Society Series B, 52: 73-104.

Dat, N.V., 1982. Tests for space-time clustering of diseases. [Ph. D. dissertation]. Department of Biostatistics, University of North Carolina, Chapel Hill.

David, F.N., Barton, D.E., 1966. Two space-time interaction tests for epidemicity. British Journal of Preventive & Social Medicine, 20: 44-8.

Donaldson, A.I., 1990. Foot-and-mouth diseaes. Surveillance, 17: 6-8.

Ederer, F., Myers, M.H., Mantel, N., 1964. A statistical problem in space and time: Do leukemia cases come in clusters? Biometrics, 20:

Edmonds, L.D., Layde, P.M., James, L.M., Flynt, J.W., Erickson, J.D., Oakley, G.P.Jr, 1981. Congenital malformations surveillance: two American systems. International Journal of Epidemiology, 10: 247-52.

Edwards, J.H., 1961. The recognition and estimation of cyclic trends. Annals of Human Genetics, 25: 83-86.

Grimson, R.C., 1979. The clustering of disease. Mathematical Biosciences, 46: 257-278.

Grimson, R.C., 1989. Assessing patterns of epidemiologic events in space-time. In: Proceeding of the 1989 Public Health Conference on Records and Statistic.

Grimson, R.C., 1993. Disease clusters, exact distributions of maxima, and P-values. Statistics in Medicine, 12: 1773-94.

Grimson, R.C., Rose, R.D., 1991. A versatile test for clustering and a proximity analysis of neurons. Methods of Information in Medicine, 30: 299-303.

Grimson, R.C., Wang, K.C., Johnson, P.W.C., 1981. Searching for hierarchical clusters of disease: Spatial patterns of sudden infant death syndrome. Social Science & Medicine, 287-93.

Hardy, R.J., Schroder, G.D., Cooper, S.P., Buffler, P.A., Prichard, H.M., Crane, M., 1990. A surveillance system for assessing health effects from hazardous exposures. American Journal of Epidemiology, 132: S32-42.

Hewitt, D., Milner, J., Csima, A., Pakula, A., 1971. On Edwards' criterion of seasonality and a non-parametric alternative. British Journal of Preventive & Social Medicine, 25: 174-6.

Hill, G.B., Spicer, C.C., Weatherall, J. A., 1968. The computer surveillance of congenital malformations. British Medical Bulletin, 24: 215-8.

Jacquez, G.M., 1993. The statistical description of disease clusters. Statistics in Medicine, 12: 1967-8.

Jacquez, G.M., 1994. Cuzick and Edwards' test when exact locations are unknown. American Journal of Epidemiology, 140: 58-64.

Jacquez, G.M., 1996. A *k* nearest neighbour test for space-time interaction. Statistics in Medicine, 15: 1935-49.

Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. Statistics in Medicine, 14. 799-810.

Kulldorff, M., 1997. A spatial scan statistic. Communications in Statistical Theory and Methods, 26(6). 1481-1496.

Knox, E.G., 1964. The detection of space-time interactions. Applied Statistics, 18: 25-29.

Knox, E.G., 1988. Detection of disease clusters. In: P. Elliott (Methodology of Enquiries into Disease Clustering. London School of Hygiene and Tropical Medicine, London, pp. 17-22.

Larsen, R.J., Holmes, C.L., Heath, C.W. Jr, 1973. A statistical test for measuring unimodal clustering: a description of the test and of its application to cases of acute leukemia in metropolitan Atlanta, Georgia. Biometrics, 29: 301-9.

Lawson, A.B., 1988. On the tests for spatial trend in a non-homogenous Poisson process. Journal of Applied Statistics, 15: 225-235.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research, 27: 209-20.

Marshall, R.J., 1991. A review of methods for statistical analysis of spatial patterns of disease. Journal of the Royal Statistical Society Series A, 154 : 421-441.

McAuliffe, T.L., Afifi, A.A., 1984. Comparison of a nearest neighbor and other approaches to the detection of space-time clustering. Computational Statistics & Data Analysis, vol.2: 125-42.

Oden, N., 1995. Adjusting Moran's I for population density [see comments]. Statistics in Medicine, 14: 17-26.

Oden, N., Jacquez, G., Grimson, R., 1996. Realistic power simulations compare point- and area-based disease cluster tests. Statistics in Medicine, 15: 783-806.

Ohno, Y., Aoki, K., Aoki, N., 1979. A test of significance for geographic clusters of disease. International Journal of Epidemiology, 8: 273-80.

Openshaw, S.,1984, The modifiable areal unit problem, Concepts and Techniques in Modern Geography no. 38, Geo Books, Norwich.

Roger, J.H., 1977. A significance test for cyclic trends in incidence data. Biometrika, 64: 152-155.

Ross, A., Davis, S., 1990. Point pattern analysis of the spatial proximity of residences prior to diagnosis of persons with Hodgkin's disease. American Journal of Epidemiology, 132: S53-62.

Rothman, K.J., 1987. Clustering of disease [editorial]. American Journal of Public Health, 77: 13-5.

Symons, M.J., Grimson, R.C., Yuan, Y.C., 1983. Clustering of rare events. Biometrics, 39: 193-205.

Tango, T., 1984. The detection of disease clustering in time. Biometrics, 40: 15-26.

Wallenstein, S., 1980. A test for detection of clustering over time. American Journal of Epidemiology, 111: 367-72.

Wallenstein, S., Neff, N., 1987. An approximation for the distribution of the scan statistic. Statistics in Medicine, 6: 197-207.

Wartenberg, D., Greenberg, M., 1990. Detecting disease clusters: the importance of statistical power. American Journal of Epidemiology, 132: S156-66.

Wartenberg, D., Greenberg, M., 1993. Solving the cluster puzzle: Clues to follow and pitfalls to avoid. Statistics in Medicine, 12: 1763-1770.

Weatherall, J.A., Haskey, J. C., 1976. Surveillance of malformations. British Medical Bulletin, 32: 39-44.

CHAPTER 3.

Detection of Spatial Clustering of Enzootic Bovine Leucosis in New Zealand

Abstract

In 1996, the New Zealand dairy industry commenced a program aimed at the eradication of Enzootic Bovine Leucosis (EBL) from the national herd. A series of studies has been conducted in order to provide important information on the epidemiology of this disease. This information has been included in the design and the continued refinement of the eradication program.

In this paper, the occurrence of EBL in New Zealand dairy cattle herds was analysed for the presence of spatial clustering. Kernel smoothing was used to generate a surface expressing prevalence of positive herds in the country from point locations of cattle herds. The spatial scan statistic was applied for statistical hypothesis testing of the presence of clustering as well as for indicating the locations of likely clusters.

Two spatial clusters of positive herds were detected: one consisting of 498 herds in the Bay of Plenty area (p =0.001) and one of 83 herds in the northern part of the South Island (p = 0.082).

Further investigations are required to determine if there are identifiable factors associated with the observed clustering of EBL positive herds in New Zealand.

Introduction

EBL is the most important of the bovine neoplastic diseases caused by Bovine Leukaemia Virus (BLV) (Ferrer, 1980). Although EBL has not caused significant economic losses to the New Zealand dairy industry, it is possible that the disease may restrict the access of New Zealand's dairy products to important markets in the future. In order to prevent the economic impact of such restrictions in the future, the need to

eradicate this disease from the national herd dairy industry has become more and more apparent. The presence of EBL in New Zealand has been recognised for some time and the disease was surveyed in 1979 (Hilbink and Penrose, 1993). A recent study based on the serological examination of random milk samples from individual cows from 1700 herds (11.3% of all dairy herds) revealed that approximately 6.5% of herds were infected with BLV, while regional herd prevalence varied significantly from 2.5 to 10% (Burton *et al.*, 1997).

In 1996, the New Zealand Dairy Board initiated an industry-funded programme to eradicate BLV from the country. As part of the programme, a series of epidemiological studies were undertaken to determine the incidence and prevalence of the disease and to identify risk factors associated with the occurrence and spread of EBL. In the present study, techniques of spatial analysis have been used to determine if the distribution of EBL-positive dairy herds in New Zealand is random or clustered. If clusters can be identified, it may be possible to refine the control programme by targeting these areas. In addition, identification of these areas may provide the opportunity to initiate focused studies on the risk factors that may influence the distribution of the disease. As much of the aetiology of EBL in New Zealand is as yet unclear, an exploratory analytical approach had to be adopted for this investigation.

Material and Methods

Classification of EBL status

As part of the EBL control scheme, all dairy herds supplying milk to a factory are screened annually using an ELISA testing protocol. Cattle in herds that return positive bulk milk tests and that participate in the Livestock Improvement Corporation's (LIC) herd testing programme are subjected to individual milk tests. Where individual milk samples indicate infection, the animals that returned a positive test are blood-sampled and serum is collected for serology to confirm their infection status. If the bulk milk test is negative, milk samples are pooled in groups of about 20 for further testing. If a pooled sample tests positive, the individuals contributing to the pool are tested individually. In herds that do not herd test, custom milk samples or blood samples are collected from each animal. Herds are declared free of BLV after the whole herd tests

negative for 3 consecutive years on bulk milk and aggregated production test samples (Hayes, 1998).

Data

The data used in the analyses was supplied by the New Zealand Dairy Industry EBL Control Scheme. For each farm involved in the scheme, the EBL status and the X-Y co-ordinates based on the New Zealand metric map grid coordinate system were provided farm's centroid. Table 17 presents the number of herds in each of the different EBL status categories in the testing period 1997-1998 from the total of 14,301 herds with a co-ordinate reference included in the data set.

Following the definition of the EBL control scheme, only *blood test positive* herds (866/14,301) were considered as EBL-positive herds in this analysis. Subsequent use of the term "positive herd" refers to these 866 herds specifically. The point locations of all dairy herds and all positive herds are presented in Figure 6.

Figure 6. Map of herd locations. The map on the left shows locations of all 14,301 herds included in the EBL Control Scheme. The map on the right shows locations of EBL positive herds in the testing period 1997-1998.

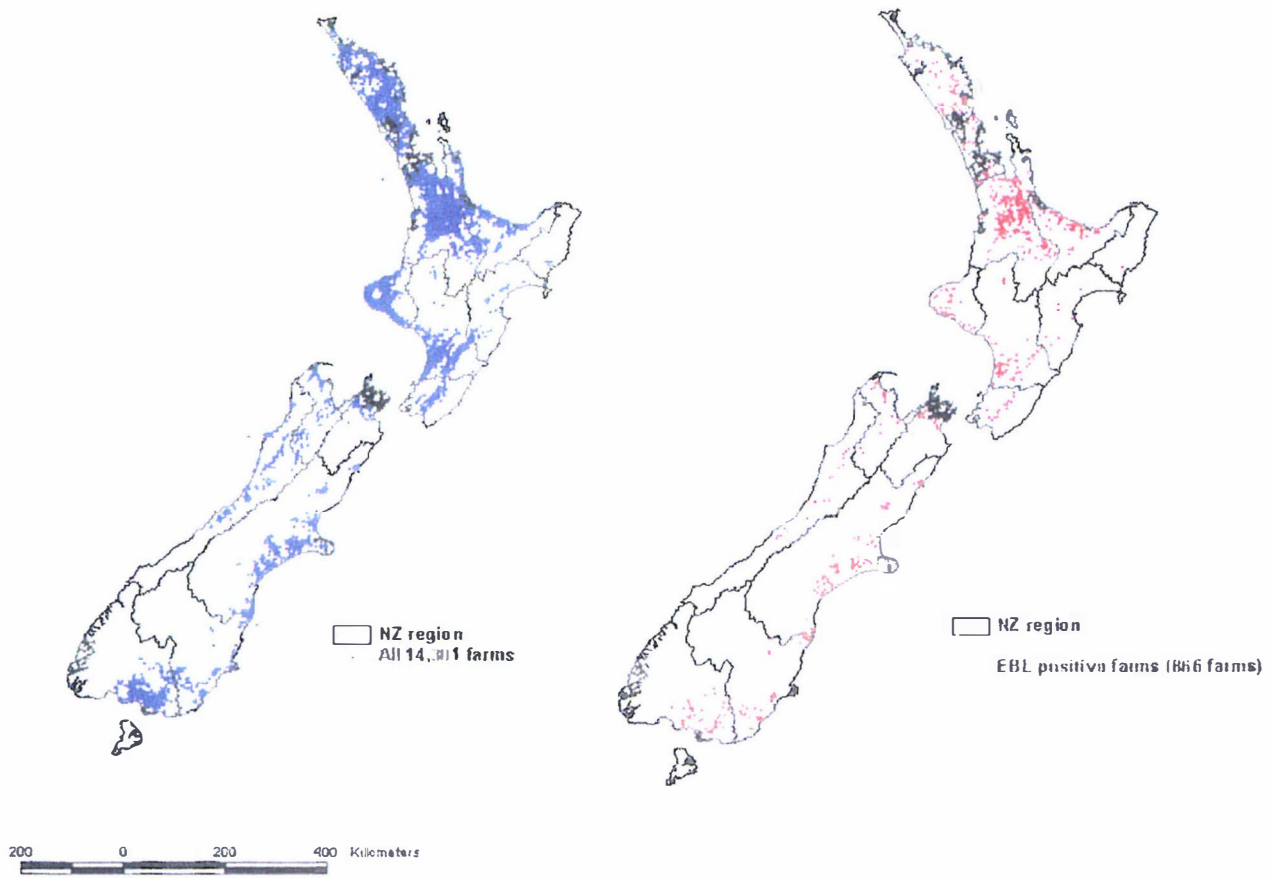


Table 17. Herd status according to EBL testing scheme for 14,301 dairy cattle herds in New Zealand in the testing period 1997-1998 (herds without locational coordinate reference were excluded).

EBL Status	*N*
Blood positive	866
EBL free	842
Individual milk positive	142
Negative year 1	5,452
Negative year 2	4,231
New location	193
Pool milk positive	108
Provisionally negative *	107
Suspect *	454
Untested	124
Vat negative	1,782
Total	14,301

* Provisionally negative: seroconversion of EBL titre from positive to negative;

* Suspect: herds introduced cattle from the positive herd or other suspected herds.

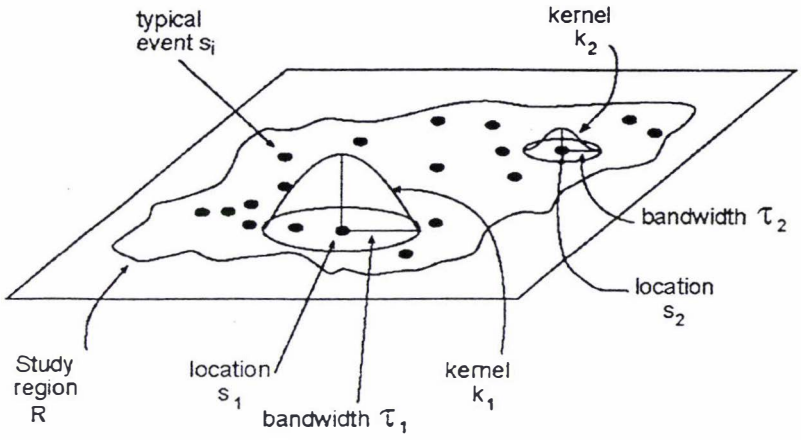
Analytical methods

Kernel smoothing

The goal of kernel smoothing of point patterns is to describe variation in the density of events across a study area. It results in a smooth map of density values where the density at each location reflects the concentration of points in the surrounding area (Bailey and Gatrell 1995).

Kernel smoothing can be thought of as a three-dimensional floating mathematical function that moves across an area. At each location defined by a suitably chosen fine grid, it estimates the density of points within a circle of a given radius (bandwidth), whereby the influence of individual points is weighted according to there distance from the centre of the circle (Figure 7). It is therefore critical to choose the appropriate bandwidth for the process being studied as well as the best probability distribution that will determine the distance weighting of observations.

Figure 7. Kernel estimation (from Bailey and Gatrell, 1995)



There is no steadfast rule for determining bandwidth, although if inappropriate, it may result in misleading density values and maps that are either too smooth or too spiky in appearance. For example, if the bandwidth is too large it will over-smooth the data and only very broad trends will be visible, whereas if the bandwidth is too small, localised variation will be maintained and interpretation will be more difficult. Choosing the correct bandwidth is crucially important to density estimation (Silverman, 1986). Often a subjective judgement is made. Silverman suggests that this might be desirable if the purpose of the investigation is to explore the data in order to develop possible hypotheses, as is the case in the current study.

Ratio of kernels

In order to interpret the kernel smoothed density map of occurrence of EBL herds, it was necessary to take into account any spatial heterogeneity in the density of the population at risk i.e. all the New Zealand dairy herds included in this analysis. Kernel density maps were therefore estimated for the positive herds as well as for the population at risk. Instead of a purely visual comparison between the two maps, the ratio of the two kernel estimates was then used to generate a map expressing the prevalence of BLV infection. Following the recommendation by Bailey and Gatrell (1995) a larger bandwidth was deliberately chosen for the kernel density estimates of the population at risk than would be normally appropriate if one were just interested in an estimate of the population density alone. This provides a surface that will be 'over smooth', and less sensitive to small-scale variation.

Statistical spatial cluster analysis

There are a wide variety of statistical methods for detecting spatial clusters (Marshall, 1991; Jacquez, 1993). The spatial scan statistic (Kulldorff and Nagarwalla, 1995) was used as it allows statistical inference with respect to the presence of spatial clustering as well as identification of the locations of any clusters.

This method generates for each location in the data set a set of circles (or windows) with ever-increasing radius. It performs a likelihood ratio test based on case numbers within and outside the circles. The null hypothesis distribution of the likelihood ratio is obtained on the basis of Monte Carlo replications. The 'most likely' cluster is defined by the highest value for a significant likelihood ratio statistic in the data set. Other statistically significant clusters are also identified as 'secondary' clusters. To limit the analysis to detection of biologically sensible cluster sizes, it was decided to use a maximum window size of 10% of the total population. The test statistic was calculated based on 999 random replications.

Geographical characteristics and variation in dairy cattle herd density were used to divide the country into 5 separate study areas for analysis purposes based on geographical and herd density considerations. Each area was analysed separately. The extent of each region is summarised in Table 18. A map of the regions is shown in Figure 8.

Table 18. Characteristics of the study areas used in the analysis.

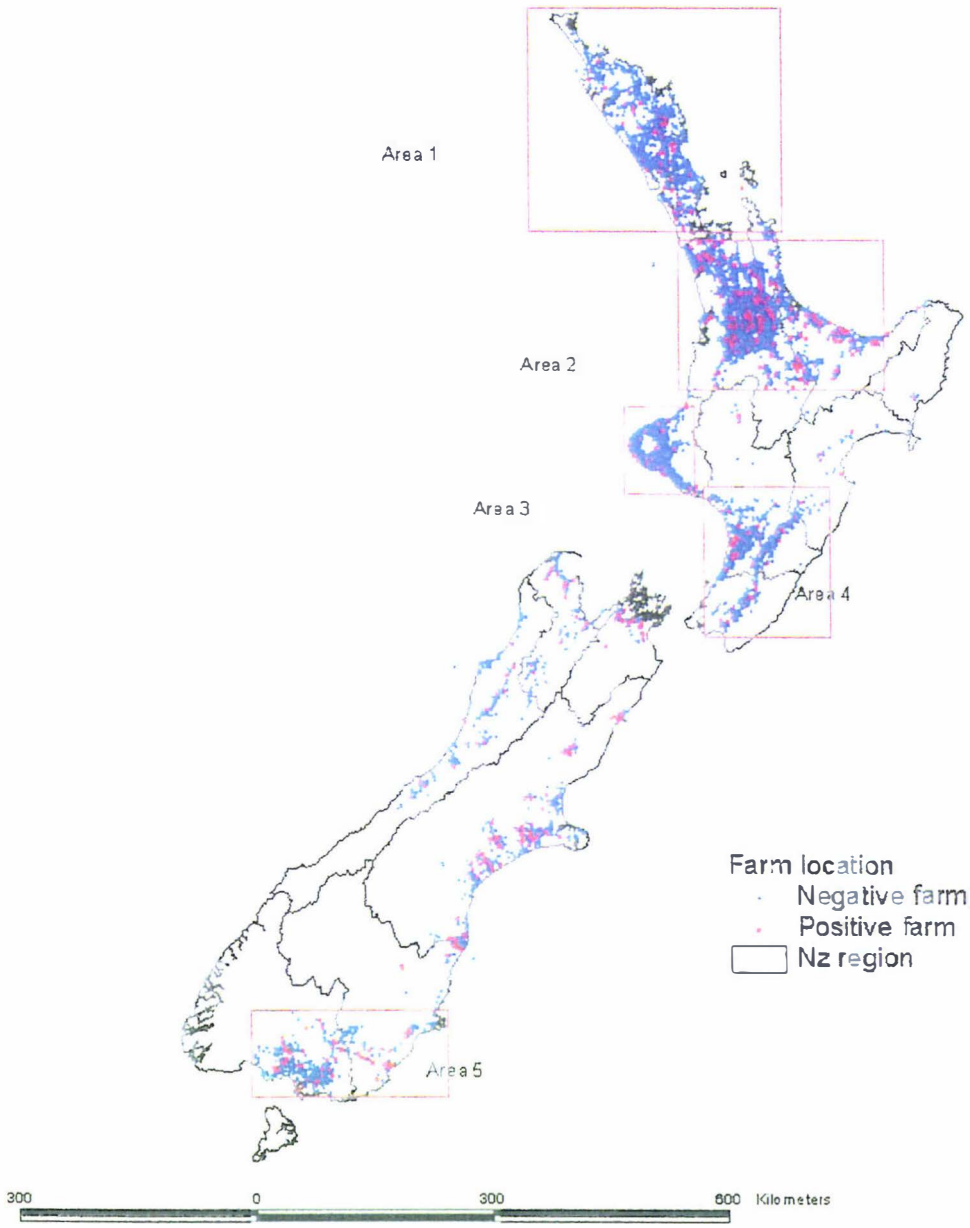
Study area	*Region*	*X, Y coordinates of study areas top right and bottom left corner*	*Number of herds (EBL positive herds)*
1	Northland	(6760000,2770000),(6480000,2450000)	1,782 (87)
2	Bay of Plenty	(6470000,2900000),(6280000,2640000)	6,599 (388)
3	Taranaki	(6260000,2660000),(6150000,2570000)	2,415 (46)
4	Wellington-Wairarapa	(6160000,2830000),(5970000,2670000)	1,338 (77)
5	Southland	(5500000,2340000),(5390000,2090000)	711 (85)

Software

Data was stored using the database management software Microsoft Access 97 (Microsoft Corporation, Redmond, WA, U.S.A.) and analysed statistically using

SaTScan® version 2.1.3 (National Cancer Institute, Bethesda, Maryland, U.S.A.). The geographic analysis system ArcView 3.1® in combination with the add-in Spatial Analyst (both Environmental Systems Research Institute., Redlands, California, U.S.A.) was used to generate the maps and perform the kernel density smoothing using.

Figure 8. Map of the location of the 5 study regions within New Zealand



Results

Visual analysis of the spatial pattern of herd density and EBL prevalence

Using the *kernel smoothing* technique, density maps were generated using different bandwidths including the default used by the ArcView software as shown in Figures 9 to 12. Bandwidth values of 10 km (Figure 9), 15 km (Figure 10), 20 km (Figure 11) and 28.6 km (Figure 12) were used. The last value was automatically calculated by ArcView whereas the others were considered to provide a reasonable compromise with respect to providing a pattern that had too many spikes or was too smooth. From these density maps, spatial patterns of herd prevalence were generated on the basis of the ratio between the density of EBL positive herds and total herds (Figure 13-14). The prevalence maps using the same bandwidth for both positive and all herds were compared with the prevalence maps that used a larger bandwidth for all herds than used for the total herds in order to compare the effect of variation in relative bandwidth.

Statistical analysis of spatial clustering of EBL positive herds

From the five study areas for which separate analyses were conducted, a significant spatial cluster could only be identified within the Bay of Plenty region. The 'most likely' cluster consists of 498 herds with 65 EBL positive herds and has a radius of 47.8 km. The overall relative risk for this cluster is estimated as 2.2 ($p = 0.001$). Results are summarized in Table 19.

Study areas were then re-aggregated into two large areas, representing the South Island and the North Island of New Zealand, and the analyses were repeated for these two data sets. The same cluster in the Bay of Plenty region was confirmed at a slightly higher likelihood ratio and overall relative risk of 25.5 and 2.6 respectively ($p = 0.001$). In addition, in the South Island data set another cluster of infected herds with a radius of 101 km was identified in the north. The overall relative risk for this cluster was estimated as 2.4 ($p = 0.08$). The results of these analyses are presented in Table 20. Maps showing the locations of these two clusters are presented in Figure 15 and 16.

Figure 9. Kernel density maps for all herds (left) and EBL positive herds (right) locations based on a bandwidth of 10 km

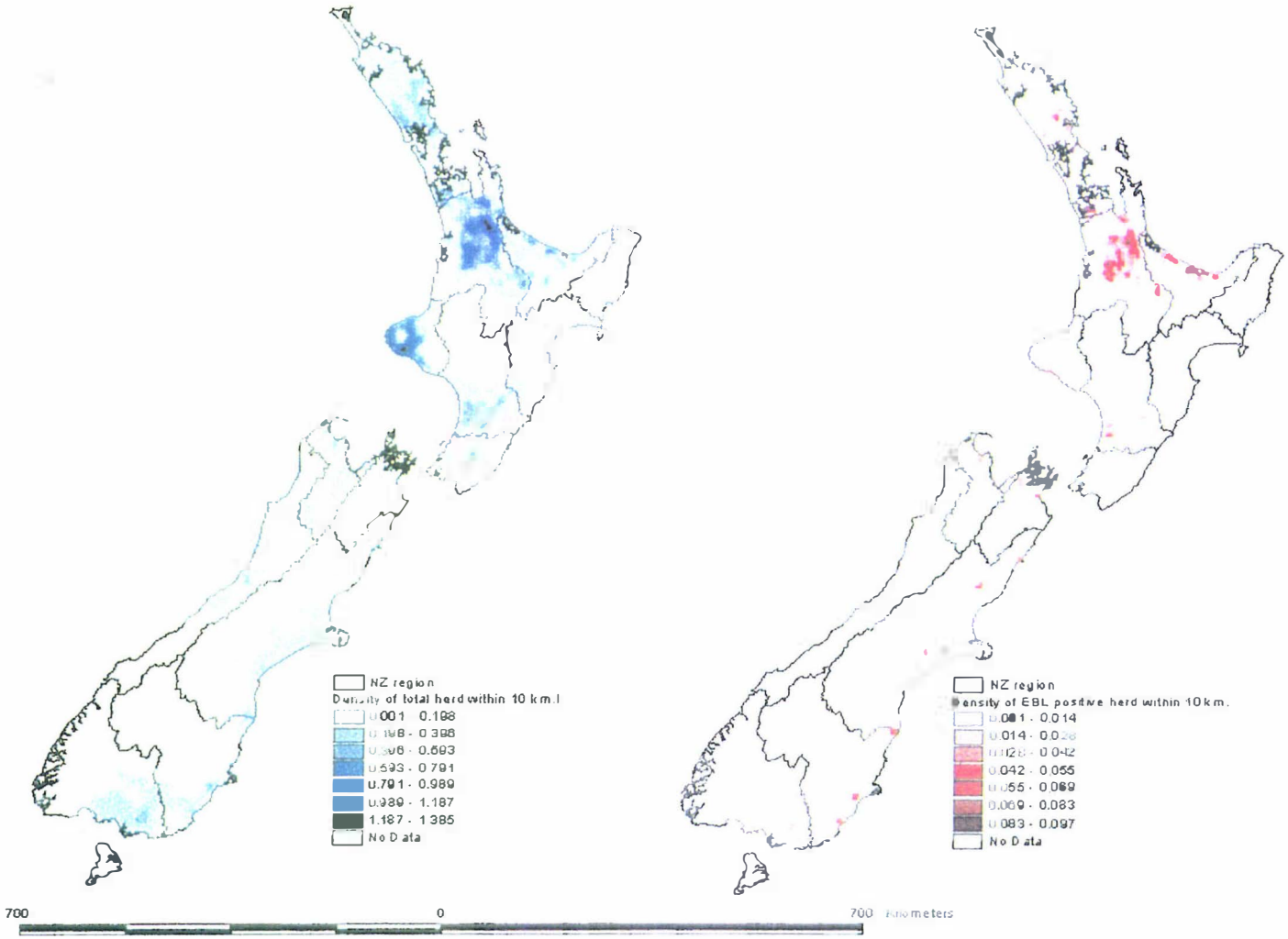


Figure 10. Kernel density maps for all herd (left) and EBL positive herd (right) locations based on a bandwidth of15 km

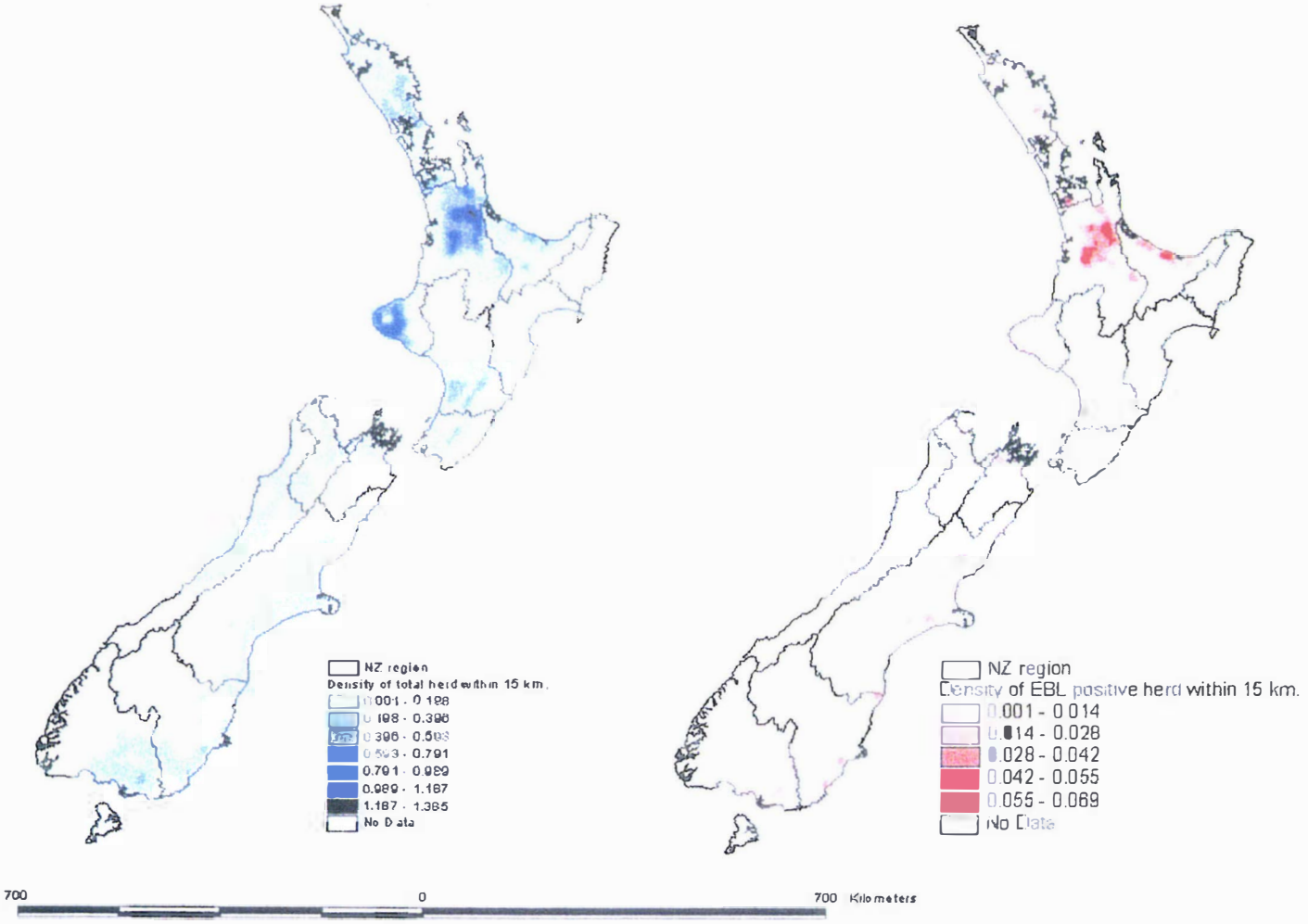


Figure 11. Kernel density maps of all (left) and EBL positive (right) herd locations using a bandwidth of 20 km

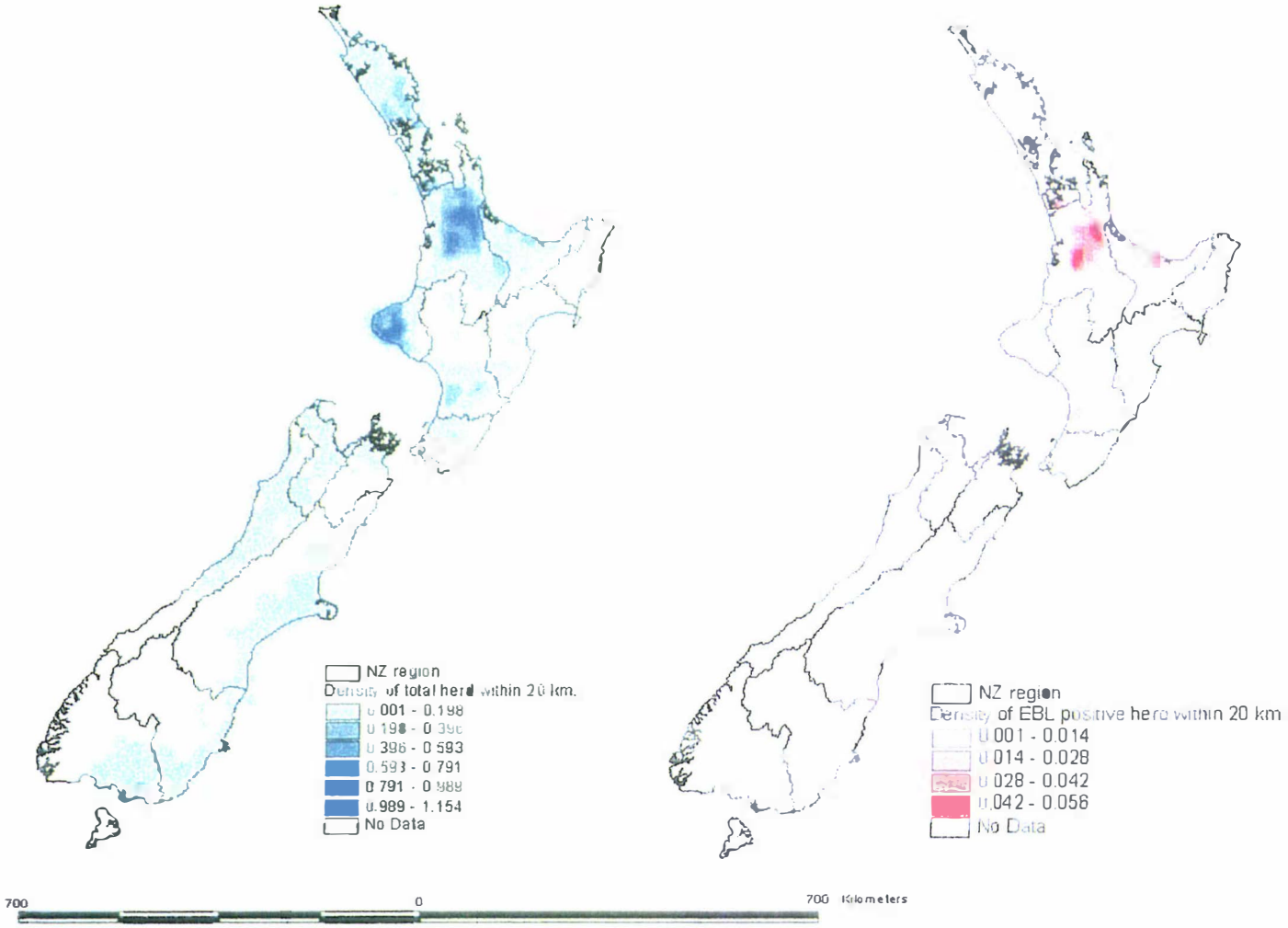


Figure 12. Kernel density maps of all herd (left) and EBL positive (right) herd locations based on a bandwidth of 28.7 km (ArcView default)

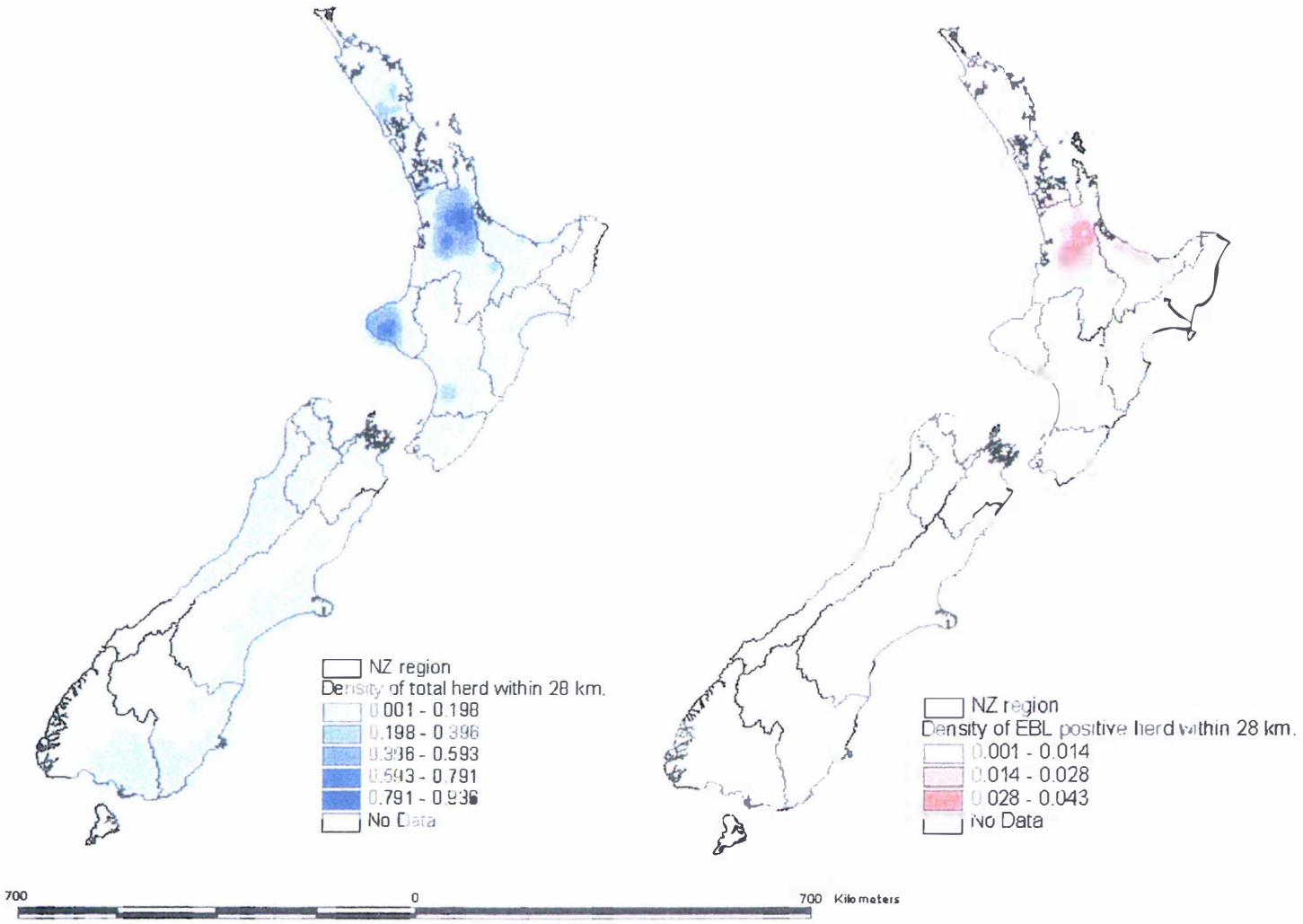


Figure 13. EBL prevalence maps based on ratios of kernel density maps for EBL positive herds and all herds using bandwidths of 20 km for both (left) and 20 km for EBL positives and 22 km for all herds (right)

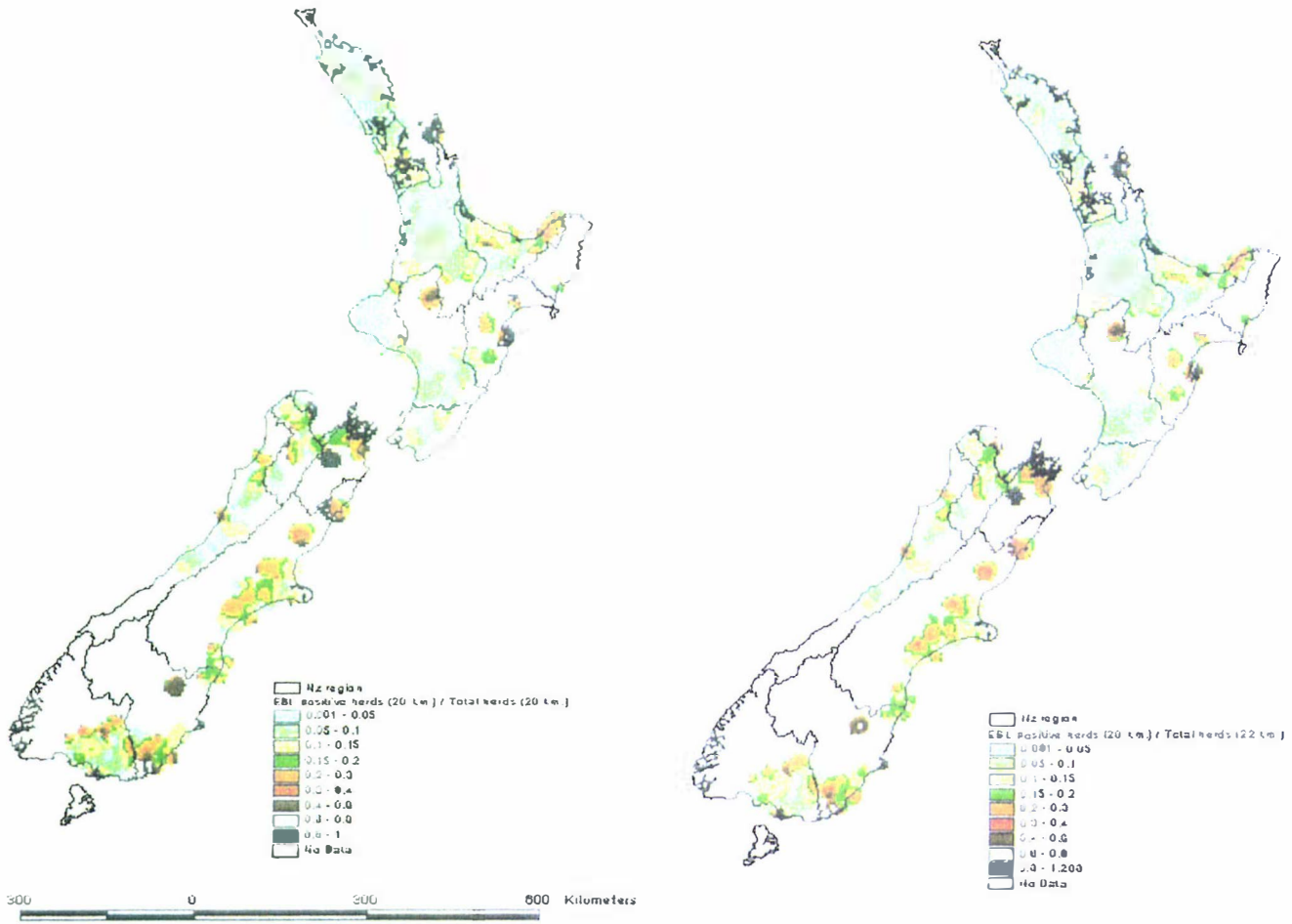


Figure 14. EBL prevalence maps based on the ratio of kernel density maps of EBL positive and all herds using a bandwidth of 28 km for both (left) and 28 km for EBL positives and 30 km for all herds (right)

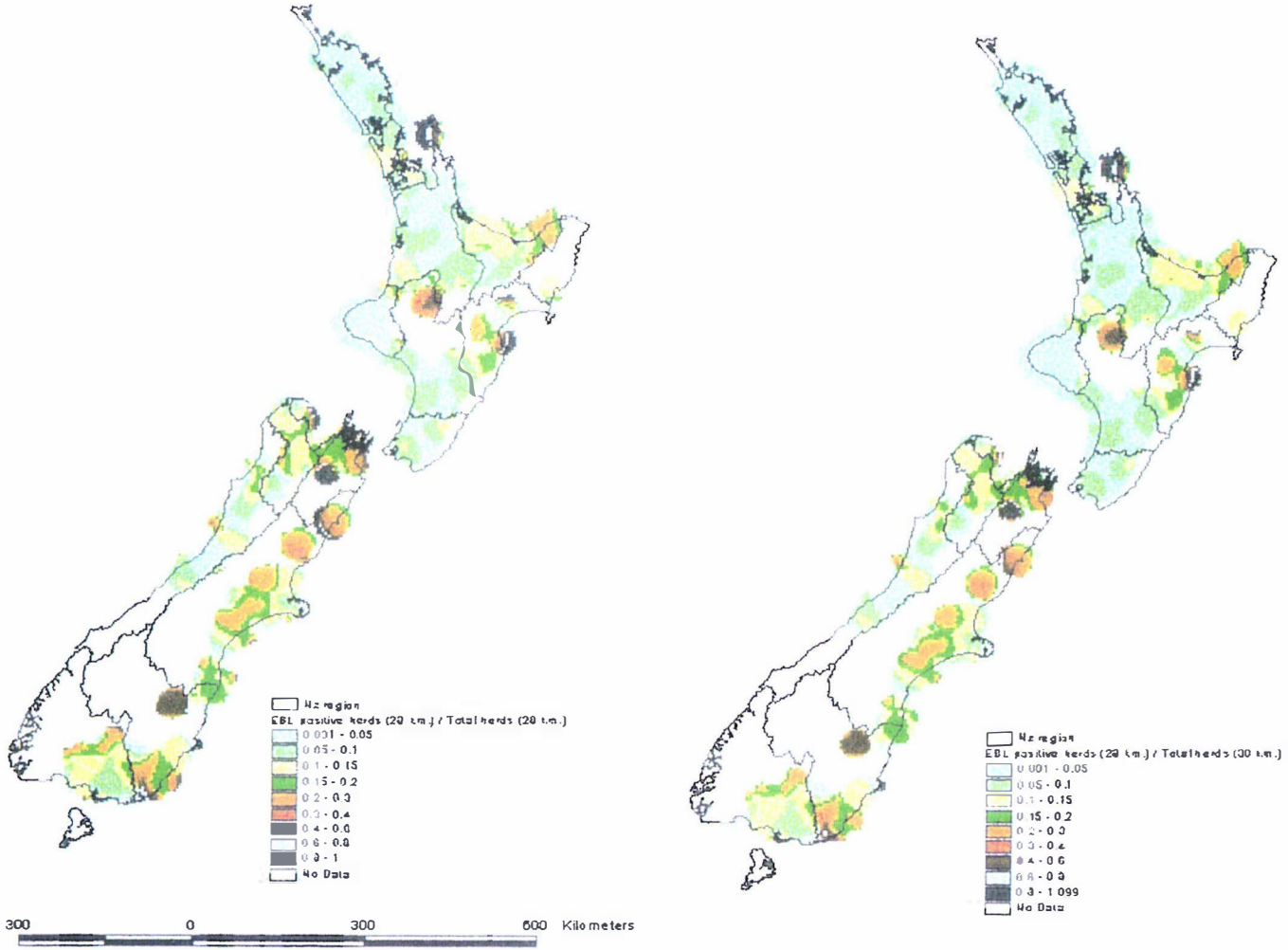


Table 19. Statistics generated by the spatial scan statistic for identified 'most likely' clusters of EBL positive herds in each of the five study regions in New Zealand

Study region	*'Most likely' cluster*						
	*Radius size** *(km.)*	*Total herds* *n*	*Cases* *n*	*Expected* *n*	*LLR** λ	*RR**	*p value*
1	2.1	10	5	0.5	8.5	10.2	0.18
2	*47.8*	*498*	*65*	*29.3*	*19.5*	*2.2*	*0.001*
3	0.8	2	2	0.04	7.96	52.5	0.18
4	0.3	2	2	0.1	5.7	17.4	0.67
5	13.4	14	7	1.7	6.3	4.2	0.59

* Radius size = radius of the circle corresponding to that potential cluster;

* LLR = log likelihood ratio;

* RR = Relative Risk for herds within the cluster compared with herds in the remainder of that pre-defined area.

Table 20. Statistics generated by the spatial scan statistic for the 'most likely' clusters of EBL positive herds in the North and South Island of New Zealand

Study region	*Most likely cluster*						
	*Radius size** *(km.)*	*Total herds* *n*	*Cases* *n*	*Expected* *n*	*LLR** λ	*RR**	*p value*
North Island	47.8	498	65	24.9	25.5	2.6	0.001
South Island	101.6	83	25	10.3	9.6	2.4	0.08

* Radius size = radius of the circle corresponding to that potential cluster;

* LLR = log likelihood ratio;

* RR = Relative Risk for herds within the cluster compared with herds in the remainder of that pre-defined area.

Figure 15. Locations of farms forming part of the significant 'most likely' cluster of EBL positive herds in the Bay of Plenty region of the North Island of New Zealand

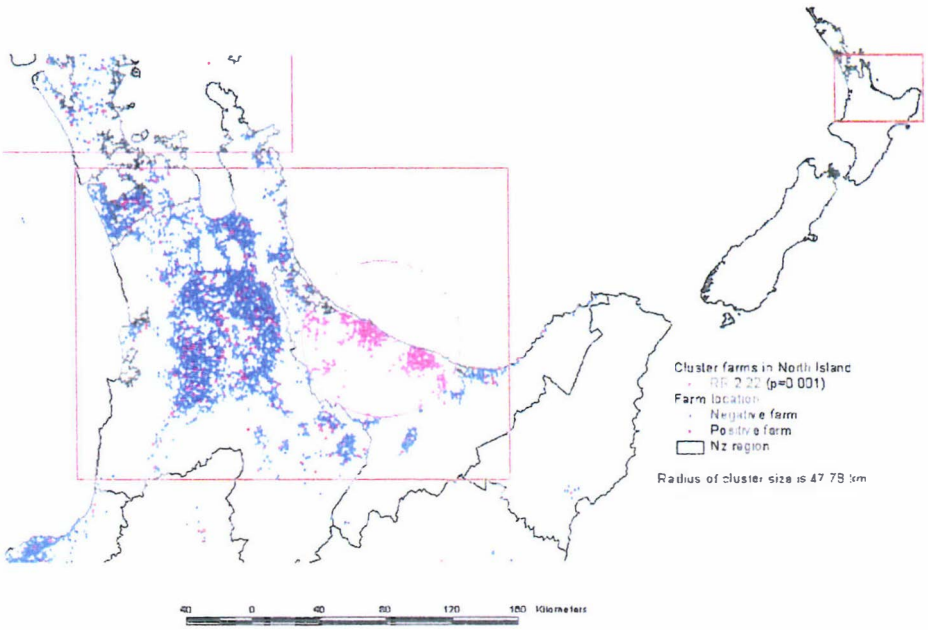
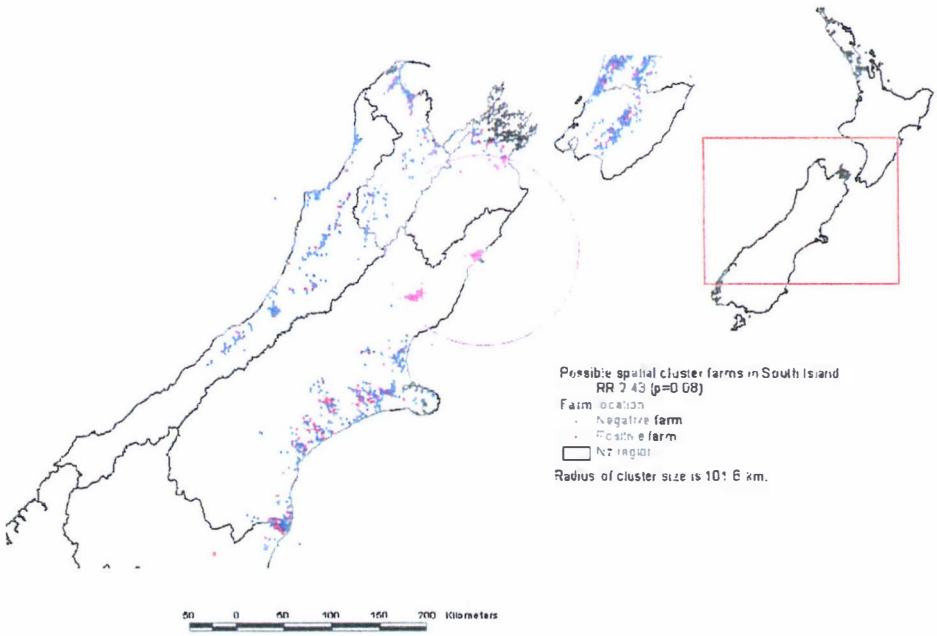


Figure 16. Locations of farms forming part of the 'most likely' cluster of EBL positive herds in the northern part of the South Island of New Zealand.



Discussion

Visualization of herd density and EBL prevalence using kernel density estimation

The point location maps of all and infected herds in Figure 6 clearly show the spatial heterogeneity of the distribution of both EBL positive as well as all dairy herds in New Zealand. Yet, it is difficult to identify any patterns within dense clusters of overlaying points. Using kernel density estimation it is possible to generate a smooth surface map. The influence of the choice of bandwidth on the appearance of the resulting smoothed map is shown in Figures 9 to 12. With bandwidths of 10 and 15 km, the resulting maps show many spikes resulting from high estimated density values. Increasing the bandwidth to 20 km resulted in a smooth map that still reflected the detail of local as well as large scale variation. Using the largest bandwidth estimated by the ArcView software, a very smooth map was generated which does not show local variation of population but offers a general impression of the overall distribution of EBL positive herd occurrence. This shows that the choice of bandwidth for kernel density estimation is a largely subjective process and that it has a strong influence of the resulting map presentations. Therefore, a decision with respect to bandwidth should consider whether the focus of the investigation is on large or small scale processes, or both. In this study, our emphasis was more on local processes and thus the application of a narrower bandwidth was appropriate. At the same time care had to be taken to ensure that the bandwidth chosen was not too small as this would result in data noise dominating the density estimates.

The EBL herd prevalence map produced on the basis of the ratio of kernel density estimates for EBL positive herds and all dairy cattle herds as the population at risk does provide a useful visual representation of spatial variation in disease occurrence particularly at the regional level. For example, in Northland herd prevalence is consistently less than 5% whereas in the Bay of Plenty region it varies between 5 and 15%. These figures are within the range reported by the New Zealand Dairy Industry EBL Control Scheme (Hayes, unpublished paper). Figures 13 and 14 demonstrate how local variation can be masked by increasing bandwidth.

On the basis of the visual inspection of EBL herd prevalence maps it was possible to tentatively identify potential clusters on the basis of observed local peaks in EBL prevalence. However, estimates produced by this technique where data is sparse may be unreliable. Estimation of confidence limits could be used to highlight these areas. Bailey and Gatrell (1995) suggest using a larger bandwidth to ensure a reasonable size denominator. Adaptive kernel estimation allows for local adjustment of bandwidth which does preserve detail in high density areas and smoothes out spurious noise in sparse data areas (see the right maps of Figure 13 and 14).

This analysis has shown that examination of kernel density maps represents a useful step in the search for clusters which often begins as an 'exploratory spatial analysis' without any prior assumptions or specific hypotheses. However, it should not be seen as an end in itself (Rushton and Lolonis, 1996).

Statistical of spatial clustering of EBL positive herds

The only significant cluster detected in this analysis consists of 498 herds in the Bay of Plenty area. Another possible cluster was identified in the northern part of the South Island. Both clusters have estimated risks for EBL positive herd status that are 2.5 times greater than that of the area outside the cluster. In the case of the Bay of Plenty cluster it is interesting to note that it is on a relatively narrow strip of coastal land, and in comparison with the one in the South Island, it is relatively compact. While it is not possible to develop any causal hypotheses on the basis of these results, it would appear worthwhile to follow up these findings with a more in-depth investigation of affected herds in both areas, but particularly in the case of the Bay of Plenty cluster. The cluster in the South Island was significant at the $p=0.08$ level and, therefore, may be due to chance.

Cluster investigations involving the use of statistical methods are easy to conduct, yet they are often difficult to interpret. The results can be used to guide further investigations, but it should be kept in mind that failure to identify clusters in the data does not necessarily mean that they are not there, because the techniques are known to have limited statistical power. The objectivity resulting from the use of a statistical analysis approach such as the spatial scan statistic without involving a post-hoc hypothesis generation will make it possible though that the use of resources for more in-depth investigations can be more easily justified (Kulldorff and Nagarwalla, 1995).

The analytical approach described in this paper provides a general framework for the detection of spatial clusters in animal populations. Both, kernel density maps of disease risk and the spatial scan statistic could become part of a disease surveillance system which would allow identification of unusual occurrences of high disease density. It does not provide information about etiological or causal mechanisms of the underlying disease process, but rather can indicate areas that should be examined more closely through specific epidemiological investigations.

Reference List

Bailey, T.C., Gatrell, A.C., 1995. Interactive spatial data analysis. Longman Scientific & Technical, New York, U.S.A.

Burton, L., Allen, G., Hayes, D., Pfeiffer, D., Morris, R., 1997. A novel approach to disease control. An industry operated programme for bovine leukaemia virus in New Zealand. 31-32. 08.08.1- 08.08.3.

Ferrer, J., F., 1980. Bovine Lymphosarcoma. In: Advances in veterinary science and comparative medicine. Academic Press, 1-68.

Hayes, D., 1998. Enzootic bovine leucosis eradication scheme. Surveillance, 25. 3-5.

Hilbink, F., 1993. Penrose, M., Prevalence of enzootic bovine leukosis in the New Zealand dairy cattle population. 20. 11-13.

Jacquez, G.M., 1993. The statistical description of disease clusters. Statistics in Medicine, 12: 1967-8.

Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. Statistics in Medicine, 14. 799-810.

Marshall, R.J., 1991. A review of methods for statistical analysis of spatial patterns of disease. Journal of the Royal Statistical Society Series A, 154. 421-441.

Rushton, G., Lolonis, P., 1996. Exploratory spatial analysis of birth defect rates in an urban population. Statistics in Medicine, 15. 717-26.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman and Hall, New York

CHAPTER 4.

Spatial Logistic Regression and Classification -Tree Models for the Development of Risk Maps of Foot-and-Mouth-Disease Outbreak Occurrence in Thailand

Abstract

If the spatial distribution of disease outbreaks can be predicted with acceptable accuracy, then intervention programmes can be targeted at the most appropriate areas, and control operations can be designed to maximise effectiveness. In this study, logistic regression and classification tree models were used to develop predictive models of the geographical patterns of foot-and-mouth-disease. The potential importance of spatial autocorrelation was assessed through a comparison of different logistic regression models. The model that appeared to take best account of spatial dependence included an indicator variable for occurrence of disease outbreaks in neighbouring areas. Receiver-operating characteristic (ROC) curves were used to describe the predictive accuracy of the model for production of risk maps. Classification tree based models were constructed to develop sets of decision rules for FMD outbreak occurrence. We compared a number of classification tree models (CART) which were based on different cost-sensitivity weightings for false-positive and false-negative model predictions. The classification tree model based on giving false negatives 5 times the weight of false positives was also used to generate a risk map. The sensitivity and specificity of the classification trees were 89% and 45%, respectively. The predictive accuracy of risk maps produced using CART and the logistic model was compared and their value in designing FMD control programme was discussed. It was shown that ROC curves are useful for choosing appropriate cut-off points for decision criteria that allow the effectiveness of control and eradication programmes to be maximised.

Introduction

Foot-and-mouth disease (FMD) is an economically important viral disease of cloven-footed animals that is endemic throughout Asia, Africa, and most of South America. The economic consequences of FMD outbreaks are severe in that importation of livestock and livestock products from areas that are not free from FMD is banned by many countries. A *National Foot-and-Mouth Disease Control and Eradication Project* has been established in Thailand with the objective to eradicate the disease from the country.

Epidemiological studies of factors related to disease outbreaks, patterns of disease occurrence and factors influencing the spread of the disease are given a high priority in this project and the information gained aids the planning and implementation of new and existing control and eradication programmes.

Modelling of spatial data

In recent years, analytical methods for identifying geographic disease patterns have been extensively employed by epidemiologists to test epidemiological hypotheses about cause-effect relationships. Geographical information systems (GIS) are used to produce maps and, more recently, such systems also allow the exploration of spatial patterns through dynamic linking of windows. From the spatial analytical perspective, there are great advantages to be had from linking statistical and mathematical models to disease surveillance databases and from data display capabilities of GIS programmes.

General concepts of spatial data

A basic property of data with a locational component is that a set of values, $\{x_i\}$, are likely to be related over space and this idea underlies the concept of the region in geography, graphically described by various authors (Cliff, 1981). Tobler (1970) refers to "the first law of geography: everything is related to everything else, but near things are more related than distant things". Stephan (1934) writes, "data of geographic units are tied together like bunches of grapes, not separate like balls in an urn".

Heterogeneity in a spatial context means that the parameters describing the data vary from place to place. Non-stationarity, or spatial heterogeneity, occurs when the process

observed in a window (or kernel) changes systematically as a result of either the presence of a trend in the data or a change in the variance.

If the distribution function of the stochastic process remains unchanged when distance changes by an arbitrary amount then the process is stationary and *spatially homogeneous*. That is, space-homogeneity is restricted to a function of the distance between the elements of the distribution in question. *Spatial dependence or spatial autocorrelation* is a special case of spatial homogeneity. It implies that the data for particular spatial units are related and similar to data for other close spatial units in a spatially identifiable way (Fotheringham and Rogerson, 1993).

Hypotheses about patterns, or estimation of relationships, may be tested using statistical models of the relevant data, for instance between some measure of disease incidence and social and/or environmental covariates. Although we may test and search for spatial autocorrelation in data, the detection of spatial patterns is often not an end in itself. Instead, interest focuses upon the development of spatial models. It should always be kept in mind that the models could not be drawn from the classical statistical models that assume independence of events. For a number of reasons, disease incidence in one zone is likely to be spatially correlated with that in neighbouring zones. Essentially, spatial models provide alternative hypotheses against which the null hypothesis of no spatial autocorrelation can be tested, thus enabling comparison of the performance of the tests against selected alternatives.

Indeed, it may be appropriate to fit a regression model and then examine the residuals for spatial dependence, or to fit the model which incorporates both regression and spatial autocorrelation. The presence of spatial autocorrelation may be attributable either to trends (or gradients) in the data or to interactions, and if gradients are suspected then a regression model is appropriate. Thus, significant autocorrelation in the original data does not imply one model rather than the other. The choice of model must involve the scientific judgement of the investigator and careful testing of the assumptions.

As mentioned earlier about the basic concept of geography, neighbouring areas tend to have similar conditions, and if available covariates do not fully reflect the conditions then the residuals from a fitted model will exhibit spatial autocorrelation. Furthermore, quite apart from the environment, the probability of occurrence of FMD in one area

might not be independent from the ones in neighbouring areas because infectious diseases tend to cluster as a result of the occurrence of risk factors. This then will generate spatial autocorrelation that can not be modelled satisfactorily by environmental covariates. By using models that allow for spatial autocorrelation, we would hope for fewer covariates in an empirical model for occurrence, and perhaps obtain a better indication of which covariates influence the distribution or occurrence of the disease under investigation.

There is an important need to examine the regression residuals for autocorrelation and not the original data. The presence of autocorrelation leads to biased estimates of the residual variance and inefficient estimates of the regression coefficients. Therefore a check for autocorrelation in the residuals should always be applied and remedial action taken when necessary. A test of autocorrelation will tell whether a given model is adequate, or whether a different form is required. There is no need to completely remove spatial autocorrelation as such, but allowances should be made for it so that valid estimation procedures can be adopted. Typical measures for expressing the autocorrelation include parameters of variables or error terms in spatially autoregressive systems such as Moran's *I* statistic, Geary's c statistic, semi-variogram models, or spatial adaptive filter parameters.

Logistic regression modelling of spatial data

Logistic regression has been applied to the modelling of spatial data, especially as part of environmental and ecological studies (Bian and West, 1997;Austin, *et* al., 1996). In recent years, reducing spatial data to its dimensionality for logistic-regression analysis has had veterinary application to the incidence of livestock disease, or the distribution of vectors capable of carrying disease (Norman, *et* al., 1996; Duchateau, *et* al., 1997). However these latter studies ignored model dependence on either unmeasured covariates or intrinsic spatial autocorrelation. Pfeiffer, *et* al. (1997) on the other hand produced a predictive model for the occurrence of theileriosis outbreaks in Zimbabwe, taking into account spatial autocorrelation based on inclusion of an indicator variable representing local regions as a random effect.

Classification tree modelling

The basic purpose of generating a classification tree model (also called classification or regression trees) can be to produce an accurate set of decision rules for classification of future cases or to understand the predictive structure of the problem (exploratory data analysis) or both as is often the case (Breiman, *et* al., 1984).

To predict systematically the class of any object, four elements are needed for the entire tree construction (Breiman, *et* al., 1984):

- A set of questions which must be formed to generate a set of data splits.
- A goodness of split criterion that can be evaluated for any split of any node. The split selection procedure can be thought of as a repeated attempt to minimise overall tree impurity.
- A stop-splitting rule which can be based on two methods: deciding not to divide a set of cases any further (stopping), or removing retrospectively some of the structure built up by recursive partitioning (pruning).
- A rule for assigning every terminal node to a class. For any set of assignment rules, the re-substitution estimate of the probability of misclassification is calculated for a case falling into a node.

During the analysis, the program determines at each step for each variable a cut point which optimally splits the population into classes and it then selects the best performing variable. It then takes the resulting sub-populations and repeats the process on each of them until no further partitioning is warranted. Tree diagrams produce logical classification rules that can be easily interpreted, communicated, and discussed. The method has also been used extensively in the medical domain.

In general, classifier learning has focused on minimum error classification. It aims to minimise the number of incorrect predictors or classifications made by classifiers. This kind of learning method ignores the differences between different types of incorrect prediction and in particular their costs that are often very relevant in real world applications. The cost of incorrect predictions can be more important than the number of incorrect predictions in medical and financial areas. In medical diagnosis, for example, the costs of false negative diagnoses are usually considered to be much higher than those of false positives. Breiman *et al.* (1984) describe two methods of

incorporating variable misclassification costs into the process of tree induction. These methods adapt the test selection criterion in the tree growing process. A number of studies on this issue have been done (Tan, 1993; Turney, 1995; Webb, *et* al., 1996; Ting, 1997; 1998). For FMD the cost of not predicting an actual outbreak (false negative) is far higher than predicting an outbreak where none does in fact occur (false positive). The differential weightings of false negative and false positive predictions were taken into account in the analyses reported here.

Study aims

The aim of this study was to produce 'risk-maps' for FMD occurrence for the northern and western parts of Thailand using statistical models based on geographically referenced data.

Material and Methods

Study area

The Kingdom of Thailand, covering an area of 514,000 square kilometres, lies in the heart of Southeast Asia and shares borders with Myanmar to the west and north, Laos to the northeast, Cambodia to the east and Malaysia to the south. The country has four distinct areas: the mountainous North, the fertile Central Plains, the semi-arid plateau of the Northeast, and the peninsular South. It contains 76 provinces (*changwat*), 844 districts (*amphoe*), and 6,404 sub-districts (*tambon*).

The area considered here comprised 67 provinces, and their associated 590 districts and did not include the FMD free zones in the southern part (region 8 and 9 of Department of Livestock Development (DLD) Regions) and the eastern region (DLD region 2) of Thailand (Figure 17).

Data layers

Geographical data

All coverages used in this study were originally captured in vector format to represent geographical features at a scale of 1:250,000. They were converted into raster-based format for the analyses because of more convenient data storage and manipulation and

easier programming for analyses. In a raster-based map, cell size is best determined based on areal dimensions which are considered to provide an adequate representation of the data. The original data were stored at the district level with the exception of the socio-economic themes which were summarised at the sub-district level. All layers were converted to raster format resulting in a total of 661 cells each representing an area of 25 x 25 kilometres. The geographical data sources that were available for this study and their characteristics are listed in Table 21. Figure 18 presents two examples of raster format maps.

FMD outbreak location data

Routine surveillance disease reporting data for foot-and-mouth-disease in Thailand from January 1995 to May 1997 were used in this analysis. A total of 113 outbreaks of FMD occurred during the period. Each record of an outbreak was recorded at a spatial aggregation at the district level and included the following information: month/year, FMD virus type (Type O, Asia1 and untyped), species affected, number of cases and district. The exact location of disease outbreaks was not available, so the number of outbreaks within each district was aggregated for the reporting period. As the resolution of the district level was considered to crude, random coordinate locations were generated for each outbreak within the corresponding district boundaries using the ArcView Avenue programming language. The outbreak locations were then aggregated into the cells of the raster cell layer.

Figure 17. Map of administrative boundaries showing area included in this study

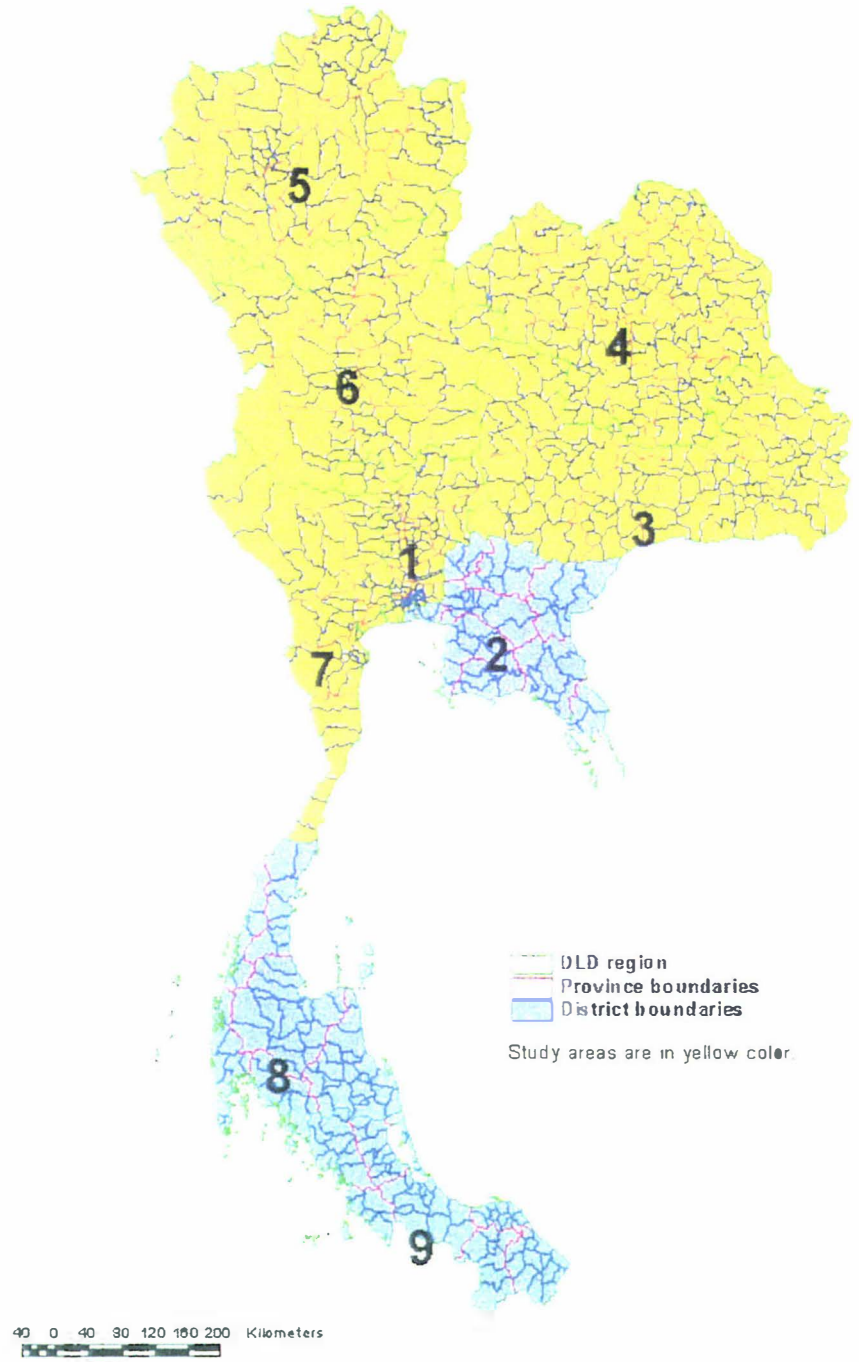


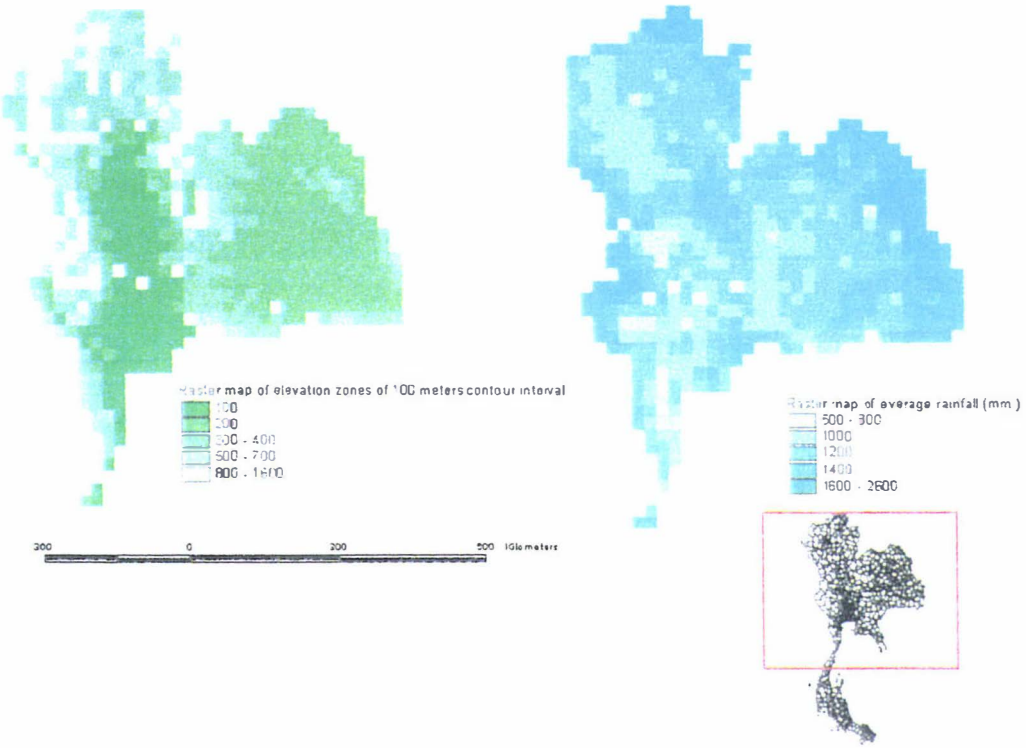
Table 21. Characteristics of geographical and attribute data in the study area..

Geographical data set (feature type)	*Coverage name*	*Description*
Administrative (Polygon)	*Amphoe* Boundaries	A total of 590 *amphoes* or districts (excluding districts in Bangkok and districts in region 2, 8, and 9)
Topography (Polygon, Arc)	Elevation	Elevation level at 100 meter contour intervals
	Streams and rivers	Classification for type of stream grouped into 3 classes
		Major river or canal
		Perennial stream
		Intermittent stream, irrigation canal, man-made canal, manmade reservoir, perennial lake, intermittent man-made canal, intermittent lake
Environmental (Polygon, Point)	Annual rainfall	Average annual rainfall measured in millimetres
	30 -year average monthly rainfall	30-year average monthly rainfall from weather recording station summarised in 3 seasons; summer, rainy, and cool season. [a]
Forestry (Polygon)	Forest	Forest area (Forest Reserve, National Park, Wildlife Sanctuary) and no forest area
Agricultural (Polygon)	Land use	Major land use grouped into 4 classes
		Agricultural land type 1
		Agricultural land type 2
		Forest land
		Miscellaneous (Golf course and recreation area, industrial land, urban and built up land, water body)
Infrastructure (Arc)	Transportation and routes	Type of transportation features grouped into 3 classes
		Two or more lanes wide
		One lane wide
		Loose surface in fair or dry weather, cart track, footpath, trail, and railroads
Rural socio-economic (Polygon)	Human	Socio-economic information as the attribute data at the *tambon* or sub-district level focuses on rural Thailand, widely grouped into 6 groups according to these kinds of factors
		Animal movement
		Farmer education
		Public relations
		Occupations of households
		Livestock
Spatial surface	Neighbourhood and distance surface	Neighbourhood and distance analysis from factors associated with outbreaks. [b]

a -: surface based on interpolation from weather recording station data using the kriging technique in ArcView® SpatialAnalyst .

[b]-: using neighbourhood analysis and distance mapping using ArcView® Spatial Analyst

Figure 18. Raster maps for elevation and rainfall in the study area



Statistical analysis

Assessment of spatial clustering

Cuzick and Edwards (1990) proposed a case-control spatial clustering test to take covariates into account by selecting controls to mimic the spatial distribution of the unaffected population at risk. In the current study, Cuzick and Edwards' test was used to test for the presence of clustering or spatial autocorrelation of the locations of FMD outbreaks and non-outbreaks by assigning them to cases and controls, respectively. Case locations were assigned the co-ordinates of the centre of a grid cell in which an outbreak occurred. An equal number of control locations was randomly selected from the remaining cells using the same method for assignment of co-ordinates.

Logistic regression analysis

Two types of dependent variable were used in the logistic regression analyses. One was the presence or absence of at least one FMD outbreak in a grid cell, coded as 1 or 0. The resulting models are called *single cell* models. The other type of dependent variable was the proportion of outbreaks in a 3*3 square grid cell area centred on the current cell (i.e. first order neighbourhood). These models are named *local region*

models and are generated in Minitab version 12 using an event/trials syntax. The proportion of outbreaks in neighbouring cells was considered as one potentially useful approach for controlling spatial autocorrelation. It has been described previously by Besag (1972) and used to estimate an autologistic model. The neighbourhood analysis function in ArcView® Spatial Analyst was used to derive as a numerator the number of adjacent cells with FMD outbreaks around each grid cell and also for the denominator the number of actual neighbouring cells for each grid cell within a 3 x 3 cell area (the latter was relevant for cells on the border of the mapped area).

The analysis was conducted in two steps. First a univariate analysis was performed to select a subset of variables which statistically significantly associated with the risk of a cell having at least one outbreak. Using the variables which came out significant in this step, a multivariate analysis was conducted to identify the best model describing the outcome variable.

Four different types of logistic regression models were constructed, and in all models specific consideration was given to potential effects of spatial autocorrelation on the regression coefficient estimates and their standard error. A stepwise variable selection process was used to identify the models *single cell (model 1)* and *local region (model 1)* including the set of variables providing the best model fit to the data. The level of statistical significance of the variables was assessed using the likelihood ratio test. At each step, any variables with a p value > 0.1 for the likelihood ratio test were excluded from the model. All variables with p values <=0.1 as determined by the univariate analysis were offered to the model selection process, except for 'total number of outbreaks in neighbouring cells', which was not offered to the *local region* model.

The Hosmer-Lemeshow statistic $\hat{C}$ (Hosmer, 1989) was used to assess goodness-of-fit for each model. It provides an easily interpretable single value for assessment, with large values of the test statistic and associated small p-values indicating poor fit. In addition, receiver operating characteristic (ROC) curves were generated to allow a graphical comparison of goodness-of-fit between different models, and to generate a tool for choosing cut-off values for decision criteria.

Assessment of autocorrelation in the residuals

The Moran's index (spatial autocorrelation coefficient) tests the null hypothesis of randomisation among residuals from the regression models as a single global statistic.

A test for autocorrelation of the model residuals will tell us whether a given model adequately takes into account spatial autocorrelation and thereby is likely to produce inefficient and unbiased estimates of regression coefficients (Cliff and Ord, 1981). High values of Moran's *I* are indicative of spatial autocorrelation. Moran scatterplots which show Moran's *I* statistic as a slope were produced to examine the degree of linear association of spatial association. This representation allows inspection of the underlying data values contributing to the global statistic, and therefore facilitates identification of outliers. It is also possible to show the result as a feature attribute in a map so that local autocorrelation can be examined for individual map locations. Using this graphical presentation, four types of spatial association can be distinguished in the set of paired values (Anselin, 1995). Associations between similar values, *high* surrounded by *high* and *small* surrounded by *small*, are interpreted as 'positive' spatial association while the other two forms representing dissimilar values (e.g. *high* surrounded by *small* and vice versa) are indicative of 'negative' spatial association.

In addition, variogram analysis was used for assessing the presence of spatial autocorrelation in the regression model residuals. The variogram is a graphical representation of the variation between sampling points separated by a given distance or lag (Isaaks and Srivastava, 1989). A continuous process without spatial autocorrelation will result in a horizontal line. Variograms which do not reach an upper bound suggest non-stationarity (Pfeiffer, 1996). In the current analysis estimates were based on the assumption that the data was isotropic. The lag distance was selected as 10km after inspection of the distribution of value pairings was inspected in order to have sufficient observations in the different lag categories. The regular and the robust version of the semivariogram were generated as suggested by Cressie (1993).

Classification tree analysis

A classification tree analysis was conducted using the occurrence of at least one FMD outbreak in a grid cell as a binary outcome variable. To generate the classification trees, all significant variables from univariate analysis were introduced into the analysis, with the exception of the variable 'total number of outbreaks in neighbouring cells' since it was highly correlated with the outcome.

The classification trees were generated using CART as the growing method and the GINI index as the impurity measure. Stopping rules for all trees were set at 10 for

maximum tree depth, 50 for the minimum number of cases in parent nodes and 25 for child nodes, and the minimum change in impurity was set at 0.0001. Ten-fold cross-validation was used to test the accuracy and stability of the trees. The sensitivity of the model to variation in misclassification cost of the ratio between false negative and false positive classifications ranging from 1 to 8 was explored using the two methods, variable misclassification costs via GINI and choice of prior, suggested by Breiman, *et* at. (1984). Prior probabilities were varied to take into account potential mis-representation of the reporting data with respect to true FMD risk in the study area.

Risk maps

The probability of a FMD outbreak occurring at each location (a cell on the grid) was estimated using the generated logistic regression and classification tree model equations and decision rules and then converted into maps of the probability of FMD outbreaks in Thailand.

The logistic regression model generates a value between 0 and 1 for each record in the data set or each grid cell, and this value represents the probability of occurrence of FMD outbreak at that cell given the predictor variable value pattern.

The classification trees were transformed to provide a probability instead of a classification by using the actual distribution of the grid cells in each leaf of the trees as the probability of occurrence of a FMD outbreak instead of assigning the leaf entirely to the most frequently occurring category (Long, *et* al., 1993).

Software

The spatial data was manipulated using ArcView® for Windows 3.1, its extensions Spatial Analyst and 3D Analyst (all three from Environmental Systems Research Institute, Redlands, California, U.S.A.) and Microsoft Access 97 (Microsoft Corporation, Redmond, WA, U.S.A.). Spatial clustering was assessed using the software Stat! (BioMedware, Ann Arbor, Michigan, U.S.A.). Statistical analyses were conducted in Minitab 12® (Minitab, State College, PA, U.S.A.) and the ROC curves were generated in NCSS 97® (NCSS Statistical Software, Kaysville, Utah, U.S.A.). The classification tree analysis was performed using the software AnswerTree version 1.0 (SPSS, Chicago, Illinois, USA). The analysis of regression residuals using the Moran scatterplot was performed with SpaceStat 1.8 (Regional Research Institute, West Virginia University, Morgantown, WV, U.S.A.) and the SpaceStat Extension for

ArcView (Regional Research Institute, West Virginia University, Morgantown, WV, USA). Variogram plots for the residuals were done using PROC VARIOGRAM in SAS for Windows Version 7.0 (SAS Institute, Cary, NC, U.S.A.).

Results

Univariate analysis

All variables were subjected to a univariate analysis using Student's t, Mann-Whitney U, or χ^2 tests to detect differences between outbreak and non-outbreak areas. Statistically significant differences at p <=0.1 were found for 28 out of the 67 variables tested (Table 22).

Based on the results of the Cuzick and Edwards' test, the distribution of FMD outbreaks showed significant spatial clustering of outbreak cells relative to the spatial distribution of non-outbreak cells ($p < 0.001$). These results suggest the presence of spatial autocorrelation in the dependent variable used in the following models.

Table 22. Variables with statistically significant difference between outbreak and non-outbreak areas.

Geographical data set	Variables	Code	Mean	*p* value*
Environmental	30-year average monthly rainfall in rainy season (mm.)	AvRainy	8.02	0.04
	30-year average monthly rainfall in cool season (mm.)	AvCool	4.28	0.02
Agricultural	Land use	Land		0.002
Socio-economic	*Farmer education*			
	Illiteracy level of sub-district	Illiterate	0.04	0.08
	Public relations [c]			
	Proportion of sub-districts with public reading places [c]	PubRead	0.94	0.07
	Proportion of sub-districts with public libraries	PubLib	0.95	< 0.001
	Proportion of sub-districts with public news broadcasts	PubNews	0.83	0.08
	Occupations of households [d]			
	Proportion of households engaged in agriculture	HHAgri	0.84	0.008
	Proportion of households for which growing rice is the main occupation	HHRice	0.20	0.001
	Proportion of households which cultivate rice once each year [c]	RiceOnce	0.65	<0.001
	Proportion of households which cultivate rice twice each year [c]	RiceTw	0.02	0.07

Geographical data set	Variables	Code	Mean	*p* value*
	Proportion of households for which growing field crops is the main occupation	HHCrop	0.08	0.03
	Annual income of households for which growing field crops is the main occupation (baht)	CropInc	46,712.05	0.01
	Annual income of households for which raising animals is the main occupation (baht)	AnimalInc	30,448.89	0.002
	Annual income of households with more than one occupation (baht) [c]	2OcInc	228,587	0.10
	Livestock			
	Number of livestock markets	LstkMrk	0.16	0.08
	Proportion of sub-districts with animal remedy depot, veterinary volunteers or livestock development volunteers	AnimalMed	0.82	<0.001
	No of public shallow or dug wells	PubShal	28.51	0.06
	Cattle (Beef, dairy, and buffalo) density per km^2	CtlDens	26.32	0.10
	Buffalo density per km^2	BufDens	9.16	0.09
	Pig density per km^2	PigDens	10.01	0.003
Spatial theme	Total number of outbreaks in neighbouring cell	SumOtb	2.52	<0.001
	Distance from livestock market (km)	DstLvMrk	49.84	<0.001
	Distance from water body such as lake, reservoir (km)	DstWtr	28.72	0.10
	Distance from transport type 1 (km)	DstTrans1	8.38	0.05
	Distance from transport type 1 and 2 (km) [e]	DstTrans12	9.38	0.02
	Distance from border to Myanmar (km)	DstMynm	264.15	0.08
	Distance from border to Laos (km)	DstLaos	168.8	<0.001

* -: *p* value from univariate comparisons between outbreak and non-outbreak areas.

c -: number of sub-districts with "yes" divided by total number of sub-districts in that district.

d -: number of households with the factor of interest divided by the total number of households in that district.

e-: variables that were not included in multiple logistic model process because they were considered to be too highly correlated with other predictors.

Logistic regression analysis

The final single cell logistic regression model called *single cell (model 1)* contained four variables: 'proportion of sub-districts with public news broadcast facilities', 'proportion of households for which growing rice is the main occupation', 'proportion of sub-districts with animal remedy depot, veterinary volunteers or livestock development volunteers', and 'total number of outbreaks in neighbouring cells'. To assess the effect of spatial dependence on the regression coefficients and model fit, additional analyses were performed in that the variables included in *single cell (model 1)* with the exception of 'total number of outbreak in neighbouring cells' were forced into *single cell (model 2)* and *local region (model 2).* The regression coefficients

for these models expressed as odds ratios are presented in Table 23. Comparison of the regression coefficient estimates shows that they are broadly similar between the models. In fact, the 95% confidence limits would overlap for all variables. The final *local region (model 1)* was derived from a separate stepwise logistic regression analysis of the dataset and six variables came into the model as shown in Table 24. From the parameters included in the *single cell (model 1)*, only the variable 'AnimalMed' was still present. Tests for significance of the models, Hosmer and Lemeshow's goodness of fit, and Moran's *I* test for spatial autocorrelation of residuals are shown in Table 25. It appears that the *single cell (model 2)* produces good fit to the data as does *local region (model 2)*, but the latter is adversely affected by autocorrelation. On the basis of this table, one would have to consider both *single cell* models being the best models. Variogram plots for the residuals of all models are shown in Figure 19.

In the *single cell (model 1)*, a factor strongly associated with the occurrence of FMD outbreaks was 'total number of outbreaks in neighbouring cell'. This can be seen as a factor expressing spatial autocorrelation. It is interesting to note that its exclusion in *single cell (model 2)* does not have a major effect on the autocorrelation in the regression residuals. The variables included in *single cell (model 1)* suggest that while the districts with a higher number of sub-districts receiving public news broadcasts were less likely to have an outbreak of FMD in the area, districts containing more households growing rice as the main occupation as well as those with more sub-districts having animal medical storehouses or livestock development volunteers were 3.5 and 5 times more likely to have FMD outbreaks, respectively.

In the context of spatial dependence, the result of the examination of standardised Pearson residuals from the different regression models using Moran's *I* statistic yielded a value of -0.035 for the *single cell (model 1)* and 0.50 for the *local region (model 1)*. It would appear that both *single cell* models are reasonably adequate while the other two *local region* models seem to have a strong spatial dependence in the residuals. This association was clearly seen from mapping the Moran scatterplot quadrant values as shown in Figure 20. The positive spatial associations are shown using a red colour shading and negative associations using a green shading. The map for *single cell (model 1)* shows a mix of the positive and negative associations, whereas the one for *local region (model 1)* is clearly dominated by positive autocorrelation. The variogram plots are flat for *single cell* models except for the steep slope between the first and

second lag which is largely the result of sparse data. In contrast, the plots for the *local region* models *(model 1 and 2)* do have a more consistent slope up until lag numbers 4 and 9, respectively. This confirms the results of Moran's *I* in indicating the presence of autocorrelation in these two models. Therefore, the regression coefficients for the *single cell* models are more likely to be unbiased and efficient than the ones for the *local region* models. This would have a significant impact on the accuracy of the probability predictions (Figure 21).

Table 23. Odds ratios with 95% confidence limits (in brackets) of variables in logistic regression models *single cell (models 1, 2)* and *local region (model 2)*

Variables	*Single cell (model 1)*	*Single cell (model 2)*	*Local region (model 2)*
PubNews	0.10 (0.03-0.41)	0.13 (0.04-0.48)	0.58 (0.36 - 0.94)
HHRice	3.47 (1.26-9.51)	4.59 (1.82-11.62)	2.27 (1.61-3.19)
AnimalMed	5.05 (1.37-18.62)	4.77 (1.53-14.88)	3.61 (2.47-5.29)
SumOtb	1.73 (1.51-1.99)	-	-

Table 24: Odds ratios with 95% confidence limits (in brackets) for variables included in final logistic regression model *local region (model 2)* using proportion of cells with outbreaks in the local region (9 grid cells) as the response variable

Variables	*Local region (model 1)*
DstMrk (km)	0.99 (0.99-0.99)
AvRainy	1.14 (1.10-1.17)
AvSummer [g]	0.61 (0.50-0.74)
Miscellaneous landuse [h]	1.62 (1.27-2.07)
Hhanimal [i]	4.33 (2.08-8.99)
AnimalMed	1.60 (1.09-2.36)

g :- 30-year average monthly rainfall in summer season

h :- compared with agricultural land and forest land as the reference group

i :- proportion of sub-district for which raising animals is the main occupation

Table 25: Summary table of deviance (G), Hosmer and Lemeshow's goodness of fit (ç), and spatial autocorrelation of residuals (Moran's I) for the four logistic models single cell (models 1, 2) and local region (models 1, 2)

Statistical parameter	*Single cell (model 1)*	*Single cell (model 2)*	*Local region (model 1)*	*Local region (model 2)*
G	97.32	25.35	257.51	81.93
degrees of freedom	4	3	6	3
P value	0.000	0.000	0.000	0.000
Hosmer-Lemeshow ĉ	14.05	8.50	2.84	21.17
degrees of freedom	8	8	8	8
P value	0.08	0.39	0.94	0.007
Moran's I	-0.04	0.05	0.50	0.37

Figure 19: Omni-directional variograms derived from the residuals of different logistic models (using a lag distance of 10km)

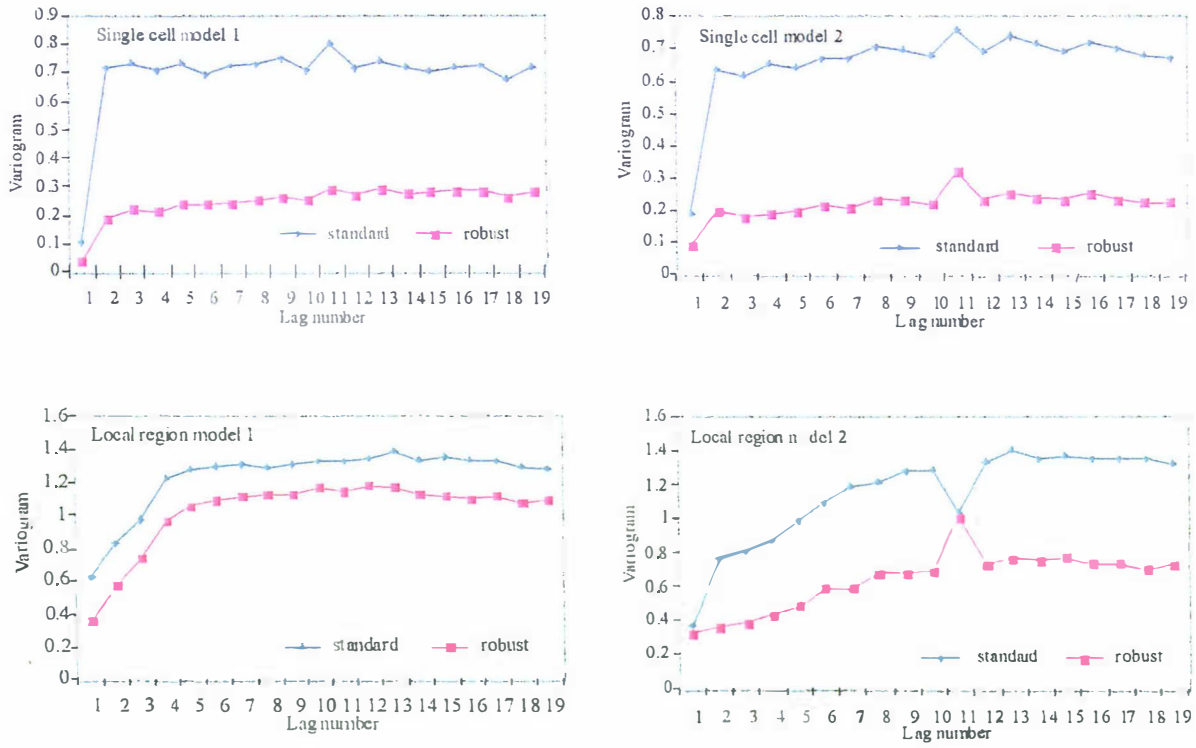


Figure 20. Maps showing the quadrants of the Moran scatterplot generated from residuals of different models.

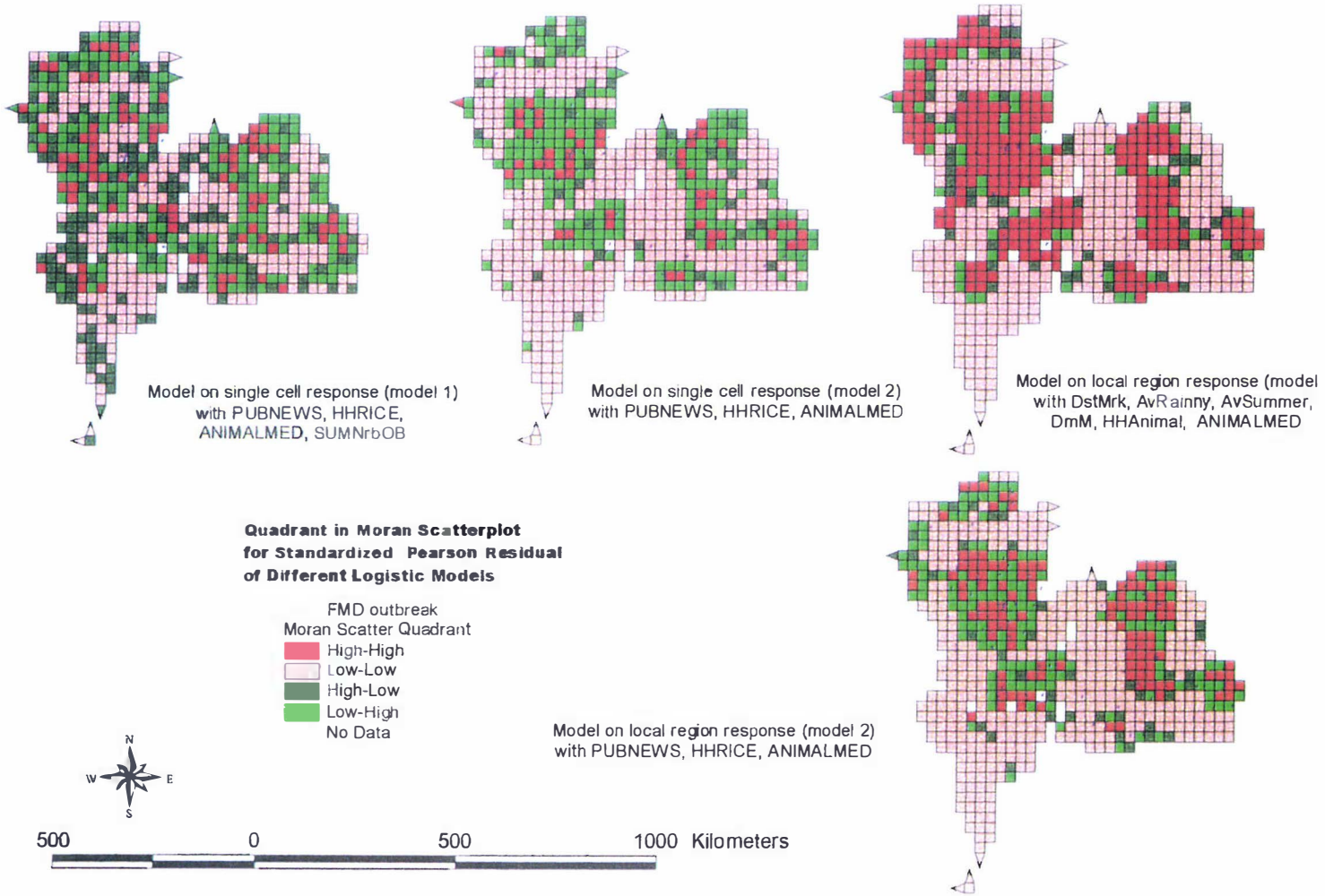


Figure 21. Maps of the probability of FMD outbreak occurrence based on different logistic models.

Classification tree analysis

The classification tree analysis where equal weighting was given to false positive and false negative classifications was unable to generate a tree given the set criteria of node impurity. After increasing the relative weighting of the costs resulting from false negatives, subsequent trees produced for different weightings contained the same variables as well as had the same structure. As intended, adjustments to the misclassification costs directly affected the misclassification risk, and sensitivity as well as specificity of the model and the outcome changed to a different FMD outcome category in some terminal nodes. However, the risks of misclassification become increasingly higher with increasing cost adjustments. Therefore, the adjusted prior method was used in combination with cost adjustment in order to reduce the risk of misclassification. This seemed justified to take account of the fact that the data set was dominated by observations (i.e. cells) which did not have FMD outbreaks.

Starting from the classification tree with equally weighted mis-classification cost, we explored the effect of adjusting prior probabilities in addition to mis-classification cost in the subsequent trees. Adjusting prior probabilities also directly affected the misclassification risk and therefore sensitivity and specificity of resulting tree models. Keeping the same tree structure, the model showed increasing sensitivity at the expense of specificity as the misclassification cost of false negatives relative to false positives increased. The misclassification risk for the subsequent trees was increased, but not as high as without adjusted prior. Tree size and the variables that came into the model were changed following cost adjustments as set out in Table 26. The ranking in this table is based on the purity gains resulting from introducing a particular variable to the tree. Summaries of all tree models are listed in Table 27.

The impurity function used to select splits relates closely to the relative risk where low impurity implies a high relative risk (Zhang, *et* al., 1996). Resubstituted relative risks (RRR) were estimated for each factor in the tree at each split. The first three variables that were included in the default tree and all the following trees after prior adjustment for cost were 'distance from the border of Laos' (DstLaos), 'proportion of sub-districts with public news broadcasts', and 'proportion of households for which growing field crops is the main occupation'. The RRR for the first split of the DstLaos factor at less than or equal 267.64 km was (87/483)/(7/178) = 4.58 compared to the group where

distance was greater than 267.6 km. The RRR for the splits of other variables are listed in Table 28.

The models need to reflect a requirement of control schemes for a conservative approach in the prediction of possible risk areas so that none are missed. It was therefore decided to weight the sensitivity of the model as high as 90 %, leaving 10% for specificity. The tree model with the cost of false negatives at 5 times the cost of false positives was chosen as the best model. The sensitivity and specificity of this tree were 89% and 45%, respectively. The model gave an overall misclassification risk of 35% and a risk of 39% following cross-validation indicating good accuracy and stability of the tree. The best classification tree model incorporating cost-sensitive adjustments in which the cost of false negatives were 5 times that of false positives is shown in Figure 22 and correct classification percentages for root, intermediate, and terminal nodes are presented in Table 28.

Table 26. Ranking of variables for different levels of cost adjustment (1-1 indicates equal weighting for false negative and false positive, 2-1 indicates the cost of a false negative to be set at 2x that of a false positive)

Variable	*Tree 1_1*	*Tree 2_1*	*Tree 3_1*	*Tree 4_1*	*Tree 5_1*	*Tree 6_1*	*Tree 7_1*	*Tree 8_1*
Cost adjusted (FN : FP)	1:1	2:1	3:1	4:1	5:1	6:1	7:1	8:1
DstLaos	1	1	1	1	1	1	1	1
PubNews	2	2	2	3	3	3	3	3
HHCrop	3	3	3	2	2	2	2	2
BuffaloDens	-	-	-	-	2	2	2	-
DstLvMrk	-	-	-	3	3	3	-	-
DstTransl	-	-	-	-	3	3	3	-
AvRainny	-	-	-	-	-	-	3	3.5*
PubShallow	-	-	-	-	-	-	-	4
AgrOccu		4	4	-	-	-	-	-

* variable was used for two splits.

Table 27. Summary data for classification trees with cost-sensitive adjustments (1-1 indicates equal weighting for false negative and false positive, 2-1 indicates the cost of a false negative to be set at 2x that of a false positive)

Summary	*Tree 1_1*	*Tree 2_1*	*Tree 3_1*	*Tree 4_1*	*Tree 5_1*	*Tree 6_1*	*Tree 7_1*	*Tree 8_1*
Cost adjusted (FN : FP)	1:1	2:1	3:1	4:1	5:1	6:1	7:1	8:1
Priors adjusted (non-outbreak: outbreak)	0.86: 0.14	0.75: 0.25	0.67: 0.33	0.60: 0.40	0.55: 0.45	0.50: 0.50	0.46: 0.54	0.43: 0.57
Resulting tree								
Number of nodes	7	9	9	9	13	13	13	13
Number of levels in the tree	3	4	4	3	3	3	3	4
Number of terminal nodes	4	5	5	5	7	7	7	7
Misclassification rate								
Risk estimate	0.14	0.23	0.30	0.36	0.35	0.33	0.32	0.27
s.e. of risk	0.01	0.01	0.01	0.01	0.02	0.03	0.02	0.03
Cross-validation								
Risk estimate	0.14	0.27	0.36	0.45	0.39	0.51	0.48	0.41
SE of risk	0.00	0.01	0.02	0.02	0.03	0.03	0.03	0.03
Sensitivity (95% CI)	0	0.17 (0.09-0.25)	0.17 (0.09-0.25)	0.15 (0.08-0.22)	0.89 (0.83-0.96)	0.97 (0.93-1.00)	0.90 (0.84-0.96)	0.97 (0.93-1.00)
Specificity (95% CI)	1	0.96 (0.95-0.98)	0.96 (0.95-0.98)	0.97 (0.96-0.99)	0.45 (0.41-0.49)	0.38 (0.34-0.42)	0.42 (0.38-0.46)	0.42 (0.38-0.46)

s.e. = standard error

Figure 22. Classification tree model (*tree 5_1*) incorporating cost-sensitive adjustment (cost of false negatives is 5 times that of false positives). *Ellipse shaped nodes indicate prediction of no outbreak. Star nodes indicate prediction of an outbreak. Within each node, the actual observed numbers of grid cells with and without outbreaks are shown. Terminal nodes are distinguished by shading.*

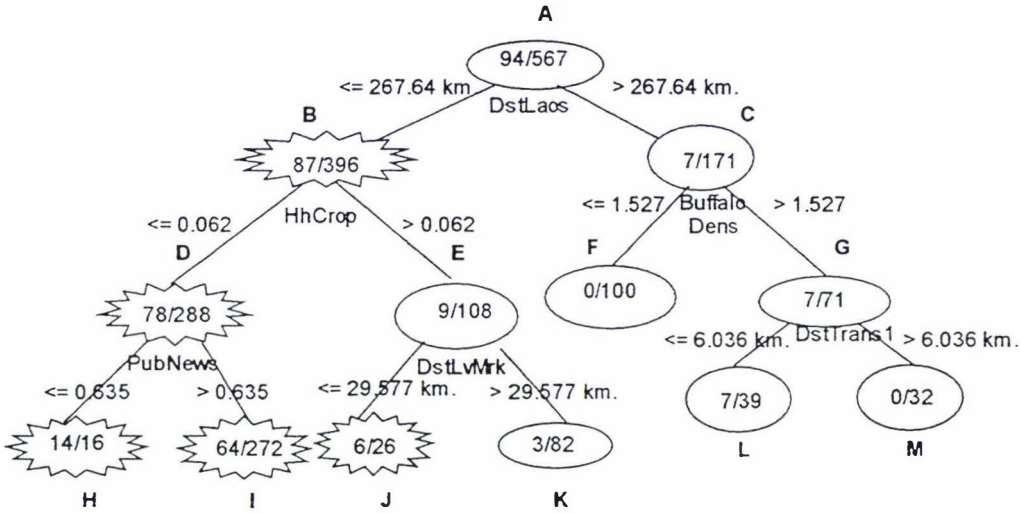


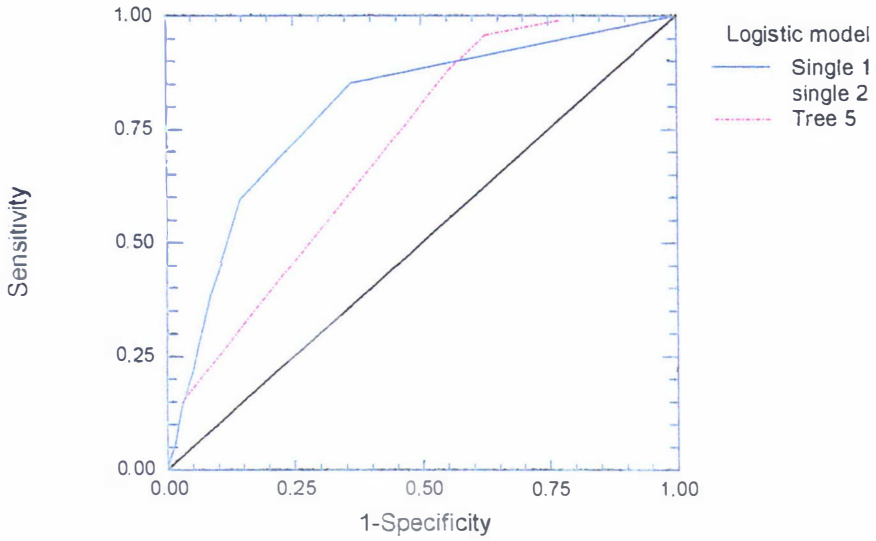
Table 28. Node statistics for preferred classification tree *tree 5_1*

Node	*Node type*	*Outcome category classified*	*% Correctly classify*	*RRR*
A	Root node	No outbreak	85.78	4.58
B	Intermediate node	Outbreak	18.01	2.77
C	Intermediate node	No outbreak	96.07	0.00
D	Intermediate node	Outbreak	21.31	2.45
E	Intermediate node	No outbreak	92.31	5.31
G	Intermediate node	No outbreak	91.03	-
H	Terminal node	Outbreak	46.67	-
I	Terminal node	Outbreak	19.05	-
J	Terminal node	Outbreak	18.75	-
K	Terminal node	No outbreak	96.47	-
F	Terminal node	No outbreak	100	-
L	Terminal node	No outbreak	84.78	-
M	Terminal node	No outbreak	100	-

ROC analysis

ROC curves for two logistic regression models and the preferred classification tree model *tree 5_1* were produced using the estimated sensitivity and 1-specificity values (Figure 23). The area under the ROC curve (AUC) for the *single cell (model 1)* was 0.79, 0.64 for the *single cell (model 2)* and 0.70 for the classification tree model *tree 5_1*. Given the lower AUC value, the logistic regression *single cell (model 2)* was not used for risk mapping purposes. It becomes clear from inspection of the curves that to achieve acceptable sensitivity levels such as 90% the percent false positive predictions could easily rise to 40%.

Figure 23. ROC curves for logistic regression and classification tree models



Risk maps

The probabilities (p) of FMD outbreaks produced from the *single cell (model 1)* and the classification tree (*tree 5_1*) were used to generate 'risk-maps' demonstrating how the risk of disease varies in space (Figure 24 and Figure 25). Separate maps based on the upper and lower 95% confidence limits were produced to indicate the accuracy of the predictions. The areas in those maps depicted in red were predicted to experience outbreaks of FMD when cut-off values for the models were set at 0.5 for the logistic model and 0.18 for the classification tree. The latter cut-off was different as a reflection of the prior cost adjustment.

Figure 24. Three-dimensional risk map showing predicted risk of FMD outbreak occurrence based on logistic regression *single cell (model 1)*

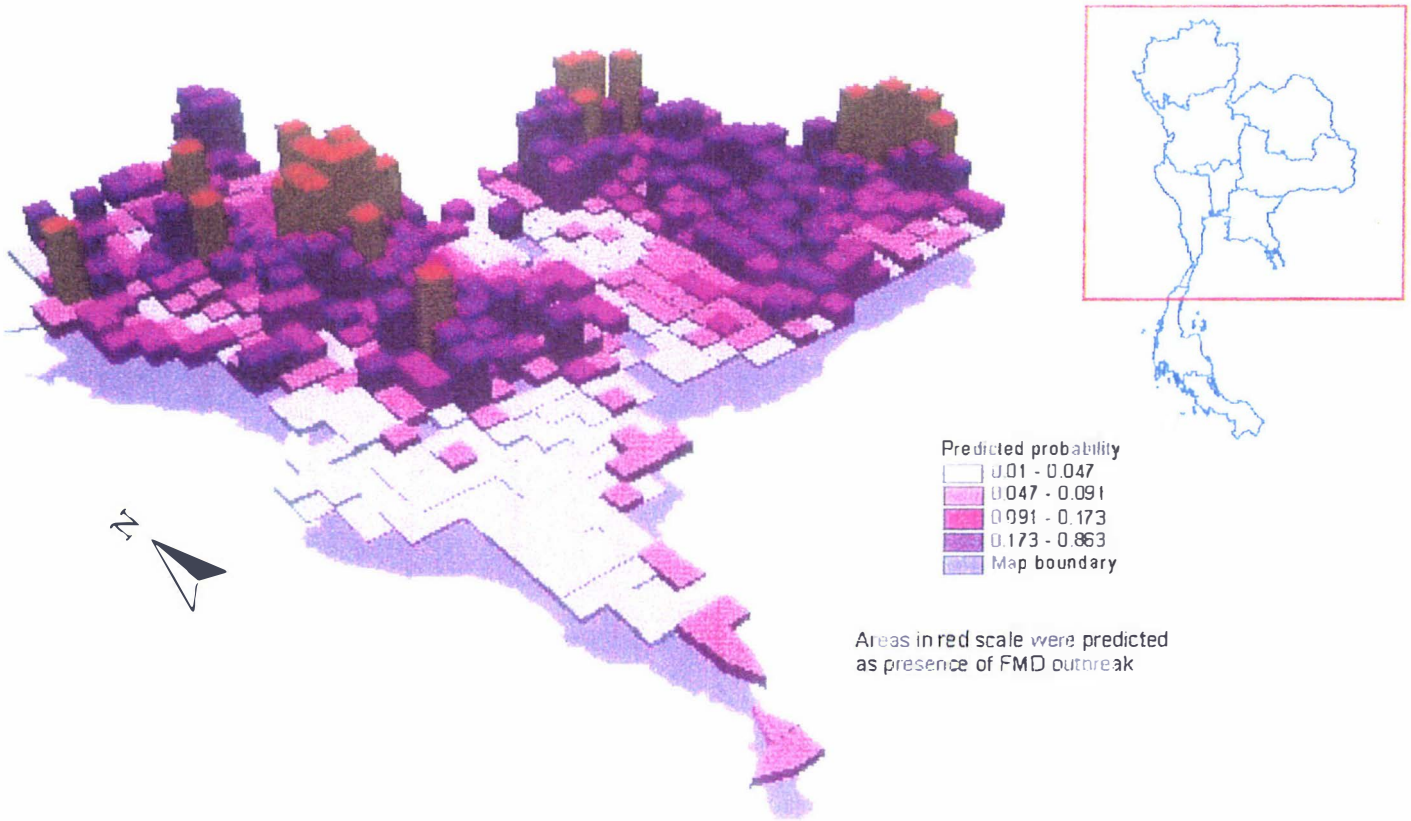
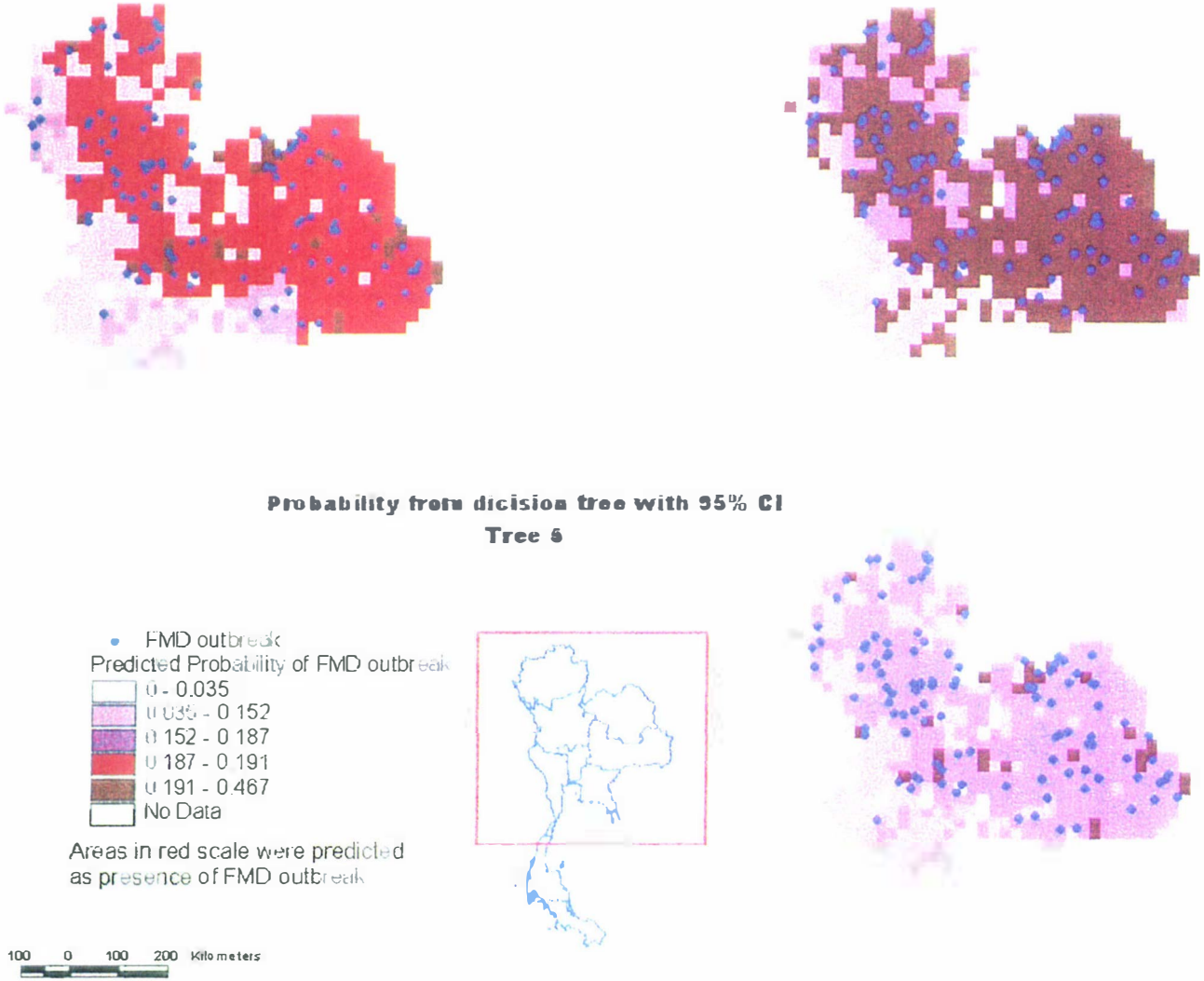


Figure 25. Map of predicted probability of FMD outbreak occurrence based on the preferred classification tree-based model *tree 5_1* (incl. 95% CI maps)



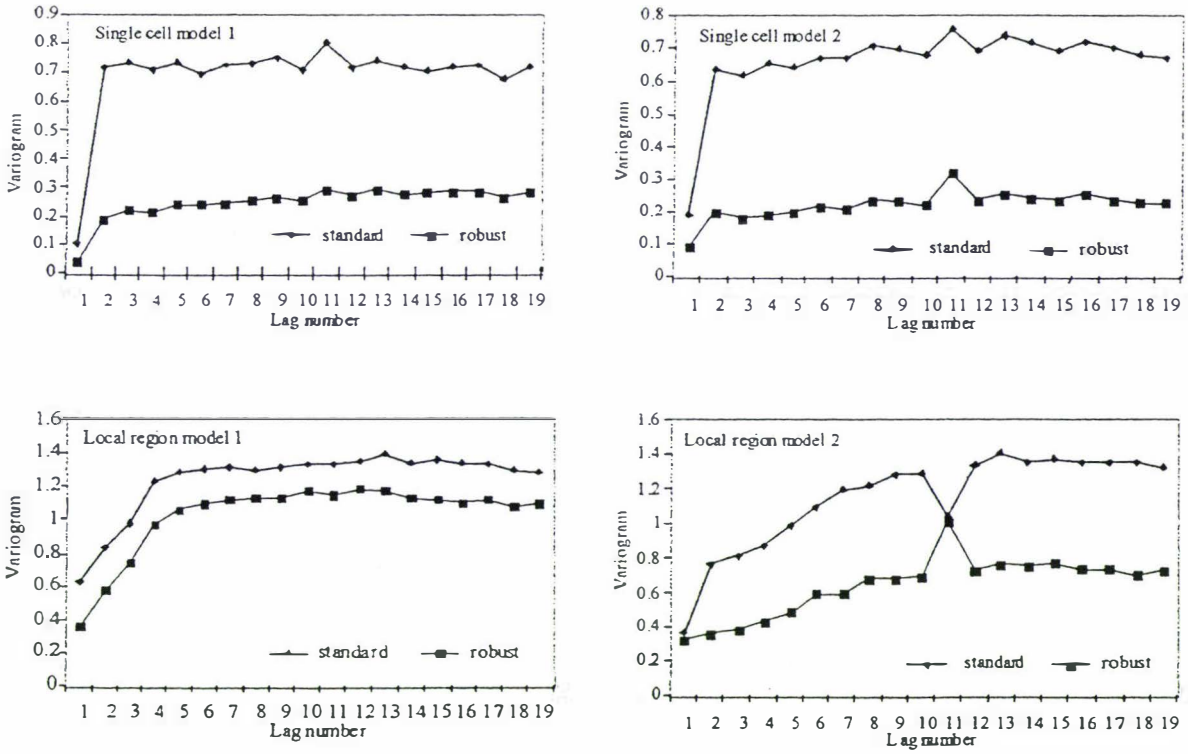
Discussion

It is important to be able to predict the distribution of disease in order that preventive programmes can be targeted at the most important areas and control operations can be designed and executed to achieve optimal cost-effectiveness. Methods for prediction of disease outbreaks on a spatial scale or geographical patterns in disease distribution have been used in the study of theileriosis (Duchateau, *et* al., 1997; Pfeiffer, *et* al., 1997), bovine anaplasmosis and babesiosis (Perez, *et* al., 1994), distribution of insect vectors of disease (Williams, *et* al., 1992), and some other parasitic zoonoses (Mott, *et* al., 1995). However, amongst these modelling approaches only Pfeiffer *et* al's models took spatial autocorrelation into account.

In the study reported here, the presence of spatial autocorrelation in the outcome variable FMD occurrence was confirmed with the Cuzick and Edwards' test. During the regression modelling spatial dependence was taken account of by introducing variables into the modelling process that represented the relationship with respect to FMD risk between any grid cell and its immediate neighbours. The variable was either used as predictor or a dependent variable. Comparisons of the Moran's *I* statistic and the variogram graphs derived from the residuals of the different logistic regression models suggest that with this particular data controlling for spatial autocorrelation through inclusion of incidence in the local region as defined by a 3*3 grid cell square does not reduce the effect of spatial autocorrelation in contrast to the analysis reported by Pfeiffer, *et* al. (1997) for theileriosis. On the contrary, the use of *local region* as a response variable did result in an increase of the spatial dependence in the residuals compared with the standard model as shown in Table 25: Summary table of deviance (G), Hosmer and Lemeshow's goodness of fit (ç), and spatial autocorrelation of residuals (Moran's I) for the four logistic models single cell (models 1, 2) and local region (models 1, 2)

Statistical parameter	*Single cell (model 1)*	*Single cell (model 2)*	*Local region (model 1)*	*Local region (model 2)*
G	97.32	25.35	257.51	81.93
degrees of freedom	4	3	6	3
P value	0.000	0.000	0.000	0.000
Hosmer-Lemeshow $\hat{C}$	14.05	8.50	2.84	21.17
degrees of freedom	8	8	8	8
P value	0.08	0.39	0.94	0.007
Moran's I	-0.04	0.05	0.50	0.37

Figure 19: Omni-directional variograms derived from the residuals of different logistic models (using a lag distance of 10km)



, when mapping the quadrants of the Moran scatterplot of the residuals (Figure 20) as well as in the variogram plots for the residuals in Figure 19.

Although the *single cell (model 1)* did not to have good overall fit as indicated by a low *p* value for the Hosmer-Lemeshow statistic $\hat{C}$, a comparison of the areas under the ROC curve suggests that this model produced the best prediction of the outcome variable. The assessment of autocorrelation in the residuals indicated that it did not appear to be adversely affected by the spatial dependence or spatial autocorrelation in the dependent

variable. The predicted probabilities from this model were used to generate a risk –map of predicted probabilities of FMD outbreak occurrence.

The most important risk factors associated with the occurrence of FMD outbreaks over the study area based on the spatial logistic regression model *single cell (model 1)* were 'proportion of sub-districts with public news broadcasts', 'proportion of households for which growing rice is the main occupation', 'proportion of sub-districts with animal remedy depot, veterinary volunteers or livestock development volunteers', and 'total number of outbreaks in neighbouring cells'. The last factor has no practical value when using the model to devise control strategies. But its inclusion in the model can be seen as reflecting clustering or contagiousness of disease. It should be noted that this variable expressing the risk in the neighbourhood as a count was better able to control for the spatial dependence in the residuals of the model than inclusion of a variable expressing the local risk of outbreaks occurring in the immediate neighbourhood as a proportion. The latter variable had been used as dependent variable in *local region model 2.* As Bailey and Gatrell (1995) suggested, if a variate interaction model helps us to understand the behaviour of the process and provides us with insights as to possible explanations for this behaviour then the modelling will have achieved its objective. In this particular case the findings lead us to suspect that some of the other predictor variables must have explained some of the spatial autocorrelation.

A factor that was sparing for outbreaks in the study area of Thailand was having a high proportion of sub-districts with public news broadcasts. Public news broadcasts have a strong public relations content including the activities of "*The National Foot-and-Mouth Disease Control and Eradication Project*". Livestock farmers, livestock vendors, the agribusiness industry, officials and governmental agencies, as well as those who are responsible for implementing the project are kept informed of the necessity to control and eradicate the disease through mass media communication such as radio, television, printed material and other audio-visual aids. This study indicates that those public relation activities may have had a positive impact on the control of the disease. It is also possible though that it may have been confounded with closeness to the borders of Myanmar and Laos which is where FMD infection is suspected to have been introduced from.

Livestock development in Thailand is often promoted with regard to its potential for increasing the country's domestic consumption and export trade. However, in rural

areas cattle and buffalo still have an important role as draught animals for rice cultivation as well as for meat and milk production. It is not surprising that the areas that have a larger proportion of households for which growing rice is the main occupation tended to have a relatively high occurrence of FMD outbreaks. The study indicates that these farmers are a risk group that the government should specifically target in the control scheme. While the use of draught animals may explain the association between FMD and rice growing, it is also possible that the association was purely a reflection of poor data quality. The disease data used in this study was based on routine monthly disease reports. There is a definite possibility that disease occurrence was underreported. In particular, some of the areas that were never reported as having FMD outbreaks may in fact have had outbreaks. Because of the importance of draught animals, farmers in the rice growing areas look after their cattle and buffaloes well. When these animals become sick, they usually report and ask for help from veterinarians, the head of the village or livestock development volunteers. This will result in a reporting bias, as in other areas of the country livestock is less relevant to farmers, who may, therefore, be less likely to report cases of FMD.

Another risk factor associated with the occurrence of FMD in Thailand is 'proportion of sub-districts with animal medical storehouses or livestock development volunteers'. This variable was included in the *single cell* and the *local region* models and had particularly high odd ratios in the *single cell* models. As discussed above, this factor could also be an indicator of reporting bias. The better the animal health infrastructure, the more likely it is that any outbreaks will be reported.

A number of other variables shown in the results of the univariate analysis may be "biologically relevant" risk factors for the occurrence of FMD although not statistically significant. For example, the number of public wells in an outbreak area is significantly higher than in an area without disease outbreak. Distance to the next livestock market, to the next main road, and from the border to Laos all were significantly shorter for areas with outbreaks when compared with those not having any reported outbreaks. Although these factors were not included in the logistic regression models, some of them became part of the classification tree models.

It was not possible to compare the performance of variables introduced into the logistic regression model with their performance when introduced into the classification tree model in the present study as different initial variables were included in the models

generated by these methods. The variable 'total number of outbreaks in the neighbouring cell' which was included in the logistic regression model and which strongly affected the coefficients of the other variables in the model was not selected in the classification tree models.

The objective of this study was not to compare the performance of these methods. However, there are some interesting points that have been discussed in many studies between these two methods.

Performance of logistic regression and classification tree models with respect to their ability to correctly classify cases has been compared in a number of papers. Long *et* al. (1993) studied the performance of both methods in the medical domain, while the methods were performed fairly similar the logistic regression method was slightly better than the classification tree. A similar conclusion was drawn by Stärk (1998) who offered a number of suggestions on the way in which the results of the logistic regression model should be interpreted. Hadorn *et* al. (1992) looked at the performance with respect to sensitivity and specificity using ROC curves derived from logistic regression, a series of CART models generated by varying the misclassification cost specifications and some other method. These authors found the performance of the logistic model to be marginally superior to that of other models. All of the above authors agree that both the classification tree and the logistic regression methods are very dependent on the quality of the underlying data.

Classification tree analysis does not provide a specific method for modelling the spatial dependence of data. As its methodology is very robust and does not require the assumption that contributing observations are independent, the derived classification trees will not be adversely affected by the presence of spatial autocorrelation in the data. Classification tree analysis offers an excellent method for efficiently developing production rules for geographically-orientated decision support systems. Classification trees provide detailed insight into the data structure as well as between-variable relationships. Compared with logistic regression it excels in detecting local complex data structure while logistic excels in detecting linear and global structure. Users can incorporate differential weightings to optimise tree structure with respect to differential mis-classification costs. The classification tree is easy both to explain and to interpret because the decision rules developed involve binary judgements which are intuitively easier to understand than equations based on a sequence of odds ratio estimates

generated by logistic regression modelling (Walker and Moore, 1988). For this reason classification trees are becoming increasingly popular, particularly in the medical domain. For exploring large datasets such as in data mining classification trees have become a standard method.

With our data set, classifiers that minimize the number of misclassification errors are inadequate for problems with variable misclassification costs. Using the GINI selection criterion in CART variables for outcome classification could not be found without weighting on misclassification cost. Trees taking account of differential cost-sensitivity in combination with altered priors allowed the ability to provide reasonable predictions of the occurrence of FMD outbreaks in selected terminal nodes. The classification trees showed an increase in misclassification risk and in sensitivity but a decreased specificity as the misclassification cost of false-negatives increased. We therefore used estimates of the sensitivity and specificity for correctly identifying outcome categories in accordance with the requirements of the FMD control scheme to compare the performance of the different classification trees based on different misclassification costs. Estimates of sensitivity and specificity have been used in a number of studies as the basis of a comparison of prognostic models (Stärk, 1998; Long, *et* al., 1993). The classification tree can be converted into a set of decision rules (Figure 22) for extrapolation beyond the data domain, keeping in mind the efficacy of prediction as quantified by the sensitivity and specificity of the tree model.

The receiver-operating characteristic (ROC) curve is a useful summary statistic of predictive performance of models generating probability-based predictions as outcome variables. It plots the various combinations of true positive and false positive risks as the test threshold that defines a positive test is being varied. The ROC curve provides a method for decision makers to consciously vary cut-off values for binary decision criteria taking into account the required level and weighting of sensitivity and specificity. This characteristic of the ROC curve can also be used to choose appropriate cut-off points for logistic regression model predictions.

Visual assessment of the risk maps shows that the logistic models predicted the presence of FMD outbreaks in only a small number of areas, whereas the classification tree indicated large areas as being at different levels of risk of FMD outbreaks. The cut-off point used in logistic regression is 0.5. With our data set this level resulted in a sensitivity and specificity of 12 % and 98%, respectively. The level of sensitivity and

specificity using the same cut-off point for the classification tree predictions resulted in 63% and 83%, respectively.

The 'risk map' of model predictions provides a visually effective method for guiding informed disease control decision making that takes into account 'likely' local outbreak risks. The decision as to which of various control strategies to implement can then be based on the level of disease risk in individual areas. The map should be considered n combination with the ROC curve which very effectively summarises the probability of missing potential outbreaks or unnecessarily applying the control measures to individual areas depending on which cut-off point has been selected.

Conclusion

This study has shown the potential for generating useful information through combination of data from a routine disease surveillance system and geographical data bases. While the different techniques discussed can be used to generate useful predictions to be incorporated in the planning of disease control efforts, the quality of the predictions will depend on the data used to generate them. In this particular situation in Thailand, before these models can be applied the accuracy of the disease reporting data should be assessed. In addition, the level of aggregation at which the data is being recorded should be determined after considering the intended use of the information. Too high a level of aggregation relative to the scale at which the underlying disease process is operating will make it difficult to derive predictive models. Ideally, outbreaks of diseases such as FMD should be recorded with an accuracy at least at the village level. The district level is not adequate if analyses as presented in this study are intended.

As result of a comparison of the different models it was decided that with this dataset the predictions generated by the classification tree model *tree 5_1* were probably most useful. Considerable effort went into the comparison of different methods for controlling spatial autocorrelation, and it was concluded that the auto-logistic or *local region* modelling approach did not seem to of benefit with this data. It is possible that the data quality and potential confounding relationships between the variables was partly responsible for this unexpected result.

In conclusion, we are moving towards a situation where, through the availability of sophisticated data analysis tools in combination with advanced data management and presentations systems such geographic information systems, disease control decision makers will be better able to make informed decisions that more effectively utilise data which often is collected at great expense, but that is frequently not used effectively.

Reference List

Anselin, L., 1995. SpaceStat version 1.80; User's guide. Morgantown, Regional Research Institute West Virginia University.

Austin, G.E., Thomas, C.J., Houston, D.C., Thompson, D.A., 1996. Predicting the spatial distribution of buzzard *buteo buteo* nesting areas using a geographical information system and remote sensing. Journal of Applied Ecology, 33: 1541-1550.

Besag, J., 1972: Nearest-neighbor systems and the auto-logistic model for binary data. Journal of Royal Statistical Society, Series B, 36: 75-83.

Bian, L., West, E., 1997. GIS modeling of elk calving habitat in a prairie environment with statistics. Photogrammetric Engineering & Remote Sensing, 63: 161-167.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Wadsworth International Group, Belmont, California.

Cliff, A.D., Ord, J.K., 1981. Spatial processes: models & applications. Pion, London.

Cressie,N.A.C., 1993. Statistics for spatial data. John Wiley, New York.

Cuzick, J., Edwards, R., 1990. Spatial clustering for inhomogenous populations. Journal of the Royal Statistical Society, Series B, 52: 73-104.

David, W.H., Lemeshow, Jr.S., 1989. Applied logistic regression. Wiley, New York.

Duchateau, L., Kruska, R.L., Perry, B.D., 1997. Reducing a spatial database to its effective dimensionality for logistic-regression analysis of incidence of livestock disease. Preventive Veterinary Medicine, 32: 207-18.

Fotheringham, A.S., Rogerson, P., 1993. Spatial analysis and GIS. Taylor & Francis, London ; Washington, DC.

Hadorn, D.C., Draper, D., Rogers, W.H., Keeler, E.B., Brook, R.H., 1992. Cross-validation performance of mortality prediction models. Statistics in Medicine, 11: 475-89.

Isaaks, E.H., Srivastava, R.M., 1989: An introduction to applied geostatistics. Oxford University Press, New York, 561pp.

Long, W.J., Griffith, J.L., Selker, H.P., Dagostino, R.B., 1993. A comparison of logistic regression to decision-tree induction in a medical domain. Computers in Biomedical Research, 26: 74-97.

Mott, K.E., Nuttall, I., Desjeux, P., Cattand, P., 1995. New geographical approaches to control of some parasitic zoonoses. Bulletin of the World Health Organization, 73: 247-57.

Norman, H.S., Sischo, W.M., Pitcher, P., Nesselrodt, A., Day, R.L., 1996. Spatial and temporal epidemiology of pseudorabies virus infection. American Journal of Veterinary Research, 57: 1563-1568.

Perez, E., Herrero, M.V., Jimenez, C., Carpenter, T. E., Buening, G.B., 1994. Epidemiology of bovine anaplasmosis and babesiosis in Costa Rica. Preventive Veterinary Medicine, 20: 23-31.

Pfeiffer, D.U, 1996. Issues related to handling of spatial data. In: Proceedings of NZVA/AVA Pan Pacific Conference, Epidemiology and State Veterinary Programmes. 83-105.

Pfeiffer, D.U., Duchateau, L., Kruska, R.L., Ushewokunze-Obatolu, U., Perry, B. D., 1997. A spatially predictive logistic regression model for occurrence of Theileriosis outbreaks in Zimbabwe. In: Epidemiologie et Sante Animale. 12.12.1-3.

Stärk, K.D.C., 1998. Systems for the Prevention and Control of Infectious Diseases in Pigs, Ph.D.Thesis. EpiCentre, Institute of Veterinary, Animal and Biomedical Science, Massey University.

Tan, M., 1993. Cost-sensitive learning of classification knowledge and its applications in robotics. Machine Learning, vol.13: 7-33.

Ting, K.M., 1997. Inducing cost-sensitive trees via instance-weighting. Dept. of Computer Science, University of Waikato.

Ting, K.M., 1998. Boosting trees for cost-sensitive classifications. Dept. of Computer Science, University of Waikato.

Turney, P.D., 1995. Cost-sensitive classification: empirical evaluation of a hybrid genetic decision tree induction algorithm. Journal of Artificial Intelligence Research, vol.2: 369-409.

Williams, B., Dransfield, R., Brightwell, R., 1992. The control of tsetse flies in relation to fly movement and trapping efficacy. Journal of Applied Ecology, 29: 163-179.

Walker, P.A., Moore, D.M., 1988. An inductive modelling and mapping tool for spatially-oriented data. Geographical Information Systems, 2: 347-363.

Webb, G.I., Foo, N., Goebel, R., 1996. Cost-sensitive specialization. In: PRICAI'96: Topics in Artificial Intelligence. 4th Pacific Rim International Conference on Artificial Intelligence. Springer-Verlag. Berlin, Germany, 23-34.

Zhang, H., Holford, T., Bracken, M.B., 1996. A tree-based method of analysis for prospective studies. Statistics in Medicine, 15: 37-49.

CHAPTER 5.

Development of a simple geographical disease reporting and analysis system for Thailand

---

Introduction

With the advent of computer-based electronic data processing, it has become possible to process information about a large number of factors associated with animal populations, disease, and geography. Computer-based systems that combine cartographic (map-making) capabilities with these information-processing capabilities are known as *geographic information systems* (GIS).

In recent time, the development of GIS has presented epidemiologists with an opportunity to explore and analyse the geographical or spatial distribution of diseases as well as any other aspects of disease distribution. The application of GIS in veterinary epidemiology and disease control has become quite common as part of projects in many countries concerning diseases such as bovine tuberculosis (Clifton-Hadley *et* al., 1993), theileriosis (Dvorak, 1993), pseudorabies (Norman *et* al., 1996), Lyme's disease (Kitron *et* al., 1997; 1998), avian influenza (Akey, 1993), trypanosomiasis (Reid *et* al., 1997), foot-and-mouth-disease (Sanson *et* al, 1993) and some parasitic zoonoses (Mott *et* al., 1995). This trend is certain to continue and expand to include more countries as the potential power, usefulness and cost-effectiveness of GIS is being realised.

Because GIS is a relatively new and high cost technique, the use of this tool in developing countries, especially as a component of animal health information systems, represents a difficult challenge. In addition, it is not sufficient to purchase the hardware and software, but it is also necessary to train staff in technical aspects of GIS. Considering all these difficulties, it does not mean that GIS is not practical or not suitable for animal health information systems in developing countries. As an initial step to allow a gradual transition towards the new technology, a comprehensive geographical epidemiological study and a reporting system can be set up.

The objective of the current study is to generate a template for the development of an information system for simple standard reporting on geographical patterns of disease occurrence and spatial cluster analyses of the major diseases in Thailand.

Review of GIS Components

GIS has served important roles as an integrating technology. Rather than being completely new, GIS has evolved by linking a number of discrete technologies into a 'whole that is greater than the sum of its parts'.

As any other information system, GIS includes four major components. These are data input: data storage, data manipulation and analysis, and data reporting. The relationship between a GIS and a traditional animal health information system is outlined in Table 29.

Table 29. Relationship between a GIS and a traditional animal health information system

Processing task	*Geographical information system*	*Animal health information system*
Data input	Geographical feature: digitising or scanning from maps, aerial photos, satellites, and other sources	Attribute data with geo-reference: populations, disease reporting, vaccination
Data storage	GIS files	Database file
Data manipulation and analysis	Spatial operations, analysis, modelling, including spatial statistic	Tabular operations, non-spatial statistic, cross-tabulate
Output and reporting	Maps, plot, display, text reports, tables	Tables, text reports

GIS is an integrative technology. Whereas other technologies might be used only to analyse aerial photographs and satellite images, to only create statistical models, or to draft maps, these capabilities are all offered together within a comprehensive GIS (Foote *et* al., 1998).

GIS, with its array of functions, should be viewed as a process rather than as merely software or hardware. GIS is a tool assisting in decision making. The way in which data

is entered, stored, and analysed within a GIS must mirror the way information will be used for a specific research or decision-making task.

Administrative Structure for Animal Disease Control and Eradication in Thailand

The country is politically divided into 76 provinces (*changwat*), 844 districts (*amphoe*), and 6,404 sub-districts (*tambon*). The governmental system is highly centralised and based in Bangkok where the central office of the Department of Livestock Development (DLD) is located. The DLD divides the country into nine Livestock Development Regions (DLD regions), which do not correspond to the regions used by the National Statistics Office of Thailand. There are DLD regional offices in each region, province, and district. The current administrative structure for animal disease control and eradication in Thailand is graphically shown in Figure 26. Figure 27 shows the procedural steps followed in the event a disease outbreak occurs (Division of Disease Control, 1996).

Figure 26. Current administrative structure of the government veterinary service in Thailand

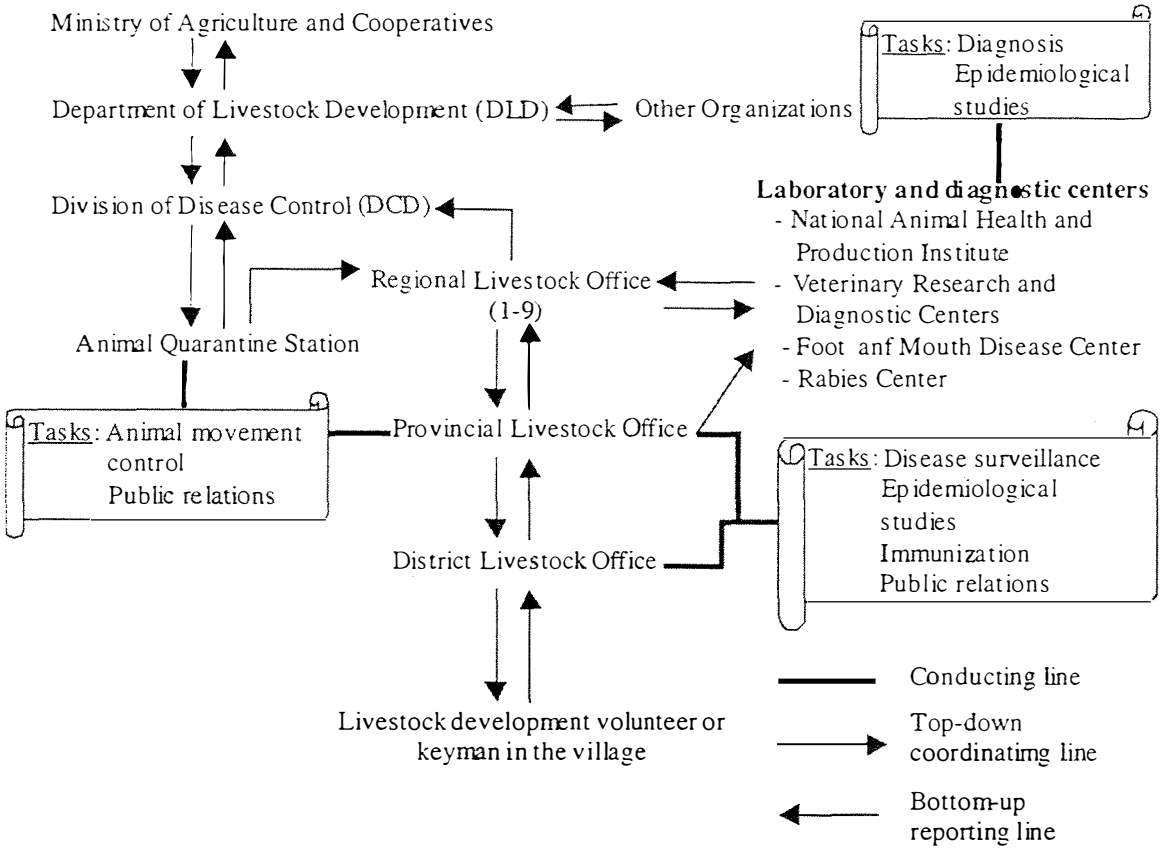
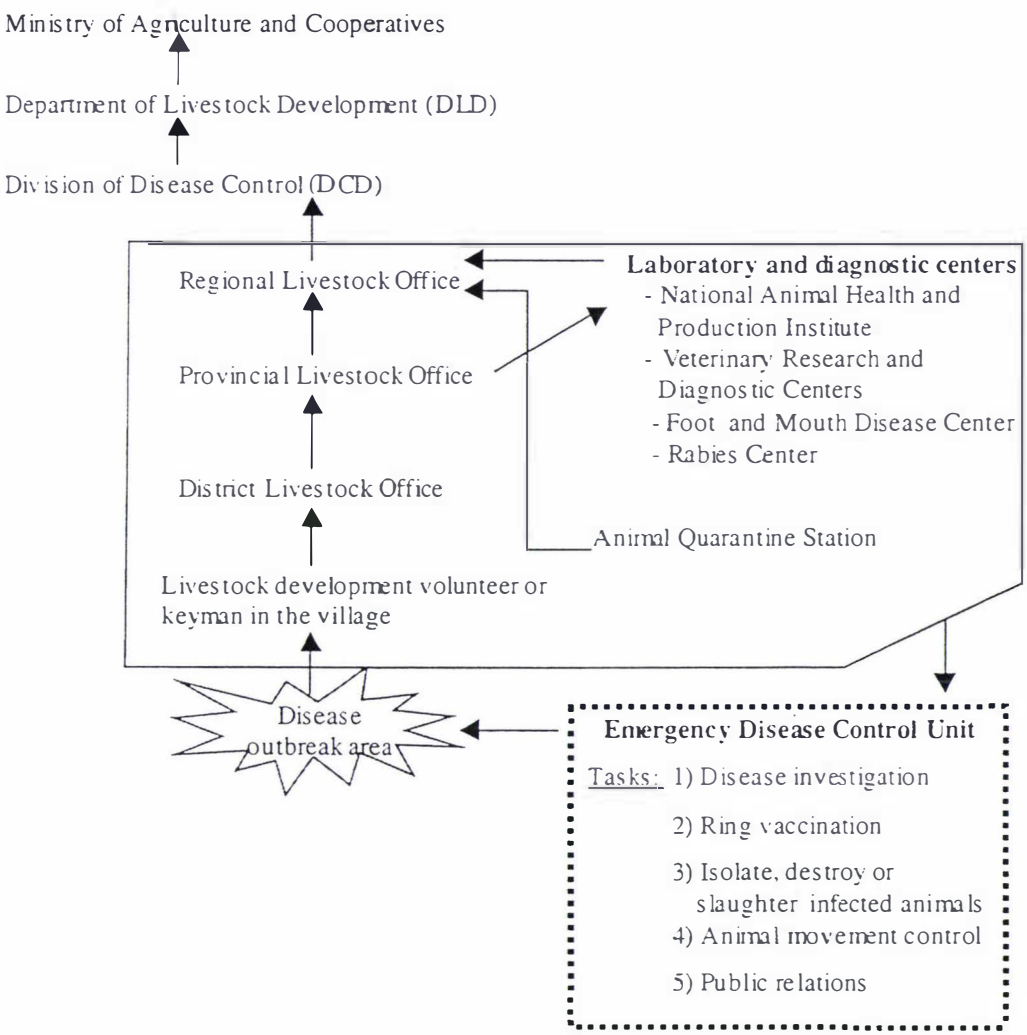


Figure 27. Flowchart of the responsibilities of the Emergency Disease Control Unit during disease outbreaks within the administrative structure of the government veterinary service



System Requirements

Software

GIS software

The GIS software ArcView 3.1® (ESRI Inc., California, U.S.A.) was chosen as the software platform for the system design. ArcView 3.1® simplifies the task of spatial data query and map generation, and it includes virtually all the key functions for manipulation and spatial analysis of geo-attribute data. Using a familiar Microsoft Windows™ interface, it allows the user to quickly and intuitively create interactive

maps. ArcView 3.1® can display an almost infinite variety of geographically linked data, view associated database tables, present graphical data displays, and develop presentation quality maps for final output.

ArcView can be used for the *ad-hoc* linking and display of any of the geographical and attribute data, and the creation of maps. More complex attributes can be derived by linking multiple files and calculating new values from existing data. It is also possible to connect to external databases through ODBC connectivity.

ArcView's programming language called Avenue™ offers the ability to develop automated procedures allowing routine tasks to be carried out quickly and simply. It allows the addition of any new functions useful for data management and analysis. Avenue is a fully integrated object-orientated programming (OOP) language and development environment (Amir, 1997). Complex programs can be developed to create a customised user-interface and automate a wide range of routine tasks. This approach is used to fit the unique needs of this present project.

Database management software

The database management software used for the entry and manipulation of animal health data in this project is Microsoft Access 97 (Microsoft Corporation, Redmond, WA, and U.S.A.) which can be easily linked to ArcView via an ODBC connection. MS Access enables the handling of very large data files. It can manage a database by providing an efficient structure for storage and retrieval of information. Because MS-Access is a relational database management system (RDBMS), this approach makes it easy to bring related data together by establishing relationships to minimise duplication of data. This process can maximise the speed and accuracy of working with the data.

Spatial cluster analysis software

SaTScan® version 2.1.3 (National Cancer Institute, 1998, Bethesda, MD, U.S.A.) is used to apply the spatial scan statistic to the data. This method is used for identification of the presence and location of possible disease clusters in a population with inhomogeneous spatial density, and to estimate likelihood ratios based on Monte Carlo replications to allow statistical inference with respect to the detected clusters (Kulldorff, 1995).

Hardware

The system was originally installed on a personal computer (PC) with an Intel Pentium microprocessor, running at 166 MHz, 32 MB of RAM as recommended for ArcView 3.1, and a 1.99 GB hard disk. The system runs under the Windows 95 operating system.

Data Requirements

In addition to hardware and software, the third essential element of a GIS is the data which comprises of spatial feature data such as maps of administrative boundaries and of attribute data such as disease reporting data.

Spatial feature datasets

All maps used in this study are in vector format and represent geographical features at 1:250,000 scale (or 1cm: 2.5 km) except for a raster map used in the spatial analysis. This raster map was converted from a district polygon coverage (but excluded districts in DLD regions 2, 8, and 9 as these regions have been declared FMD free zones by the government) to a grid cell theme using a resolution of 25 x 25 kilometre (km) for each grid cell square. The two geographical data sets used in this project are listed in Table 30.

Table 30. Spatial datasets used as part of this project

Spatial dataset category	*Data (feature type)*	*Description*
Administrative	DLD Region (polygon)	Boundaries of DLD region
	Province (polygon)	Boundaries of province
	District (Point and polygon)	Boundaries and centroid of district (*amphoe*)
	Sub-district (point)	Centroid of sub-district or (*tambon*)
	Grid cell (25 x 25 km)	Converted from district polygon coverage excluding districts in DLD regions 2, 8, and 9
Infrastructure	Transportation and route (arc)	Type of transportation features

Attribute data

Geographical attribute data

The attribute tables of geographical features and associated data tables were linked using a geo-reference code. The coding scheme for administrative purposes used in this project has been defined by the National Rural Development Committee (NRD2C), except for the DLD region identification code. The unique identification code for each province, district, and sub-district is used to connect to the tables, which provide geographical attribute data such as names of sub-districts, districts, and provinces for the whole country.

The transportation routes related to nine types of transportation features were aggregated into four categories: highway, main road, secondary road and minor road, and railroad.

Animal health attribute data

An expert consultancy report commissioned by FAO (1994) describes a number of specific data elements which should be part of an effective animal health information system. The four elements used as part of the current animal disease control information system in Thailand are 1) livestock demographic, 2) disease specific monitoring, 3) laboratory diagnostic and 4) animal movement data. In this project, only the first three data elements are used.

The livestock population profile is updated annually by the District Livestock Office. Livestock population data at the level of district and sub-district were used in the project. Since the livestock census data in Thailand is currently collected at the district level, the livestock population in sub-districts was estimated as the product between the livestock density for each district (animals per square km.) and the area (square km.) of the sub-district.

Disease-reporting data is also collected at the district level. The district livestock officer must report the details of disease outbreaks for all notifiable diseases as soon as the outbreak occurs. At the same time, the district livestock officer has to send specimens from the infected animal to the laboratory diagnostic centre for confirmation and virus typing in the case of a FMD outbreak. All information flows are from the regional office to the central office for analysis, interpretation, and reporting of disease patterns.

The results are being used to develop a disease control work plan for the next year (Figure 27).

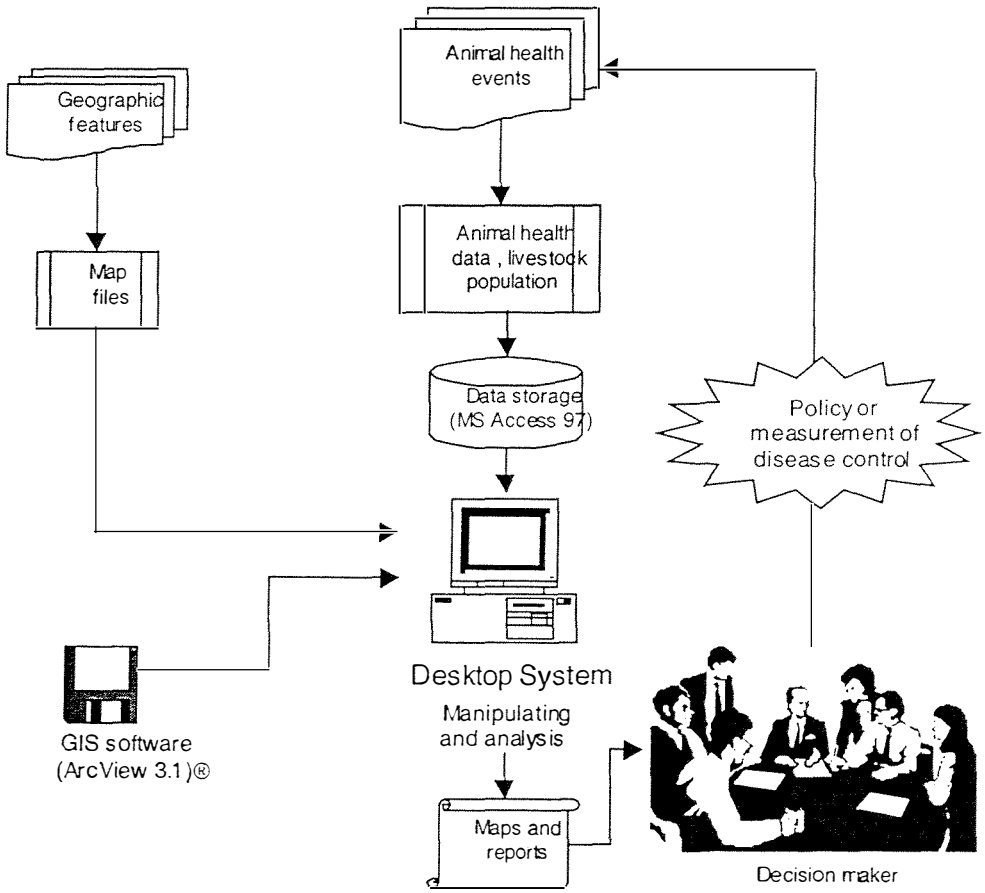
All data are kept in an Access database which can be accessed from ArcView so that up-to-date data is available at all times.

Project Overview

The power and flexibility of a GIS can help to achieve the aim of animal health improvement by providing better information to decision-makers (Figure 28). Knowledge about the spatial component of disease distribution has the ability to considerably enhance the effectiveness of most functions of an animal health information system, and in addition it can also provide entirely new areas of functionality.

The potential applications for GIS in animal disease control in Thailand range from use in epidemiological studies to animal disease surveillance. The main two areas of use in epidemiological field studies include the visual display of geographical patterns and spatial analysis. In the area of surveillance, GIS allows production of maps of disease occurrence and it can be part of a sophisticated animal disease information system (Pfeiffer, 1994).

Figure 28. Components of a simple geographical disease reporting and analysis system for Thailand



The system comprises of three main functions; a routine report generator, outbreak management, and data analysis tools. The functionality includes simple data visualisation techniques as well as analysis for spatial cluster detection.

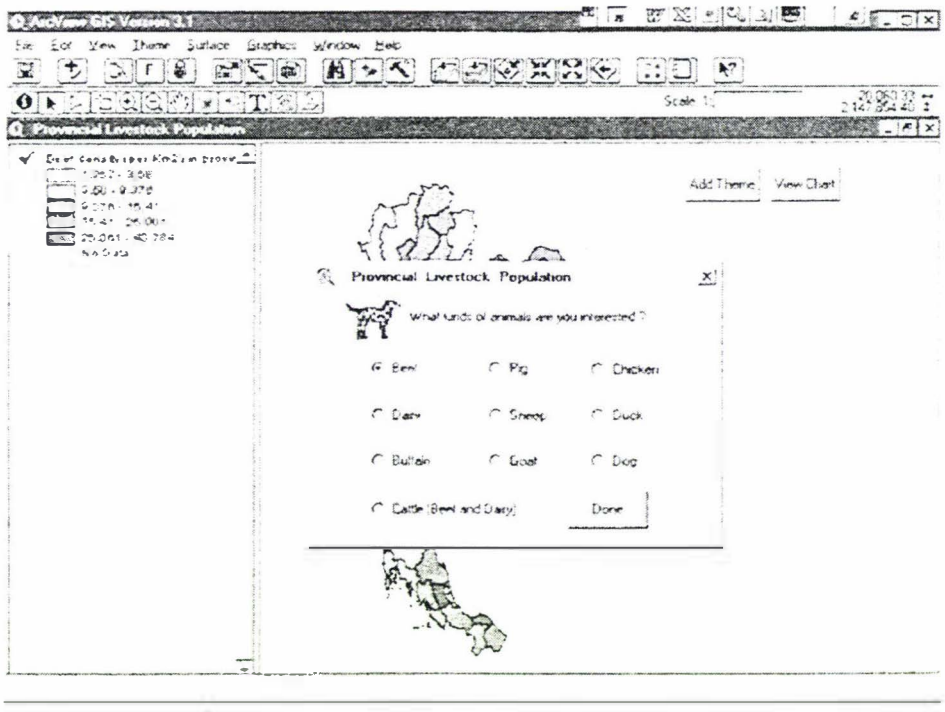
Routine report generator

The routine report menu includes maps for visualisation of livestock density and disease occurrence. The generation of routine reports is important for administrative management, control program planning, priority setting, and to satisfy international reporting requirements and thereby represents a key function in any animal health information system.

Livestock population

The system allows production of maps showing livestock density for any relevant animal species by choosing data aggregation at the area, district or provincial level. However, if the actual livestock population for a specific district is required, a bar chart showing the livestock population numbers for the district of interest can be generated interactively by clicking with the cursor on the relevant location on the map (Figure 29).

Figure 29. Provincial livestock population view (includes "AddTheme" and "ViewChart" buttons for visualising the density of any livestock species at the province level)



Disease occurrence

Data used to produce disease occurrence maps were collated on the basis of the current passive reporting system. The regional offices have to report all outbreaks of notifiable diseases according to the Animal Epidemic Act B.E. 2499 (in 1956 AD). This data can be used to estimate the cumulative incidence of infected animals. This information should be interpreted with caution since under-reporting is common with passively acquired data. With the system, maps of cumulative incidence for any diseases are drawn based on the data which are stored in an Access database file to be updated regularly.

In this study, the cumulative incidence by area, district or province, over a given period of time was used to display the disease occurrence. This is done by aggregating the number of new cases reported in each outbreak at the area level, then dividing by the number of animals at risk in that area.

Selection of appropriate category cut-points can be performed in a number of ways (ArcView ® GIS, 1996). For the present study, the quantile classification method was used to display the disease map. Each class presented on the map is assigned the same number of observations. Quantile classes can therefore be misleading because low values are often included in the same class as high values. However, this distortion can be overcome by increasing the number of classes. The quantile classification method is useful for emphasising the relative position of a feature among other features.

For example, an FMD cumulative incidence map is easily created. The system allows the user to choose an administrative boundary level of interest, such as district or province, and a specific duration of time. It then draws cumulative incidence maps at the selected level of spatial aggregation by obtaining the relevant FMD reporting data for that period from the Access database and summarising incidence data for each outbreak (Figure 30). A bar chart for selected areas can be viewed as well (Figure 31).

Figure 30. Map of cumulative incidence of FMD. User can choose the district or province level of aggregation and a period of time for disease reporting.

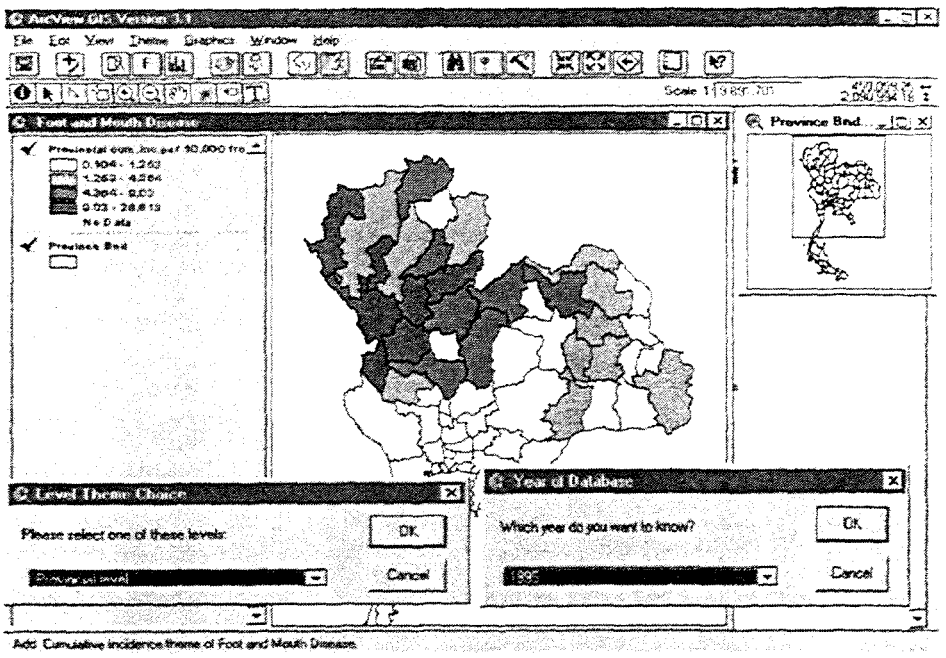
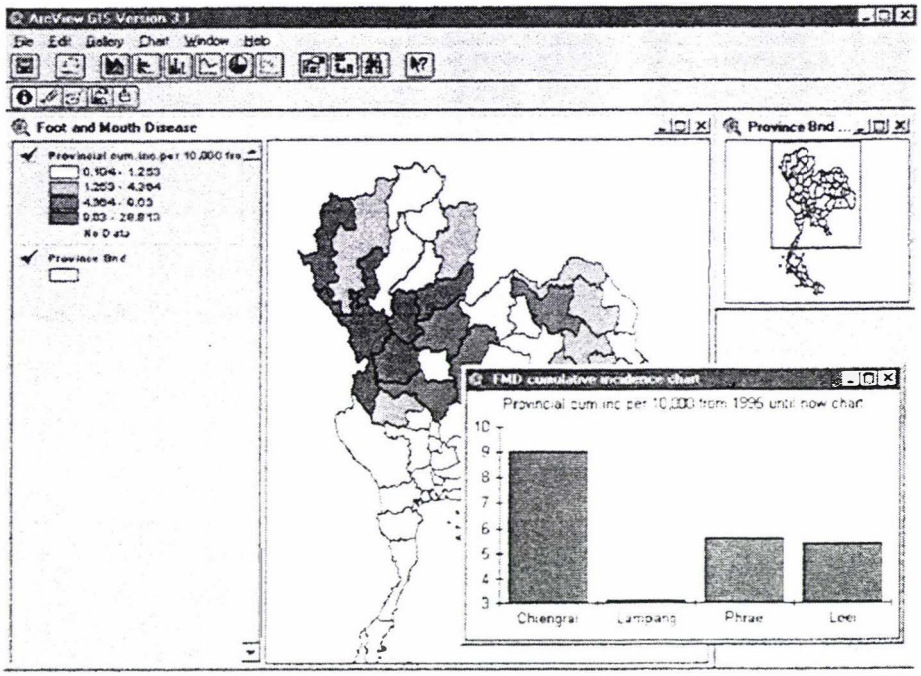


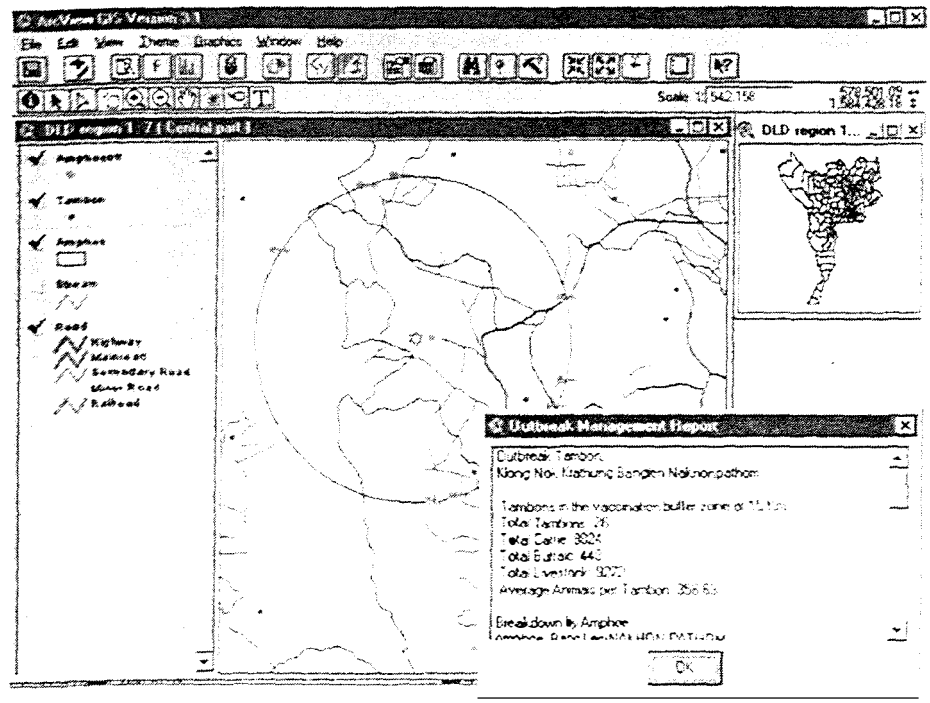
Figure 31. Interactively generated bar charts of cumulative incidence for selected provinces shown in yellow on background map



Outbreak management

Primary information for the current disease outbreak should be collected immediately in order to allow effective planning of control activities. The outbreak management strategy designed specifically for the current system in Thailand consists of five activities (Figure 27). It was adapted from Cameron's system (Cameron *et* al., 1997), which is able to assist the conducting unit (Emergency disease control unit) in the management of the disease outbreak by providing the basic information for the tasks of ring vaccination and animal movement control. This action allows the user to choose the radius of the vaccination buffer zone, draws the corresponding map and generates a message box with statistics on livestock population in the sub-district where the outbreak occurs as well as other sub-districts within the buffer zone. The name of the District Livestock Office and all road intersections with the buffer zone are also presented (Figure 32).

Figure 32. Outbreak management system allowing the user to choose the radius of the vaccination buffer zone. This operation will draw the corresponding maps and generate a message box with various relevant statistics



Data analysis

The data analysis menu consists of two main functions: prediction of FMD outbreaks and spatial cluster analysis.

Prediction of FMD outbreaks

The system estimates the probability of occurrence of FMD outbreaks in any grid cell area based on the analysis result from the classification tree model (CART) described in Chapter 4. The disease surveillance data on FMD outbreaks in Thailand from January 1995 to May 1997 were used to construct the CART model for prediction of FMD outbreaks for each grid cell based on a selection of risk factors accessible as attribute data (Figure 33).

Factors associated with occurrence of FMD outbreaks can vary over time such as the density of buffalo in that area in turn affecting the probability of occurrence of FMD outbreaks. This problem can be addressed by the user specifying the current information for that area before estimating the predicted risk of FMD outbreaks (Figure 34).

Figure 33. Probability of FMD outbreak and 95% confidence intervals based on a classification tree model shown for a selected grid cell

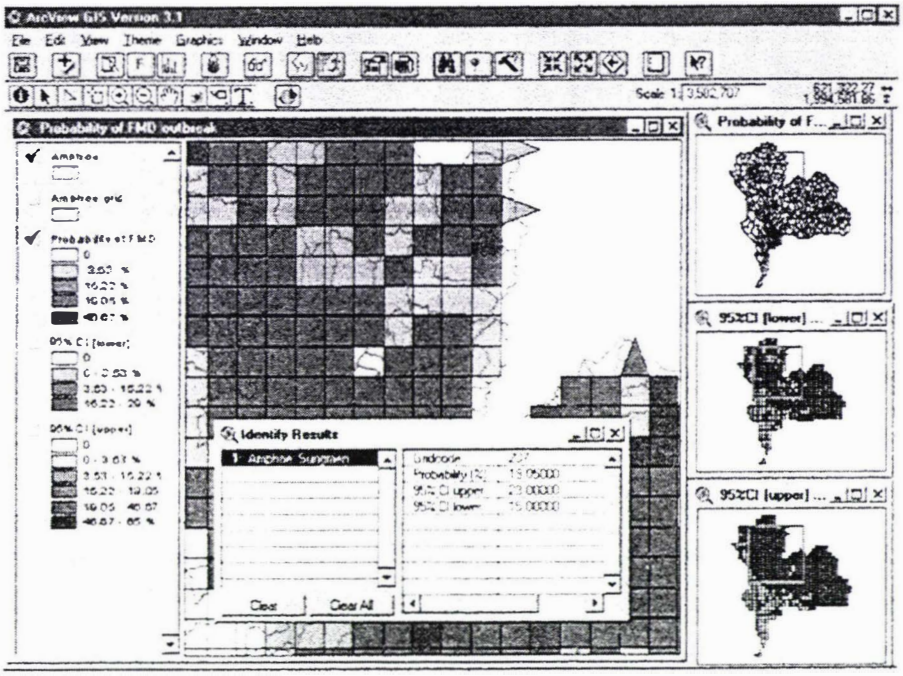
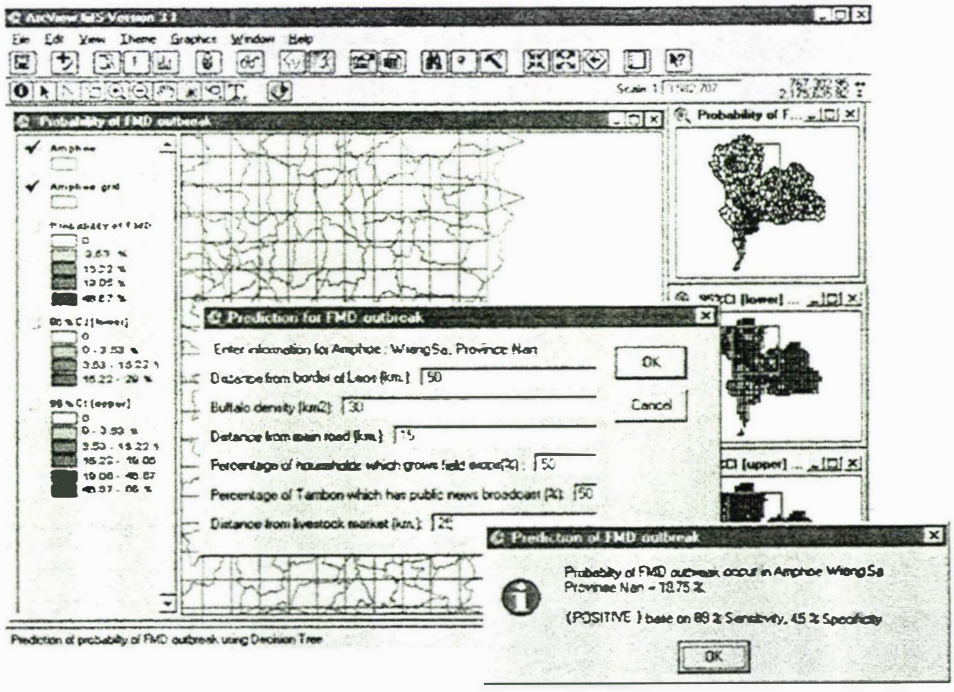


Figure 34. Occurrence of FMD outbreaks at a selected grid cell with user defined information in dialog box and message box showing the resulting predicted outbreak risk based on the classification tree model

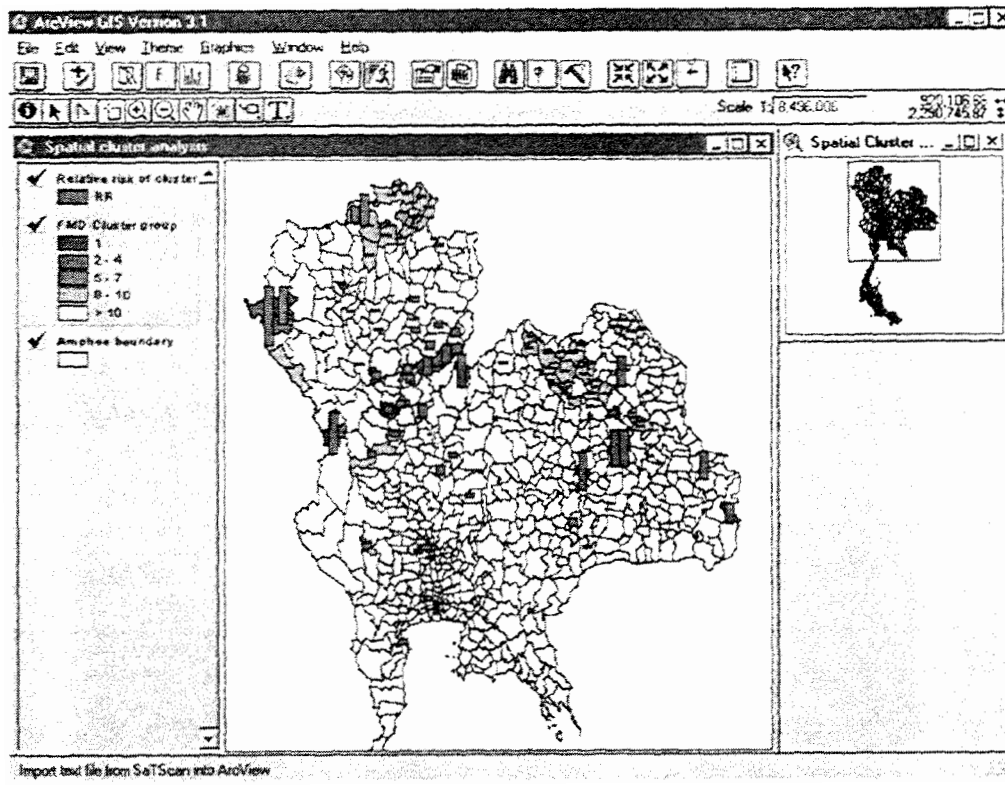


Spatial cluster analysis

On-going research is essential for the effective implementation of disease control programs. The examination of geographical location as a factor in disease occurrence or spatial pattern is part of the disease distribution study. The spatial patterns of disease distribution can be investigated routinely using a GIS.

Using the Avenue programming language the software, SaTScan®, is called to perform spatial cluster analysis within ArcView®. SaTScan® conducts the spatial analysis scanning for high rate disease clusters using a Poisson model. Disease reporting data from an up-to-date database can be transferred to SaTScan® via ArcView®. The data transfer involved minor data manipulation to make adjustments to the text file formats generated by these two programmes. The analysis results from SaTScan® are used to produce maps of relative risk estimates for each district and identified cluster (Figure 35). This type of connectivity has been named bi-directional integration by Anselin (1992).

Figure 35. Map showing the output generated by SaTScan® and displayed using ArcView®.



Discussion

The complexity, cost, and lack of appropriate digital data represent a difficult challenge for the use of GIS in veterinary applications. This situation is likely to change as soon as GIS has proliferated into other areas, and digital data becomes more easily accessible. The key will be shared use of data. This will then allow increasing use of GIS in the veterinary field for animal disease control. For example, Sanson *et* al. (1993) developed EpiMAN, an advanced decision support system for management of an outbreak of foot-and-mouth-disease. This system combines database management systems, GIS, expert system elements, and simulation models of foot-and-mouth-disease. It is certain that the cost of establishing such as system is generally high but should be seen in the context of the costs poorly managed disease outbreaks may be incurring. The cost of GIS technology has been reviewed by a number of authors (Sharma, 1997; Tim, 1995; Harrison *et* al., 1997; Akey, 1993). Although most came to the conclusion that GIS is cost effective, this issue is still a significant problem for developing countries such as Thailand. In fact, it remains questionable whether immediate implementation of such complex animal health information systems should be a goal for countries such as Thailand. A gradual process starting with simple systems and leading towards more sophisticated systems is probably a more realistic option. The usefulness of GIS for national animal disease control has already been recognised by the Thai government.

As in many other countries, GIS technology has been applied to many areas within Thailand, such as environmental management and military activities. A basic digital geographical database at a national scale has become available. This will provide the basis for applying this technology to animal health management in Thailand.

The current project developed a simple geographical disease reporting and analysis system for Thailand. This simple system aims at improving the understanding of the geographical patterns of major animal diseases in Thailand such as FMD, and it can be used to produce custom maps for routine reporting for decision makers. The system functions are largely automated and only require the operator to possess basic computer skills. Thus this system is within the capabilities of virtually all government veterinary services.

As with all information systems, availability and quality of animal health data used in the system is one of the most important issues, especially disease specific surveillance data. These data are used in spatial analysis and mapping with a GIS at the national level. Therefore, the improvement of the quality and detail of animal health surveillance data should be considered.

It has been stated as part of the national animal disease control and eradication strategy for Thailand that ring vaccination is one strategy for controlling disease outbreaks. Ring vaccination involves strategic vaccination of animals at risk in areas surrounding an outbreak to provide a barrier against spread of infection. The main aim of vaccination is to reduce the amount of circulating virus by decreasing the number of susceptible animals in the vaccination zone and consequently reducing the amount of virus excreted (Donaldson *et* al., 1992). Movement restrictions have to be placed on the vaccinated animals in the vaccination buffer zone because these animals can become silent excretors if they were infected before full immunity developed, or they can become carriers if they were infected later (Donaldson *et* al., 1992). Checkpoints have to be set up at major livestock transport routes passing through the outbreak area to control animals moving in or out of this area.

With livestock husbandry systems in developing countries such as Thailand, the best locational unit for specifying an outbreak area, in the epidemiological sense, is the village since the livestock are kept in a contiguous area and are often mixed together while grazing during the day. The village is also the smallest unit of the administrative system in Thailand. Geographical data of village locations at the national scale is not available at present, therefore the sub-district was used to specify outbreak locations. This is too crude a scale relative to the scale at which outbreaks will occur and the biological mechanisms relevant to the epidemiology of FMD. But it should still be possible to use the current scale used for recording locational data about disease occurrence to demonstrate the use of the system in the case of disease outbreaks for providing information to the emergency disease control unit to assist when planning ring vaccination and animal movement control tasks.

The simple geographical reporting and analysis system developed as part of this project fulfils the following tasks:

- Production of livestock population maps

- Cumulative incidence maps of the distribution of major diseases
- Provision of information for ring vaccination and animal movement control tasks in the case disease outbreak,
- Perform spatial cluster analysis using the spatial scan statistic and
- Production of outbreak risk maps for visual presentation of potential disease clusters

Accurate maps have to be generated regularly and passed on to the decision-makers, not just in the event of a disease outbreak. It is hoped that the maps and spatial cluster analysis results can provide useful information about disease patterns for decision-makers to plan disease control strategies more effectively and more quickly. The current system provides relatively simple reporting and analysis functionality and it is hoped that it will provide the basis for a gradual transition towards a more sophisticated decision support system.

Reference List

Akey, B.L., 1993. Establishment and use of a geographic information system (GIS) to control avian influenza. In: Proceedings of the Annual Meeting of the United States Animal Health Association. 64-8.

Anselin L., 1992. Spatial data analysis with GIS: An introduction to application in the social sciences. Volume 92-10. Santa Barbara, California: National Center for Geographic Information and Analysis. Technical Report 92-10.

Cameron, A.R., Sharma, P., Chamnanpood, P., 1997. Implementation of a simple outbreak response management system for developing countries using a low-cost GIS. In: Epidemiologie et Sante Animale, No. 31/32, 02.14.1-02.14.3.

Clifton-Hadley, R.S., Thrusfield, M.V., 1993. The use of a geographical information system (GIS) in the control and epidemiology of bovine tuberculosis in south-west England. In: Society for Veterinary Epidemiology and Preventive Medicine. Roslin, UK. 166-79.

Division of Disease Control, Ministry of Agriculture and Cooperatives, 1996. Co-ordinating system for animal disease control and eradication in Thailand. In: Animal Disease Prevention, Control, and Eradication as a Part of Animal Health

Work Plan; Department of Livestock development; according to The Eighth Plan of National Economic and Social Development (1997-2001) [in Thai]. Thailand.

Donaldson, A.I., Doel, T.R., 1992. Foot-and-mouth disease: the risk for Great Britain after 1992. Vet. Rec. 131, 114-20.

Dvorak, K.A., 1993. Spatial projection of socioeconomic data using geographic information systems: results from a Kenya study in the strategic implementation of a livestock disease control intervention. Social Science Research for Agricultural Technology Development: Spatial and Temporal Dimensions. CAB International, Wallingford, Oxon, UK. 37-50.

FAO, 1994. Report of the FAO expert consultation on the need for information systems to strengthen veterinary services in developing countries. Rome, Italy.

Foote, K. E., Lynch, M., 1998. Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions. The Geographer's Craft Project, Department of Geography, University of Texas at Austin.

Harrison, S.R., Sharma, P.C., Tisdell, C.A., 1997. Linking GIS with economic models for managing livestock health: possibilities and constraints. In: Epidemiologie Et Sante Animale. No. 31/32, 12.11.1-12.11.3. 2

Kitron, U., Kazmierczak J.J., 1997. Spatial analysis of the distribution of Lyme disease in Wisconsin. American Journal of Epidemiology 145, 558-66.

Kitron, U., 1998. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. Journal of Medical Entomology 35, 435-45.

Kulldorff, M., Nagarwalla, N., 1995. Spatial disease clusters: detection and inference. Statistics in Medicine, 14: 799-810.

Mott, K.E., Nuttall, I., Desjeux, P., Cattand, P., 1995. New geographical approaches to control of some parasitic zoonoses. Bulletin of the World Health Organization 73, 247-57.

Norman, H.S., Sischo, W.M., Pitcher, P., Nesselrodt, A., Day, R.L., 1996. Spatial and temporal epidemiology of pseudorabies virus infection. American Journal of Veterinary Research, 57, 1563-1568.

Reid, R.S., Wilson, C.J., Kruska, R.L., Mulatu, W., 1997. Impacts of tsetse control and land-use on vegetative structure and tree species composition in south-western Ethiopia. Journal of Applied Ecology 34, 731-747.

Sanson, R.L., Morris, R.S., Stern, M.W., 1993. EpiMAN: a spatial decision support system for use in an exotic disease emergency. Agricultural Systems and Information Technology Newsletter 5, 20-22.

Sharma, V.P., Srivastava, A., 1997. Role of geographic information system in malaria control. Indian Journal of Medical Research 106: 198-204.

Tim, U.S., 1995. The application of GIS in environmental health sciences: opportunities and limitations. Environmental Research 71, 75-88.

General Discussion and Conclusions

Different methods for investigating and presenting spatial animal health data have been presented in this thesis. Epidemiological data was first analysed using a range of different spatial analysis methods. Results from some of these analyses were then used for the development of a simple spatial information system. Animal disease data from both Thailand and New Zealand were explored and analysed using the spatial analysis methods.

This chapter discusses the methodology and conclusions reached from this thesis. It also briefly outlines future opportunities for development of geographical animal health information systems in Thailand.

The first three chapters presented how spatial analysis can be used to improve understanding of the patterns of disease distribution through detection of disease clustering. The pattern of disease events can be studied and aggregated to identify the pattern of the distribution of cases in any particular place and/or time. Any disease event occurring in both, the space and the time dimension, may express any of three types of patterns: clustering in time, clustering in space and clustering in space-time or space-time interaction. A number of methods were used to explore the clustering of FMD in Thailand and EBL in New Zealand. It is important to choose the most appropriate method of analysis for any disease cluster investigation, and it will usually involve consideration of a range of different cluster scenarios taking into account the strengths and weaknesses of each method. This places considerable responsibility on the investigator who needs to have a thorough understanding of the data and the epidemiology of the disease being investigated. A range of analytical methods should be applied with due consideration of their strengths and weaknesses relative to the disease process before deciding on the final method of choice.

Considerable attention must be paid to confounding. Since its presence may lead to false positive or false negative results of the statistical spatial data analysis. Although confounding is a complex and difficult issue to address, it is an important component in the interpretation of cluster studies.

Searching for spatial clustering is the first step in spatial data exploration. Studies of disease clusters are valuable since the result of this analysis can then be used to generate specific hypotheses for further investigation.

Spatial data modelling can be used if the investigators would like to explain or predict the spatial patterns of disease occurrence by identifying important risk factors. Chapter four presented the use of logistic regression and tree-based models for the prediction of geographical patterns of FMD outbreaks in Thailand. The potential presence of spatial dependence or spatial autocorrelation must be taken into account when modelling spatial data. The validity of any final model depends upon the choice of an appropriate form of the first-order component of spatial variation and the appropriate covariate parameters used to control for the presence of spatial dependence when fitting the model.

Tree-based models were also constructed to determine decision rules for FMD outbreak prediction using the classification tree (CART) method. Cost-sensitivity concepts were introduced during the model construction through weighting the value of sensitivity and specificity. With this technique, the differential effect of false positive or false negative predictions can be taken into account rather than simply an equal weighting of the number of incorrect predictions. For FMD the cost of not predicting an actual outbreak is far higher than predicting a "non-event" outbreak. This could be easily taken into account with the tree-based analysis approach.

A useful summary statistic of predictive model performance is the receiver-operating characteristic (ROC) curve. The ROC curve plots the various conjunctions of sensitivity and false positive rates as the threshold defining an event/non-event is varied. This characteristic of the ROC curve can also be used to choose appropriate cut-off points for the model that maximise cost-effectiveness of disease control and eradication programmes taking into account the consequences of false positive and false negative model predictions.

The predicted probability values generated by the different types of spatial models were then used to produce risk maps. Each grid cell represented the probability of occurrence of an FMD outbreak at that cells taking into account any attribute information. The model generated using CART was incorporated in the development of the simple geographical animal health reporting and analysis system. Finally, observed and predicted geographical patterns of disease occurrence can be displayed using these spatial analysis techniques.

Chapter five outlines a simple geographical analysis and reporting system which could be introduced as the basis of an animal health information system in Thailand. Simple standard reporting of geographical patterns of disease occurrence and results of spatial disease cluster detection can be generated easily and quickly using this system. The system presented in this thesis is based on the GIS software ArcView 3.1®, the database management software MS Access 97®, the spatial cluster analysis software SaTScan® version 2.1.3 and integrated using the Avenue programming language. The animal health and disease reporting data are stored in an Access database that can be accessed from ArcView®, the advantage being that the dataset used in ArcView is up-to-date at all times.

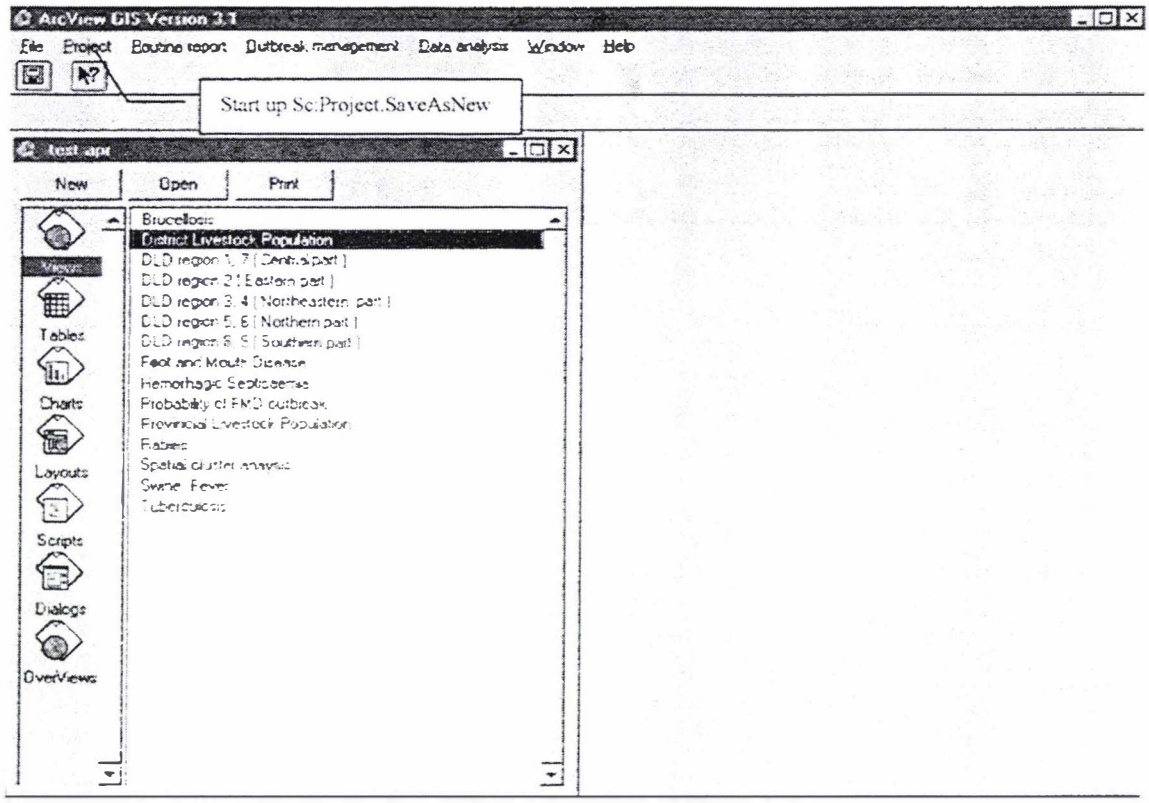
Availability and quality of the data used in the system is one of the most important issues, especially in the case of disease specific monitoring data. At the national level these data provide the basis for spatial analysis and mapping. Therefore, the improvement of animal health monitoring data should be considered a priority so that these data can be used to generate meaningful analyses.

It is hoped that the maps generated and analysis results by such an information system can provide useful information on the disease patterns for decision-makers to improve the effectiveness of planning of control strategies for any animal diseases.

Appendix

List of Interfaces and Associated Scripts

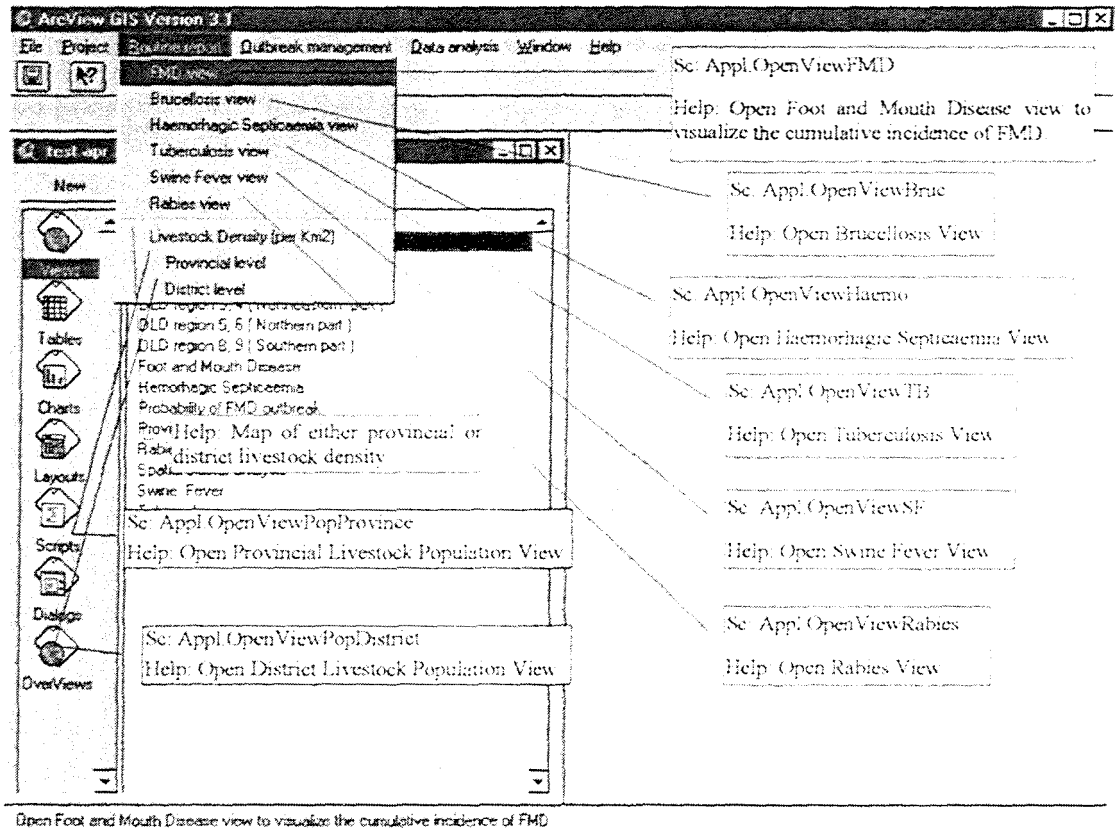
Interface 1 Graphical user interface (GUI) at the beginning of the application. User will be forced to save the default project as new project at the start-up.



Project.SaveAsNew

```
' Forcing Save As at the start-up of the project.
theProject = av.GetProject
defaultName = FileName.GetCWD.MakeTmp ("GIS","apr")
' Open the file window to get a project name.
newProject = FileDialog.Put
(defaultName,"*.apr","create a NEW PROJECT")
NewName="E:/Maps/test.apr".AsFileName
WrongName=(NewName=NewProject)
  if (Nil <> newProject) then
   if (not(WrongName) )then
     theProject.SetFileName(newProject)
     theProject.Save
     Exit
   end
 MsgBox.Error("You can not use this name for your project","Wrong name")
 theProject.Close
 Else
   ' If the user clicked on Cancel, close the project.
 theProject.Close
end
```

Interface 2 "Routine report menu" provides maps for disease status and livestock density visualization (associated scripts for each interface control in callout square).



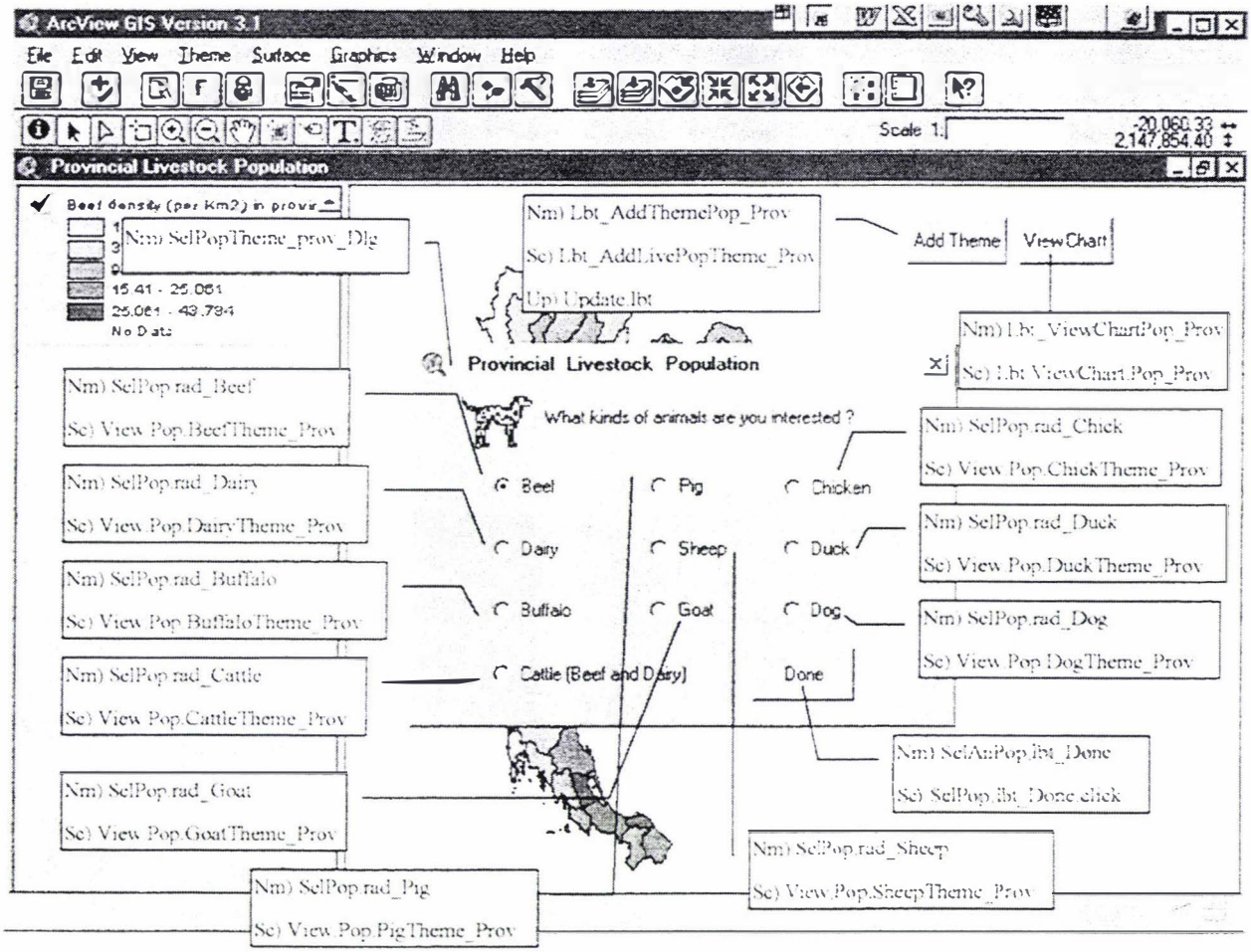
Appl.OpenViewFMD

```
thisProject=av.GetProject
theView=thisProject.FindDoc("Foot and Mouth Disease")
theViewWindow=theView.GetWin
theViewWindow.open
```

Appl.OpenViewPopProvince

```
thisProject=av.GetProject
theView=thisProject.FindDoc("Provincial Livestock Population ")
theViewWindow=theView.GetWin
theViewWindow.open
theViewWindow.maximize
```

Interface 3 Provincial livestock population view comes with "AddTheme" and "ViewChart" buttons to visualize the density of any livestock species at the level of province (associated scripts for each interface control in callout square).



Lbt_AddLivePopTheme_Prov

```
' View.OpenSelPopTheme_Prov_Dlg.Button
' Run the dialog
av.GetProject.FindDialog ("SelPopTheme_Prov_Dlg").Open
```

Update.Lbt

```
theView=av.GetActiveDoc
Lbt_AddTheme=self.GetObjectTag
Lbt_AddTheme.GetControl
```

View.Pop.BeefTheme_Prov

```
'This script used to add a new beef population theme to the current view. Then draw it from shape file.
'This script joined the data of livestock population from Access database by SQL connection into the
feature table of selected into 5 strata.
theView=av.GetActiveDoc
'Establish a data source and load it into the view.
```

```
aSrcName=SrcName.Make("E:/Maps/thai/thaicwatShape/Province.shp")
aTheme=Theme.Make(aSrcName)
theView.AddTheme(aTheme)
aTheme.SetActive(true)
BeefTheme=aTheme.SetName("Beef density (per Km2) in province")
BeefTheme=theView.GetThemes.Get(0)
'If Oracle server is not available, then display a list of available databases to the user for selection.
mySQLConnection = SQLCon.Find ("oracle")
mySQLConnection = SQLCon.Find ("MS Access 97 Database")
mySQLConnection.Login("Tippawon/Took")
 astring="Select * from LivestockProvince"
 theVTab=VTab.MakeSQL(mySQLConnection, astring)
  'Join the theme with the livestock table data
 ProvinceJoinField=theVTab.FindField("ChangwatID")
 BeefVTab=BeefTheme.GetFTab
 for each oneVTab in {BeefVTab}
  theJoinField = oneVTab.FindField("Changwat")
  oneVTab.Join (theJoinField,theVTab,ProvinceJoinField)
  end
 'Set theme's legend. 'Use  Natural for the legend type
BeefLegend=BeefTheme.GetLegend
BeefLegend.SetLegendType
(#LEGEND_TYPE_COLOR)
' Add a normalization by type Field. But First eliminate a known null value in the normalization field
BeefLegend.SetNullValue("SqKm",0)
BeefLegend.SetNormType(#LEGEND_NORMTYPE_FIELD)
BeefLegend.SetNormFieldName("SqKm")
BeefLegend.Natural(BeefTheme,"Beef",5)
'Load the symbol legend from legend file(SymbLeg.avl)
aLegendFile = "E:/Maps/Legend/SymbLeg.avl".AsFileName
Beeflegend.Load(aLegendFile,#LEGEND_LOADTYPE_SYMBOLS)
'Display the Beef population theme and varify that the legend is visible.
BeefTheme.SetVisible(true)
BeefTheme.SetLegendVisible(true)
aViewDisplay=theView.GetWin
aViewDisplay.maximize
```

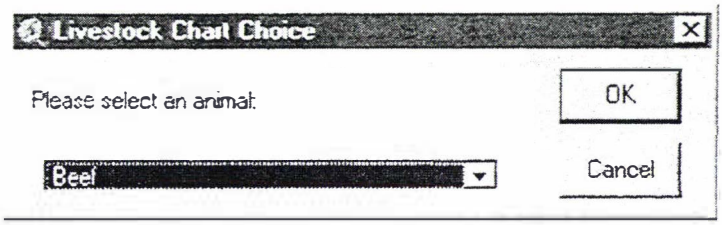
SelPop.lbt_Done.click

```
' SelPop.lbt_Done.click
'Attached to the Done button to close the dialog
self.GetDialog.Close
```

Lbt.ViewChart.Pop_Prov

```
AnimalList={"Cattle (Beef&Dairy)", "Beef", "Dairy", "Buffalo",
        "Sheep", "Goat", "Pig", "Chicken", "Duck", "Dog"}
WhichAnimal=MsgBox.ChoiceAsString
(AnimalList,"Please select an animal: ","Livestock Chart Choice")
'stop the excute if user clicks on the cancel button.
if (nil=WhichAnimal) then
exit
end
if(WhichAnimal="Beef") then
 aScriptBeef=av.FindScript("Ch.Pop.Beef_Prov")
 av.Run("Ch.Pop.Beef_Prov","")
elseif (WhichAnimal="Dairy") then
 aScriptDairy=av.FindScript("Ch.Pop.Dairy_Prov")
 av.Run("Ch.Pop.Dairy_Prov","")
elseif (WhichAnimal="Buffalo") then
 aScriptBuffalo=av.FindScript("Ch.Pop.Buffalo_Prov")
```

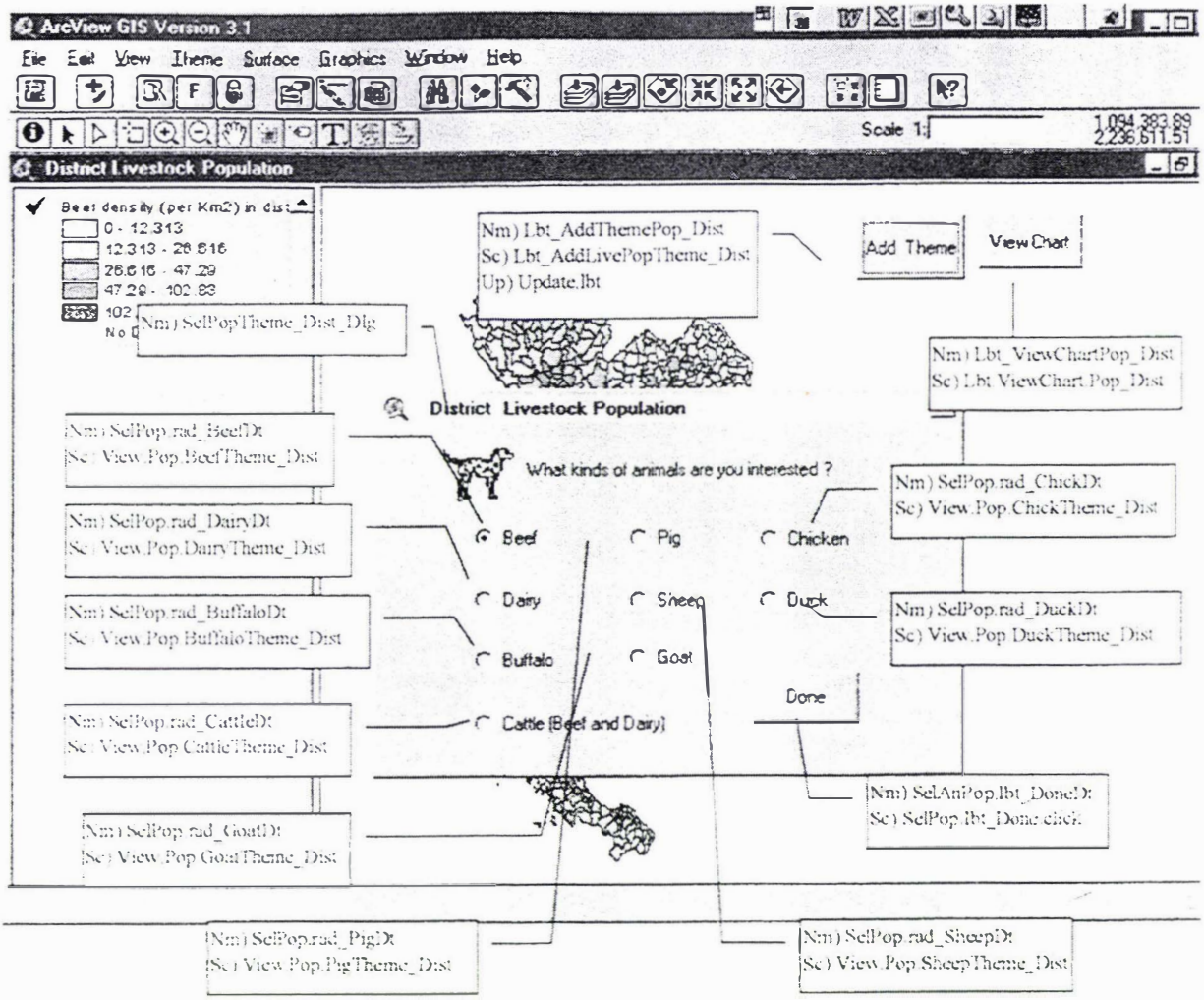
```
  av.Run("Ch.Pop.Buffalo_Prov","")
 elseif (WhichAnimal="Cattle (Beef&Dairy)") then
  aScriptCattle=av.FindScript("Ch.Pop.Cattle_Prov")
  av.Run("Ch.Pop.Cattle_Prov","")
 elseif (WhichAnimal="Sheep") then
  aScriptSheep=av.FindScript("Ch.Pop.Sheep_Prov")
  av.Run("Ch.Pop.Sheep_Prov","")
 elseif (WhichAnimal="Goat") then
  aScriptGoat=av.FindScript("Ch.Pop.Goat_Prov")
  av.Run("Ch.Pop.Goat_Prov","")
 elseif (WhichAnimal="Pig") then
  aScriptPig=av.FindScript("Ch.Pop.Pig_Prov")
  av.Run("Ch.Pop.Pig_Prov","")
 elseif (WhichAnimal="Chick") then
  aScriptChick=av.FindScript("Ch.Pop.Ckick_Prov")
  av.Run("Ch.Pop.Chick_Prov","")
 elseif (WhichAnimal="Duck") then
  aScriptDuck=av.FindScript("Ch.Pop.Duck_Prov")
  av.Run("Ch.Pop.Duck_Prov","")
 else
  aScriptDog=av.FindScript("Ch.Pop.Dog_Prov")
  av.Run("Ch.Pop.Dog_Prov","")
end
Ch.Pop.Beef_Prov
' Name:
Chart.Population.Beef in
provinces
' Title: Displays a column
chart of the number of beef
population
```




```
' Topics: Charts, Views
" Description: Generates a column chart for the selected provincial areas in a Provincial livestock
population view. A new Chart document is created to display the number of beef population in that areas.
' Requires: A livestock population theme must be the active document.
" Self:
" Returns:
' Gather information from the current Document and retrieve the basic information.
theView = av.GetActiveDoc
theTheme = theView.GetThemes.Get(0)
if (nil=theTheme) then
MsgBox.Warning
("Unable to find any Livestock Population Themes "
+NL+"You have to add the Livestock Population Theme first"
+NL+"then select interesting provinces","")
 exit
 end
  theVTab = theTheme.GetFTab
'SelectedRecords=theVTab.GetSelection
if (selectedRecords.count=0) then
'no records were selected
'Therefore,no chart display
MsgBox.Error ("Please select about 5 interesting areas.","")
exit
end
' Retrieve the required fields for charts.
fieldList1 = { theVTab.FindField("Beef")}
if (fieldList1.Get(0) = nil) then
  MsgBox.Error ("Unable to find fields.","")
  exit
end
```

```
'Create the charts and set their properties.
columnChart = Chart.Make (theVTab, fieldList1)
columnChartWin = columnChart.GetWin
columnDisplay = columnChart.GetChartDisplay
columnDisplay.SetType (#CHARTDISPLAY_COLUMN)
columnDisplay.SetStyle (#CHARTDISPLAY_VIEW_SIDEBYSIDE)
columnChart.SetName ("Livestock Population Chart (Beef)")
columnChart.SetSeriesFromRecords (True)
columnChart.GetChartLegend.SetVisible (true)
columnChart.GetYAxis.SetAxisVisible (true)
columnChart.GetYAxis.SetLabelVisible (False)
columnChart.GetYAxis.SetTickLabelsVisible (True)
columnChart.GetXAxis.SetLabelVisible (False)
columnChart.GetXAxis.SetName ("Province")
columnChart.SetSeriesFromRecords(true)
nameField=theVTab.FindField("Province")
columnChart.SetRecordLabelField(nameField)
columnChart.GetChartLegend.SetLocation(#ChartDISPLAY_LOC_BOTTOM)
columnChart.GetYAxis.SetMajorGridVisible(False)
columnChart.GetTitle.SetName ("Beef Population in Province")
if (columnDisplay.IsOK.Not) then
  proceed = MsgBox.YesNo
  ("column chart may have an inconsistency status" ++columnDisplay.GetStatus
    +NL+"Do you want to continue?", "", False)
  if (Not proceed) then
   exit
  end
end
`Add the charts to the project and open them.
thisProject=av.GetProject
thisProject.closeAll
TheView.GetWin.Open
columnChartWin.Open
av.TileWindows
```

Interface 4 District livestock population view comes with "AddTheme" and "ViewChart" buttons to visualize the density of any livestock species at the level of district (associated scripts for each interface control in callout square).



Lbt_AddLivePopTheme_Dist

```
' View.OpenSelPopTheme_Dist_Dlg.Button
' Run the dialog
av.GetProject.FindDialog ("SelPopTheme_Dist_Dlg").Open
```

Update.Lbt

```
theView=av.GetActiveDoc
Lbt_AddTheme=self.GetObjectTag
Lbt_AddTheme.GetControl
```

View.Pop.BeefTheme_Dist

```
theView=av.GetActiveDoc
'Establish a data source and load it into the view.
aSrcName=SrcName.Make("E:/Maps/thai/thaiamphShape/amphur.shp")
aTheme=Theme.Make(aSrcName)
theView.AddTheme(aTheme)
```

```
aTheme.SetActive(true)
BeefTheme=aTheme.SetName("Beef density (per Km2) in district")
BeefTheme=theView.GetThemes.Get(0)
' If Oracle server is not available, then display a list of available databases to the user for 'selection.
mySQLConnection = SQLCon.Find ("oracle")
mySQLConnection = SQLCon.Find ("MS Access 97 Database")
mySQLConnection.Login("Tippawon/Took")
 astring="Select * from Livestock"
 theVTab=VTab.MakeSQL(mySQLConnection, astring)
  'Join the theme with the livestock table data
 AmpherJoinField=theVTab.FindField("AmphurID")
 BeefVTab=BeefTheme.GetFTab
 for each oneVTab in {BeefVTab}
   theJoinField = oneVTab.FindField("Amphur")
   oneVTab.Join (theJoinField,theVTab,AmpherJoinField)
   end
'Set theme's legend.
'Use  Natural for the legend type
BeefLegend=BeefTheme.GetLegend
BeefLegend.SetLegendType
(#LEGEND_TYPE_COLOR)
' Add a normalization by type Field
'But First eliminate a known null value in the normalization field
BeefLegend.SetNullValue("SqKm",0)
BeefLegend.SetNormType(#LEGEND_NORMTYPE_FIELD)
BeefLegend.SetNormFieldName("SqKm")
BeefLegend.Natural(BeefTheme,"Beef",5)
'Load the symbol legend from legend file(SymbLeg.avl)
aLegendFile = "E:/Maps/Legend/SymbLeg.avl".AsFileName
Beeflegend.Load(aLegendFile,#LEGEND_LOADTYPE_SYMBOLS)
'Display the Beef population theme
'and varify that the legend is visible.
BeefTheme.SetVisible(true)
BeefTheme.SetLegendVisible(true)
aViewDisplay=theView.GetWin
aViewDisplay.maximize
```

Lbt.ViewChart.Pop_Dist

```
AnimalList={"Cattle (Beef&Dairy)", "Beef", "Dairy", "Buffalo",
        "Sheep", "Goat", "Pig", "Chicken", "Duck", "Dog"}
WhichAnimal=MsgBox.ChoiceAsString
(AnimalList,"Please select an animal: ","Livestock Chart Choice")
'stop the excute if user clicks on the cancel button.
if (nil=WhichAnimal) then
exit
end
if(WhichAnimal="Beef") then
  aScriptBeef=av.FindScript("Ch.Pop.Beef_Dist")
  av.Run("Ch.Pop.Beef_Dist","")
elseif (WhichAnimal="Dairy") then
  aScriptDairy=av.FindScript("Ch.Pop.Dairy_Dist")
  av.Run("Ch.Pop.Dairy_Dist","")
elseif (WhichAnimal="Buffalo") then
  aScriptBuffalo=av.FindScript("Ch.Pop.Buffalo_Dist")
  av.Run("Ch.Pop.Buffalo_Dist","")
elseif (WhichAnimal="Cattle (Beef&Dairy)") then
  aScriptCattle=av.FindScript("Ch.Pop.Cattle_Dist")
  av.Run("Ch.Pop.Cattle_Dist","")
elseif (WhichAnimal="Sheep") then
  aScriptSheep=av.FindScript("Ch.Pop.Sheep_Dist")
```

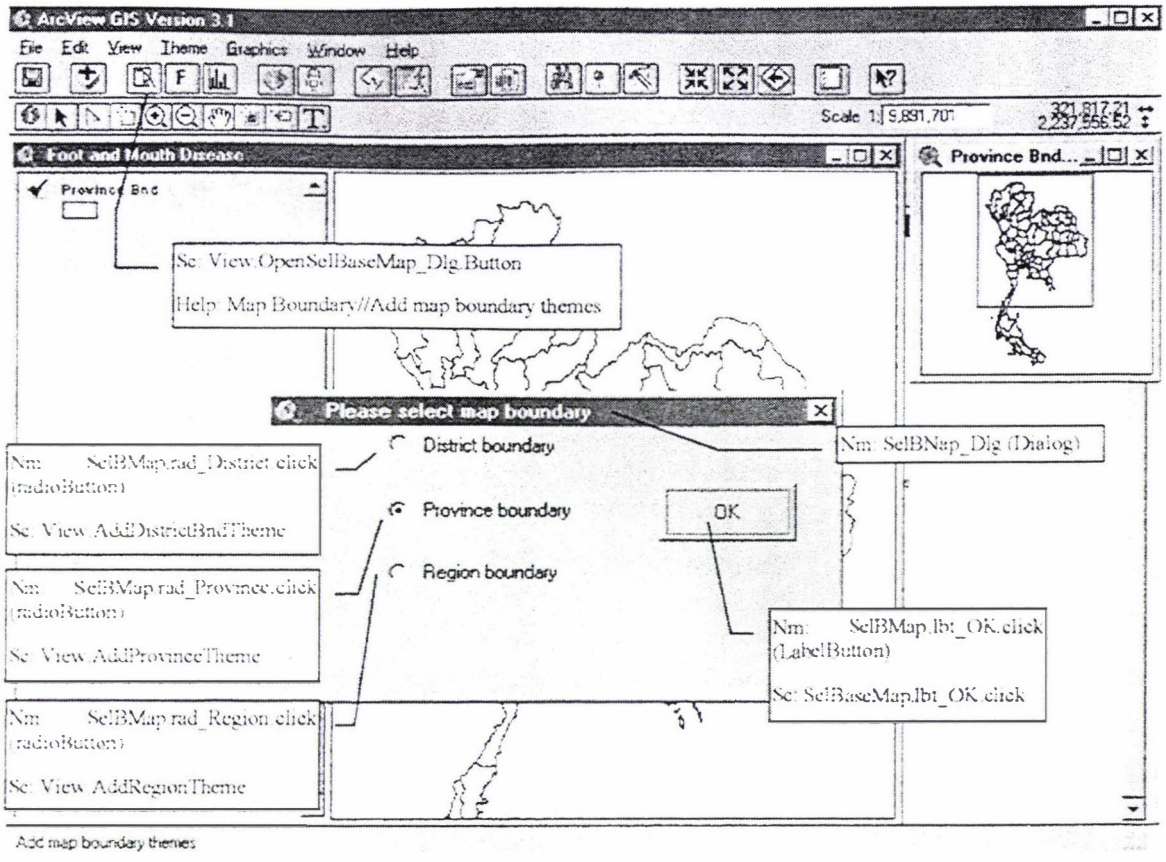
```
  av.Run("Ch.Pop.Sheep_Dist","")
elseif (WhichAnimal="Goat") then
  aScriptGoat=av.FindScript("Ch.Pop.Goat_Dist")
  av.Run("Ch.Pop.Goat_Dist","")
elseif (WhichAnimal="Pig") then
  aScriptPig=av.FindScript("Ch.Pop.Pig_Dist")

  av.Run("Ch.Pop.Pig_Dist","")
elseif (WhichAnimal="Chick") then
  aScriptChick=av.FindScript("Ch.Pop.Ckick_Dist")
  av.Run("Ch.Pop.Chick_Dist","")
elseif (WhichAnimal="Duck") then
  aScriptDuck=av.FindScript("Ch.Pop.Duck_Dist")
  av.Run("Ch.Pop.Duck_Dist","")
else
  aScriptDog=av.FindScript("Ch.Pop.Dog_Dist")
  av.Run("Ch.Pop.Dog_Dist","")
end
```

Ch.Pop.Beef_Dist

```
'Gather information from the current Document and retrieve the basic information.
theView = av.GetActiveDoc
theTheme = theView.GetThemes.Get(0)
if(nil=theTheme) then
MsgBox.Warning
("Unable to find any Livestock Population Themes."
+NL+"You have to add Livestock Population Theme first"
+NL+"then select interesting districts","")
 exit
 end
  theVTab = theTheme.GetFTab
SelectedRecords=theVTab.GetSelection
if (selectedRecords.count=0) then
'no records were selected. Therefore,no chart display
MsgBox.Error ("Please select about 5 interesting areas.","")
exit
end
'Retrieve the required fields for charts.
fieldList1 = { theVTab.FindField("Beef")}
if (fieldList1.Get(0) = nil) then
  MsgBox.Error ("Unable to find fields.","")
  exit
end
'Create the charts and set their properties.
columnChart = Chart.Make (theVTab, fieldList1)
columnChartWin = columnChart.GetWin
columnDisplay = columnChart.GetChartDisplay
columnDisplay.SetType (#CHARTDISPLAY_COLUMN)
columnDisplay.SetStyle (#CHARTDISPLAY_VIEW_SIDEBYSIDE)
columnChart.SetName ("Livestock Population Chart (Beef)")
columnChart.SetSeriesFromRecords (True)
columnChart.GetChartLegend.SetVisible (true)
columnChart.GetYAxis.SetAxisVisible (true)
columnChart.GetYAxis.SetLabelVisible (False)
columnChart.GetYAxis.SetTickLabelsVisible (True)
columnChart.GetXAxis.SetLabelVisible (False)
columnChart.GetXAxis.SetName ("District")
columnChart.SetSeriesFromRecords(true)
nameField=theVTab.FindField("District")
```

```
columnChart.SetRecordLabelField(nameField)
columnChart.GetChartLegend.SetLocation(#ChartDISPLAY_LOC_BOTTOM)
columnChart.GetYAxis.SetMajorGridVisible(False)
columnChart.GetTitle.SetName ("Beef Population in District")
if (columnDisplay.IsOK.Not) then
  proceed = MsgBox.YesNo  ("column chart may have an inconsistency"
  +NL+"Status:"++columnDisplay.GetStatus
  +NL+"Do you want to continue?", "", False)
  if (Not proceed) then
   exit
  end
end
' Add the charts to the project and open them.
thisProject=av.GetProject
thisProject.closeAll
TheView.GetWin.Open
columnChartWin.Open
av.TileWindows
```

Interface 5 ***Foot and Mouth Disease view works with three buttons which will be activated when this view is opened. "Map boundary" button allows user to choose the level of map boundary and also gives the overview of that map level (associated scripts for each interface control in callout square).***



View.OpenSelBaseMap_Dlg.Button

```
' View.SelBaseMap_Dlg_open.Button
' Run the dialog
av.GetProject.FindDialog ("SelBMap_Dlg") Open
```

View.AddDistrictBndTheme

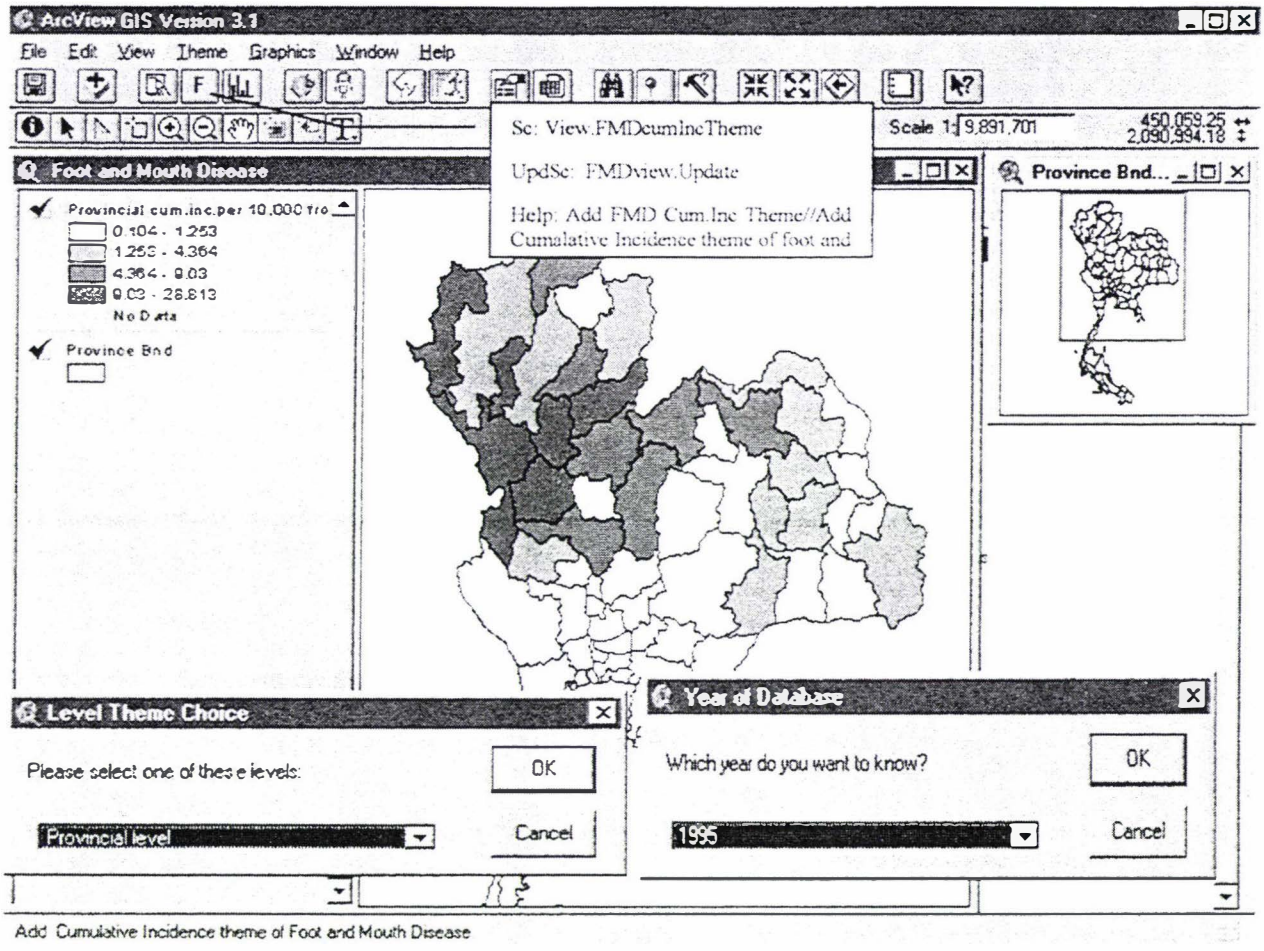
```
' A script used to add a new ThaiAmpherTheme to the current view, create a single symbol legend to
show the 'District boundary, and draw it from shape file of Thaiamph (Amphur.shp). A legend of single
symbol is created 'based on the SngBnd.avl file.
theView = av.GetActiveDoc
' Create the SourceName...
theSrc = SrcName.Make("E:\Maps\Thai\ThaiamphShape\Amphur.shp")
' Use the SourceName to make a theme...
aTheme = Theme.Make(theSrc)
' Add the theme to the view...
theView.AddTheme(aTheme)
' Set a new name for the theme...
aTheme.SetName("District Bnd")
' Get the theme legend and create a single symbol...
theLegend = aTheme.GetLegend
aLegendFile="E:\maps\legend\SngBnd.avl".AsFileName
TheLegend.Load(aLegendFile,#LEGEND_LOADTYPE_ALL)
```

```
'Update the legend in the view TOC...
aTheme.InvalidateLegend
'Draw the theme...
aTheme.SetVisible(True)
aTheme.SetActive (True)
aViewDisplay=theView.GetWin
'Create Overview
'check if an overview window already exists for this view; if so, exit
DistOver=av.GetProject.Finddoc("District Bnd Overview")
  if(not(DistOver=nil )) then
    Msgbox.info(" District Bnd Overview already exists","")
    DistOver.GetWin.Activate
    return NIL
  end
'get the themes to use for the overview
LstTheme = {aTheme}
'get the themes to use for the overview
LstTheme = {aTheme}
'attach the update scripts
theView.SetUpdateScript("Overview.Update")
theView.SetOpenScript("Overview.OpenClose")
theView.SetCloseScript("Overview.OpenClose")
'create the overview view
over = View.MakeWithGUI("Overview")
over.SetTOCWidth(0)
over.SetTOCUnResizable(TRUE)
over.SetName(aTheme.GetName ++ "OverView")
over.SetProjection(TheView.GetProjection.Clone)
'set the extent of the overview and set theme properties
r = Rect.MakeEmpty
for each t in lstTheme
  r = r.UnionWith(t.ReturnExtent)
  t.SetVisible(TRUE)
  t.GetThreshold.SetMaximumOn(FALSE)
  t.GetThreshold.SetMinimumOn(FALSE)
  over.AddTheme(t.Clone)
end
over.GetDisplay.SetExtent(r.Scale(1.1))
'position the window of the overview view
w = over.GetWin
w.Resize(175, 175)
ext = av.ReturnExtent
w.MoveTo(ext.GetX - 185, 5)
'link the view to the overview
over.SetObjectTag(theView)
'add the overview to the project and open it
av.GetProject.AddDoc(over)
over.GetWin.Open
'draw the graphic and bring the main view to the front
av.Run(theView.GetUpdateScript, theView)
theView.GetWin.Activate
```

SelBaseMap.lbt_OK.click

```
'SelBaseMap.lbt_OK.click
'Attached to the OK button to close the dialog
self.GetDialog.Close
```

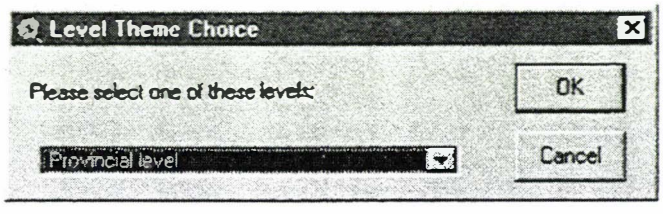
Interface 6 "F" button is used to add foot and mouth disease cumulative incidence map and allows user to choose the level of district or province and year of disease reporting (associated script for each interface control in callout square).



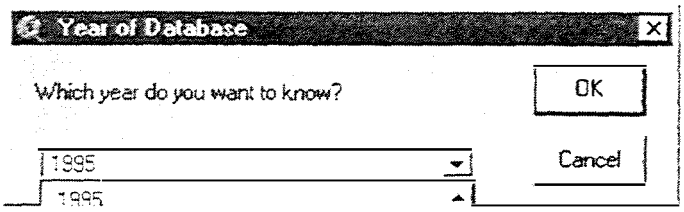
View.FMDcumIncTheme

'This script used to add a FMD cumulative incidence theme to the current view and allowed user to choose interested level, District and Provincial level. Then draw it from shape file of selected level. This script gets the information of FMD from Access database by SQL connection and summarized all incidence in each outbreak into selected level (District or provincial). Then joined the summarized data of FMD into the feature table of selected shape file. A Natural legend is created into 4 strata.

```
'Retrive the boundary base map.
'Let user choose the interested level
LevelList={"District level","Provincial level"}
WhichLevel=MsgBox.ChoiceAsString
(LevelList,"Please select one of these levels: ","Level Theme Choice")
'stop the excute if user clicks on the cancel button.
if (nil=WhichLevel) then
exit
end
theView=av.GetActiveDoc
if(WhichLevel="Provincial level") then
' If user choose Provincial level
' Add Province boundary theme as a base map.
```



```
' Establish a data source and load it into the view.
aSrcName=SrcName.Make("E:/Maps/thai/thaicwatShape/Province.shp")
aTheme=Theme.Make(aSrcName)
theView.AddTheme(aTheme)
aTheme.SetActive(true)
FMDTheme=aTheme.SetName("Provincial cum.inc.per 10,000 ")
FMDTheme=theView.GetThemes.Get(0)
' Retrive the data base of FMD
' Verify that there is a FMD table in this project or not
thisProject=av.GetProject
FMDTable=thisProject.FindDoc("FMD")
if(FMDTable=nil) then
' Add table from SQL conection
mySQLConnection = SQLCon.Find ("oracle")
mySQLConnection = SQLCon.Find ("MS Access 97 Database")
mySQLConnection.Login("Tippawon/Took")
 astring="Select * from FMD9597"
 theVTab=VTab.MakeSQL(mySQLConnection, astring)
 FMDTable=Table.Make(theVTab)
 FMDTable.SetName("FMD")
end
' Retrive the basic information of data table
FMDVTab=FMDTable.GetVTab
abitMap=FMDVTab.GetSelection
aYear = FMDVTab.FindField("Year")
'summarize FMD incidence in each province
'Let user choose the interested year
YearList={"1995","1996","1997","All"}
WhichYear=MsgBox.ChoiceAsString
(YearList," Which year do you want to know?"," Year of Database")
'stop the excute if user clicks on the cancel button.
if(nil=WhichYear) then
exit
end
aQStr= "[Year]>=1995"
'if user choose the year of 1995
if( WhichYear="1995") then
   aQStr= "[Year]=1995"
FMDVTab.Query(
```




```
aQStr,abitmap,#VTAB_SELTYPE_NEW)
 provinceFld = FMDVTab.FindField("Province")
 incFld     = FMDVTab.FindField("Incidence")
 cwatFld    = FMDVTab.FindField("Changwat ID")
 aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/ProIncF".AsFileName,dBASE,
cwatFld, {provinceFld,incFld},
        {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
 sumFMDTable=Table.make(aNewVtab)
 sumFMDTable.SetName ("sumProvIncFMD 1995")
   FMDTheme=aTheme.SetName("Provincial cum.inc.per 10,000 in 1995")
    elseif(WhichYear="1996")then
    aQStr= "[Year]=1996"
   FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
   provinceFld = FMDVTab.FindField("Province")
   incFld     = FMDVTab.FindField("Incidence")
   cwatFld    = FMDVTab.FindField("Changwat ID")
   aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/ProIncF".AsFileName,dBASE,
cwatFld, {provinceFld,incFld},
        {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
   sumFMDTable=Table.make(aNewVtab)
```

```
    sumFMDTable.SetName ("sumProvIncFMD 1996")
   FMDTheme=aTheme.SetName("Provincial cum.inc.per 10,000 in 1996")
      elseif (WhichYear="1997")then
        aQStr= "[Year]=1997"
       FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
       provinceFld = FMDVTab.FindField("Province")
       incFld     = FMDVTab.FindField("Incidence")
       cwatFld    = FMDVTab.FindField("Changwat ID")
       aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary
data/ProIncF".AsFileName,dBASE, cwatFld, {provinceFld,incFld},
         {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
      sumFMDTable=Table.make(aNewVtab)
      sumFMDTable.SetName ("sumProvIncFMD 1997")
    FMDTheme=aTheme.SetName("Provincial cum.inc.per 10,000 in 1997")
      else
       FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
       provinceFld = FMDVTab.FindField("Province")
       incFld     = FMDVTab.FindField("Incidence")
       cwatFld    = FMDVTab.FindField("Changwat ID")
       aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/ProIncF".AsFileName,dBASE,
cwatFld, {provinceFld,incFld},
         {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
       sumFMDTable=Table.make(aNewVtab)
       sumFMDTable.SetName ("sumProvIncFMD")
  FMDTheme=aTheme.SetName("Provincial cum.inc.per 10,000 until now")
end
  FMDTheme=theView.GetThemes.Get(0)
'Join the theme with the livestock table data
theVTab=sumFMDTable.GetVTab
 'Set Alias for the provice name field
ProvinceNameFld=theVTab.FindField("First_Province")
ProvinceNameFld.SetAlias("Name")
cwatJoinField=theVTab.FindField("Changwat ID")
 FMDVTab=FMDTheme.GetFTab
 for each oneVTab in {FMDVTab}
  theJoinField = oneVTab.FindField("Changwat")
  oneVTab.Join (theJoinField,theVTab,cwatJoinField)
  end
FMDTable=thisProject.FindDoc("FMD")
FMDVTab=FMDTable.GetVTab
 allSelected=FMDVTab.GetSelection
 allSelected.ClearAll
'Clear all selection from the FMD table
'Set theme's legend.
'Use  Natural for the legend type
FMDLegend=FMDTheme.GetLegend
FMDLegend.SetLegendType
(#LEGEND_TYPE_COLOR)
FMDLegend.Natural(FMDTheme,"Sum_Incidence",4)
' Set a colorramp
theColorRamp = SymbolList.GetPreDefined(#SYMLIST_TYPE_COLORRAMP).Get(0)
FMDLegend.GetSymbols.RampSavedColors(theColorRamp)
'Load the symbol legend from legend file(FMD4Lev.avl)
aLegendFile = "E:/Maps/Legend/FMD4Lev.avl".AsFileName
FMDlegend.Load(aLegendFile,#LEGEND_LOADTYPE_SYMBOLS)
'Display the FMD theme
'and varify that the legend is visible.
FMDTheme.SetVisible(true)
FMDTheme.SetLegendVisible(true)
exit
```

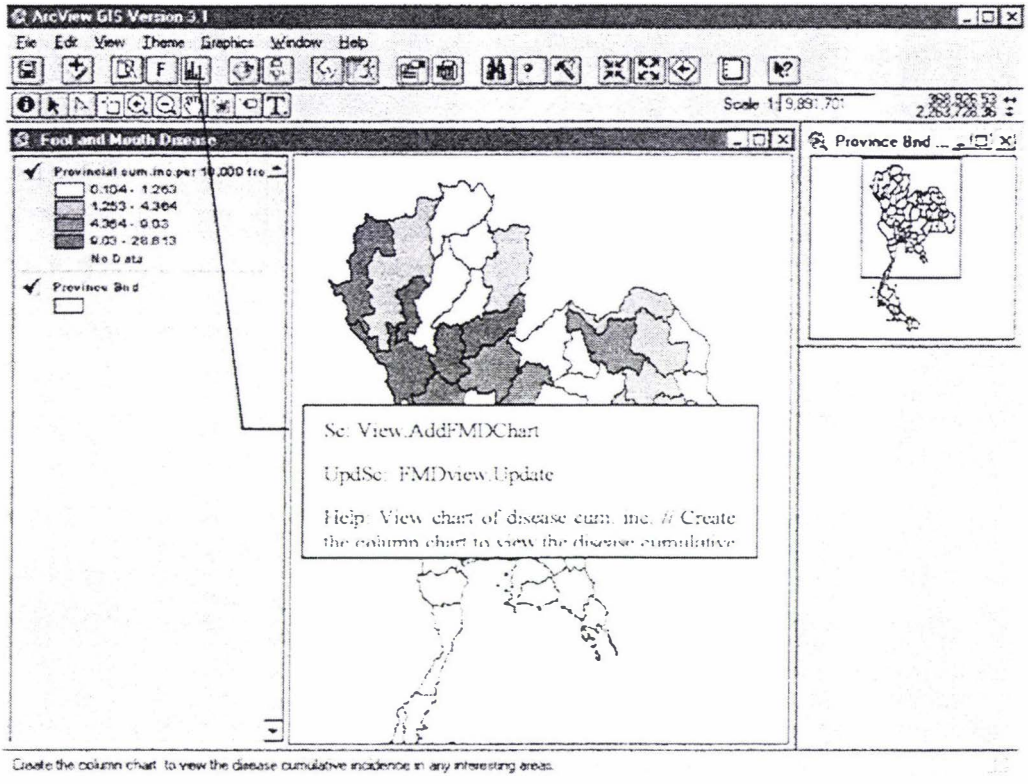
```
else
'If user choose the district level
' Add District boundary theme as a base map.
' Establish a data source and load it into the view.
aSrcName=SrcName.Make("E:/Maps/thai/thaiamphShape/amphur.shp")
aTheme=Theme.Make(aSrcName)
theView.AddTheme(aTheme)
aTheme.SetActive(true)
FMDTheme=aTheme.SetName("District cum.inc. per 10,000 ")
FMDTheme=theView.GetThemes.Get(0)
' Retrive the data base of FMD
' Verify that there is a FMD table in this project or not
thisProject=av.GetProject
FMDTable=thisProject.FindDoc("FMD")
if (FMDTable=nil) then
' Add table from SQL conection
mySQLConnection = SQLCon.Find ("oracle")
mySQLConnection = SQLCon.Find ("MS Access 97 Database")
 mySQLConnection.Login("Tippawon/Took")
 astring="Select * from FMD9597"
 theVTab=VTab.MakeSQL(mySQLConnection, astring)
 FMDTable=Table.Make(theVTab)
 FMDTable.SetName("FMD")
end
' Retrive the basic information of data table
FMDVTab=FMDTable.GetVTab
abitMap=FMDVTab.GetSelection
aYear = FMDVTab.FindField("Year")
'summarize FMD incidence in each province
'Let user choose the interested year
YearList={"1995","1996","1997","All"}
WhichYear=MsgBox.ChoiceAsString
(YearList," Which year do you want to know?"," Year of Database")
'stop the excute if user clicks on the cancel button.
if (nil=WhichYear) then
exit
end
aQStr= "[Year]>=1995"
'if user choose the year of 1995
if ( WhichYear="1995") then
   aQStr= "[Year]=1995"
 FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
  districtFld = FMDVTab.FindField("AmphurID")
  incFld     = FMDVTab.FindField("Incidence")
  amphFld    = FMDVTab.FindField("District")
  aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/DstIncF".AsFileName,dBASE,
districtFld, {amphFld,incFld},
         {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
  sumFMDTable=Table.make(aNewVtab)
  sumFMDTable.SetName ("sumDistrictIncFMD 1995")
FMDTheme=aTheme.SetName("District cum.inc. per 10,000 in 1995")
     elseif (WhichYear="1996")then
   aQStr= "[Year]=1996"
   FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
   districtFld = FMDVTab.FindField("AmphurID")
   incFld     = FMDVTab.FindField("Incidence")
   amphFld    = FMDVTab.FindField("District")
   aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/DstIncF".AsFileName, dBASE,
districtFld, {amphFld,incFld},        {#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
   sumFMDTable=Table.make(aNewVtab)
```

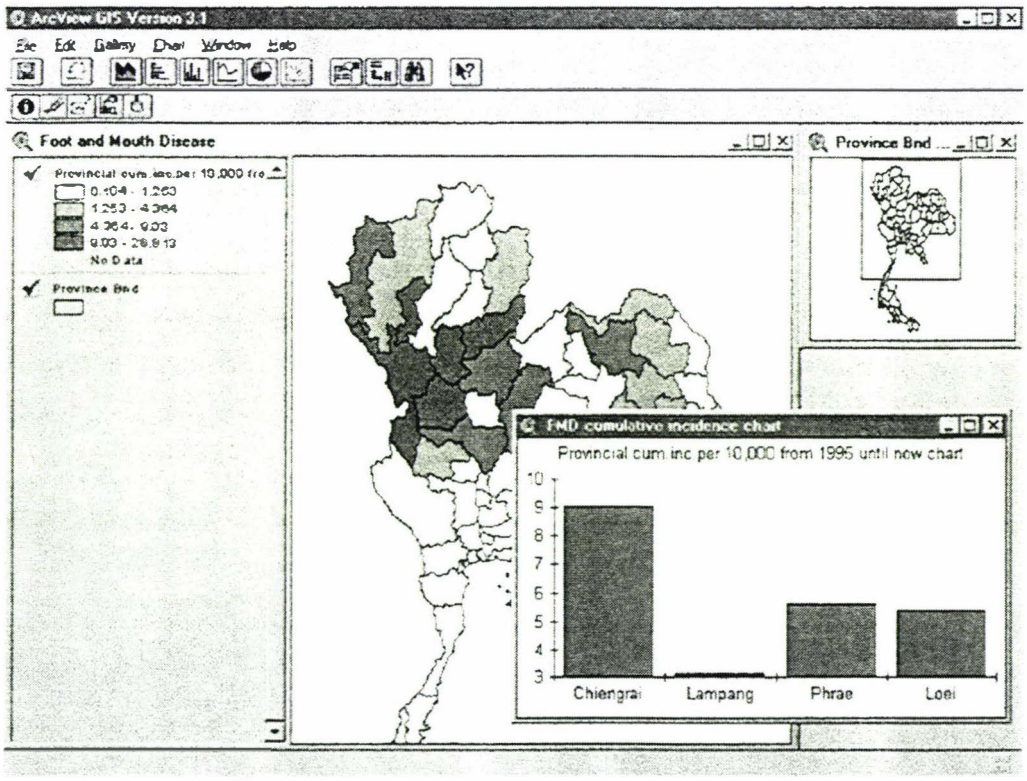
```
    sumFMDTable.SetName ("sumDistrictIncFMD 1996")
FMDTheme=aTheme.SetName("District cum.inc. per 10,000 in 1996")
        elseif (WhichYear="1997")then
        aQStr= "[Year]=1997"
       FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
       districtFld = FMDVTab.FindField("AmphurID")
       incFld     = FMDVTab.FindField("Incidence")
       amphFld    = FMDVTab.FindField("District")
       aNewVtab   = FMDVTab.Summarize("E:/maps/thai/summary data/DstIncF".AsFileName,
dBASE, districtFld, {amphFld,incFld},
{#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
       sumFMDTable=Table.make(aNewVtab)
       sumFMDTable.SetName ("sumDistrictIncFMD 1997")
 FMDTheme=aTheme.SetName("District cum.inc. per 10.000 in 1997")
else
         FMDVTab.Query( aQStr,abitmap,#VTAB_SELTYPE_NEW)
         districtFld = FMDVTab.FindField("AmphurID")
         incFld     = FMDVTab.FindField("Incidence")
         amphFld    = FMDVTab.FindField("District")
        aNewVtab = FMDVTab.Summarize("E:/maps/thai/summary data/DstIncF".AsFileName,
dBASE, districtFld, {amphFld,incFld},
{#VTAB_SUMMARY_FIRST,#VTAB_SUMMARY_SUM })
         sumFMDTable=Table.make(aNewVtab)
         sumFMDTable.SetName ("sumDistrictIncFMD")
FMDTheme=aTheme.SetName("District cum.inc. per 10,000 up to now")
end
FMDTheme=theView.GetThemes.Get(0)
'Join the theme with the livestock table data
 theVTab=sumFMDTable.GetVTab
'Set Alias for the district name field
districtNameFld=theVTab.FindField("First_District")
districtNameFld.SetAlias("Name")
AmpherJoinField=theVTab.FindField("AmphurID")
 FMDVTab=FMDTheme.GetFTab
 for each oneVTab in {FMDVTab}
  theJoinField = oneVTab.FindField("Amphur")
  oneVTab.Join (theJoinField.theVTab,AmpherJoinField)
  end
FMDTable=thisProject.FindDoc("FMD")
FMDVTab=FMDTable.GetVTab
 allSelected=FMDVTab.GetSelection
 allSelected.ClearAll
'Clear all selection from the FMD table
'Set theme's legend.
'Use  Natural for the legend type
FMDLegend=FMDTheme.GetLegend
FMDLegend.SetLegendType
(#LEGEND_TYPE_COLOR)
FMDLegend.Natural(FMDTheme,"Sum_Incidence",4)
theColorRamp = SymbolList.GetPreDefined(#SYMLIST_TYPE_COLORRAMP).Get(0)
FMDLegend.GetSymbols.RampSavedColors(theColorRamp)
'Load the symbol legend from legend file(FMD4Lev.avl)
aLegendFile = "E:/Maps/Legend/FMD4Lev.avl".AsFileName
FMDlegend.Load(aLegendFile,#LEGEND_LOADTYPE_SYMBOLS)
'Display the FMD theme
'and varify that the legend is visible.
FMDTheme.SetVisible(true)
FMDTheme.SetLegendVisible(true)
end
```

FMDView.Update

```
theView = av.GetActiveDoc.GetName
TheFMDView="Foot and Mouth Disease"
SELF.SetEnabled(theView=TheFMDView)
```

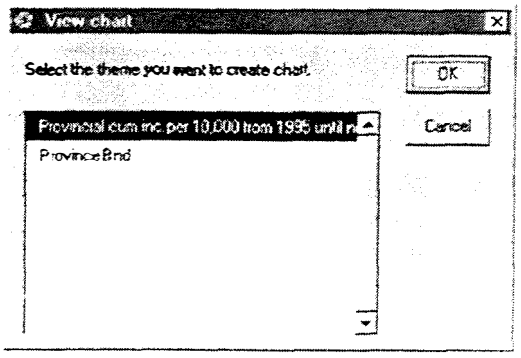
Interface 7 "View chart" button is used to display a column chart of cumulative incidence for interesting areas (associated script for interface control in callout square).

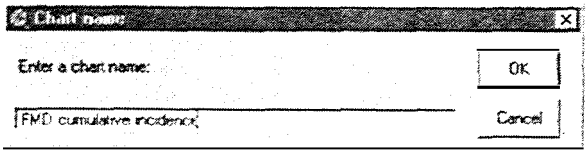




View.AddFMDChart

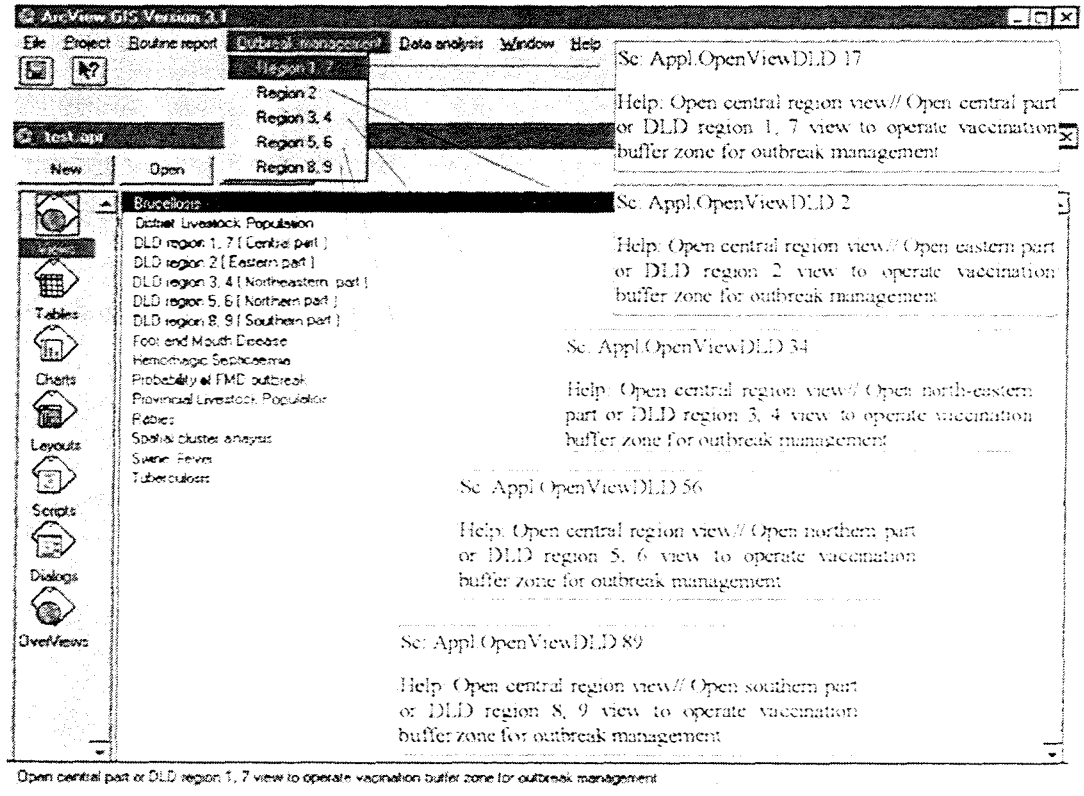
```
' Generates a column chart for the selected areas in active theme. A new Chart document is created
' to display cumulative incidence in those areas.
' Requires:  FMD theme in any level theme must be the active document.
' Retrieve the basic information.
theView = av.GetActiveDoc
ThemeList=TheView.GetThemes
FMDThemeList={}
For each t in ThemeList
 FMDThemeList.Add(t)
 end
 If (FMDThemeList.count=0) then
  MsgBox.error("no theme in the FMD
view","Error")
  exit
 else
  TheTheme=MsgBox.List(FMDThemeList, "Select the theme you want to create chart.","View chart")
 end
theFTab = theTheme.GetFTab
SelectedRecords=theFTab.GetSelection
if (selectedRecords.count=0) then
'no records were selected
'Therefore,no chart display
MsgBox.Error ("Please select the interested areas.","")
exit
end
' Retrieve the required fields for charts.
fieldList1 = { theFTab.FindField("Sum_Incide")}
if (fieldList1.Get(0) = nil) then
 MsgBox.Error ("Unable to find fields.","")
 exit
end
' Create the charts and set their properties.
columnChart = Chart.Make (theFTab, fieldList1)

columnChartWin = columnChart.GetWin
columnDisplay = columnChart.GetChartDisplay
columnDisplay.SetType (#CHARTDISPLAY_COLUMN)
columnDisplay.SetStyle (#CHARTDISPLAY_VIEW_SIDEBYSIDE)
NameChart=MsgBox.input("Enter a chart name: ","Chart name","")
columnChart.SetName (NameChart.AsString++"chart")
columnChart.SetSeriesFromRecords (True)
columnChart.GetChartLegend.SetVisible (False)
columnChart.GetYAxis.SetAxisVisible (true)
columnChart.GetYAxis.SetLabelVisible (False)
columnChart.GetYAxis.SetTickLabelsVisible (True)
columnChart.GetXAxis.SetLabelVisible (false)
columnChart.SetSeriesFromRecords(false)
nameField=theFTab.FindField("Name")
columnChart.SetRecordLabelField(nameField)
columnChart.GetXAxis.SetMajorGridSpacing (1)
columnChart.GetTitle.SetName (theTheme.GetName++"chart")
if (columnDisplay.IsOK.Not) then
  proceed = MsgBox.YesNo
  ("column chart may have an inconsistency"
  +NL+"Status:"++columnDisplay.GetStatus
  +NL+"Do you want to continue?", "", False)
  if (Not proceed) then
   exit
  end
end
```





```
'
' Add the charts to the project and open them.
thisProject=av.GetProject
thisProject.AddDoc (columnChart)
columnChartWin.open
```

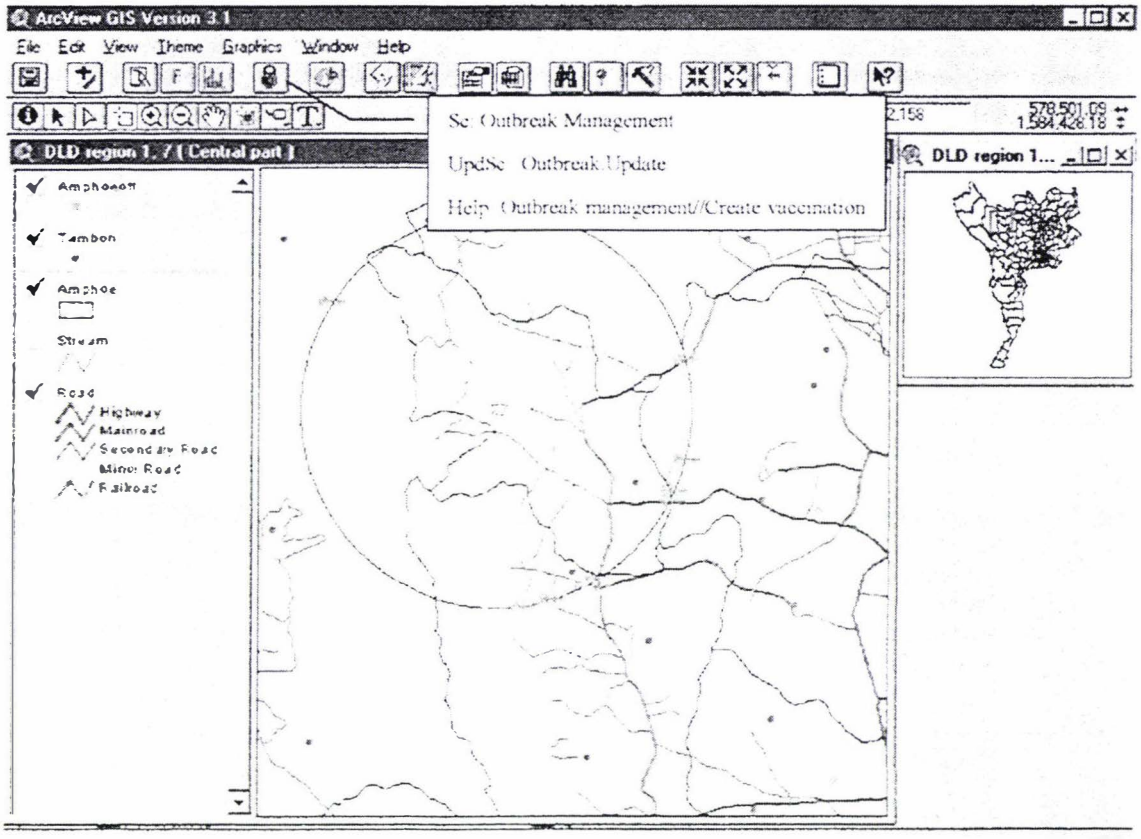
Interface 8 "Outbreak management menu" let user open DLD region view to operate vaccination buffer zone for outbreak management (associated script for each interface control in callout square).

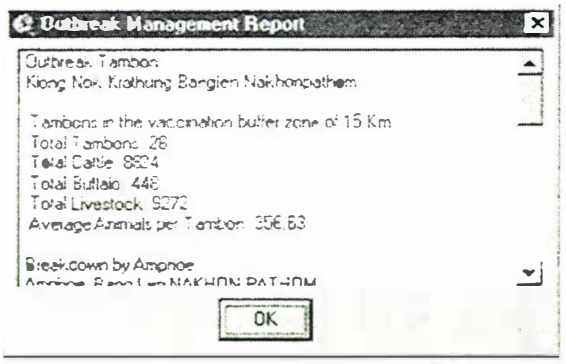


Appl.OpenViewDLD17

```
'This script opens view named DLD region 1,7(Central part) from the project menu. This script is
'executed from menu bar on the project's user interface.
'Required: Foot and Mouth Disease View Document from the defualt Project
thisProject=av.GetProject
theView=thisProject.FindDoc("DLD region 1, 7 ( Central part )")
theViewWindow=theView.GetWin
theViewWindow.open
```

Interface 9 "Outbreak management" button will be activated when one of DLD region view is opened. This operation allows user to choose the distance of vaccination buffer zone. This operation will draw the graphics and give message box with statistics on livestock numbers and District Livestock Office within buffer zone (associated script for interface control in callout square).





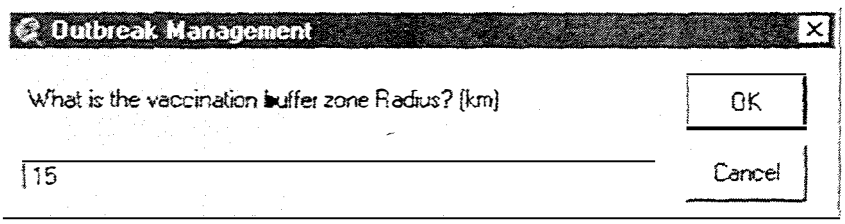
Outbreak.Update

```
theView = av.GetActiveDoc.GetName
OutbreakViewPattern= Pattern.Make ("DLD region*")
SELF.SetEnabled(theView=OutbreakViewPattern)
```

OutbreakManage

```
'* Author: Angus Cameron, 1996
'* Modified: Tippawon Teekayuwat, 1998
'Set up themes and field names
TheView = av.GetActiveDoc
theTheme = TheView.findTheme("Tambon")
AmphoeTheme = TheView.FindTheme("Amphoe")
AmphoeoffTheme = TheView.FindTheme("Amphoeoff")
RoadTheme = TheView.FindTheme("Road")
TmbNameField = TheTheme.GetFTab.FindField("Tambon_e")
```

```
TmbAmpField = TheTheme.GetFTab.FindField("District")
TmbCwatField = TheTheme.GetFTab.FindField("Province")
'Get selected Tambon
TS = theTheme.GetFTab.GetSelection
if (TS.count = 0) then
MsgBox.Error ("Please click on the tambon where the outbreak has occurred.","Outbreak management")
exit
elseif (TS.count >1) then
for each Tmbl in TS
if (TS.count > 1) then
TS.clear(Tmbl)
end
end
end
ShapeField = TheTheme.GetFTab.FindField("Shape")
OBTmbRec = TS.GetNextSet(-1)
OBTambon = TheTheme.GetFTab.ReturnValue(ShapeField,OBTmbRec)
OBTmbString = "Outbreak Tambon: "+nl+TheTheme.GetFTab.ReturnValueString(TmbNameField,
OBTmbRec)++","++
TheTheme.GetFTab.ReturnValueString(TmbAmpField,
OBTmbRec)++","++
TheTheme.GetFTab.ReturnValueString(TmbCwatField,
OBTmbRec)+nl+nl
TheTheme.GetFTab.SetSelection(TS)
TheTheme.GetFTab.UpdateSelection
'Get Selection Radius and select Tambons
Radius = ""
While (Radius.IsNumber.Not)
Radius = MsgBox.Input("What is the vaccination buffer zone radius? (km)",
"Outbreak Management", Radius)
end
if (radius = nil) then
exit
end
```

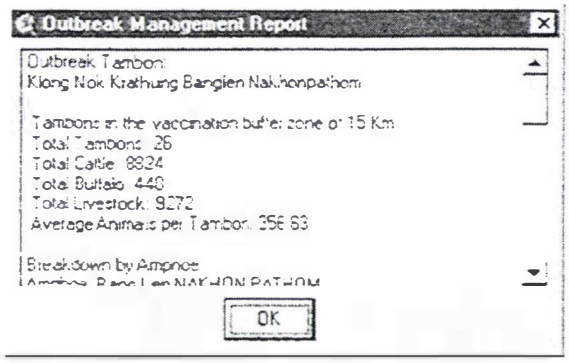


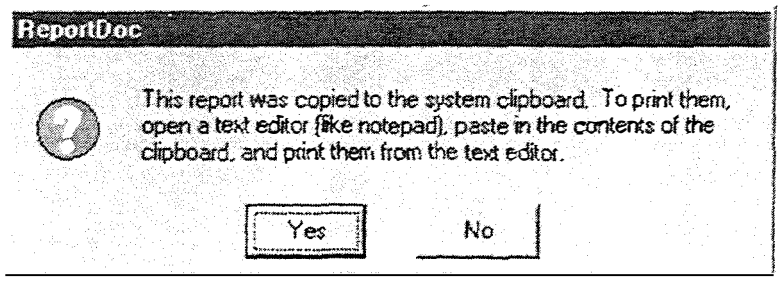

```
TheDistance = Units.Convert( Radius.AsNumber, #UNITS_LINEAR_KILOMETERS,
av.GetActiveDoc.GetDisplay.GetUnits )
TheTheme.SelectByTheme(TheTheme,#FTAB_RELTYPE_ISWITHINDISTANCEOF,TheDistance,
#VTAB_SELTYPE_NEW)
'Draw buffer zone and outbreak Tambon
myODB = ODB.Open("c:\ESRI\AV_GIS30\ARCVIEW\symbols\municip1.avp".asFilename)
MarkerList = myODB.Get(0)
TambonMarker = MarkerList.Get(7)
RoadBlockMarker = MarkerList.Get(47)
TheCircle = GraphicShape.Make(circle.Make(OBTambon, TheDistance))
TheTambon = GraphicShape.Make(OBTambon)
RoadBlockMarker.SetColor(Color.GetMagenta)
TambonMarker.SetColor(Color.GetBlue)
TheTambon.SetSymbol(TambonMarker)
TheView.GetGraphics.Add(TheCircle)
TheView.GetGraphics.Add(TheTambon)
TheView.GetDisplay.Flush
```

```
TS = TheTheme.GetFTab.GetSelection
TotalTmbls = TS.count
'Export Buffer Zone Tambon details to tabular report
CattleField = theTheme.GetFTab.FindField("Cattletmb")
BuffaloField = theTheme.GetFTab.FindField("Buffalotmb")
PreviousTable = av.GetProject.FindDoc("Vaccination Buffer Tambons")
If (PreviousTable <> nil) then
av.GetProject.RemoveDoc(PreviousTable)
end
ReportVTab = TheTheme.GetFTab.Export("E:\Maps\workDr\outbreak report.dbf".asFileName, dBASE,
true)
ReportTable =Table.Make(ReportVTab)
ReportTable.SetName("Vaccination Buffer Tambons")
'Calculate Total number of cattle and buffalo
Script.The.SetNumberFormat( "d" )
CattleSum = 0
BuffaloSum = 0
Counter=0
av.ShowMsg("Calculating totals...")
av.ShowStopButton
for each rec in TS
counter = counter + 1
Progress=100*(counter/TS.Count)
If (av.SetStatus(Progress).not) then
break
end
theValue1 = theTheme.GetfTab.ReturnValueNumber( CattleField, rec )
theValue2 = theTheme.GetFTab.ReturnValueNumber( BuffaloField, rec )
if ( not ( theValue1.IsNull ) ) then
CattleSum = theValue1 + CattleSum
end
if ( not ( theValue2.IsNull ) ) then
BuffaloSum = theValue2 + BuffaloSum
end
end
av.ClearStatus
av.ClearMsg
ReportString = "Tambons in the vaccination buffer zone of" ++Radius.AsString++ "Km."+nl+
" Total Tambons: "+TS.Count.AsString + nl +
" Total Cattle: "+CattleSum.AsString +nl +
" Total Buffalo: "+BuffaloSum.AsString+nl +
" Total Livestock: "+(BuffaloSum+CattleSum).AsString+nl+
" Average Animals per Tambon: "+
((BuffaloSum+CattleSum)/TS.Count).SetFormat("d.dd").AsString+nl
AmphoeList = nl+"Breakdown by Amphoe"+nl
'Find affected Amphoes
NameField = AmphoeTheme.GetFTab.FindField("Amphoe_E")
CwatField = AmphoeTheme.GetFTab.FindField("Province_E")
AmphoeoffField = AmphoeTheme.GetFTab.FindField("Amphoe_")
AmphoeoffShapeField = AmphoeoffTheme.GetFTab.FindField("Shape")
AmphoeTheme.SelectByTheme(TheTheme,#FTAB_RELTYPE_COMPLETELYCONTAINS,0,
#VTAB_SELTYPE_NEW)
AmphoeBitMap = AmphoeTheme.GetFTab.GetSelection
AmphoeShapeField = AmphoeTheme.GetFTab.FindField("Shape")
'Calculate Amphoe Breakdown
Counter=0
av.ShowMsg("Calculating Amphoe breakdown...")
av.ShowStopButton
for each amp in AmphoeBitMap
counter = counter + 1
```

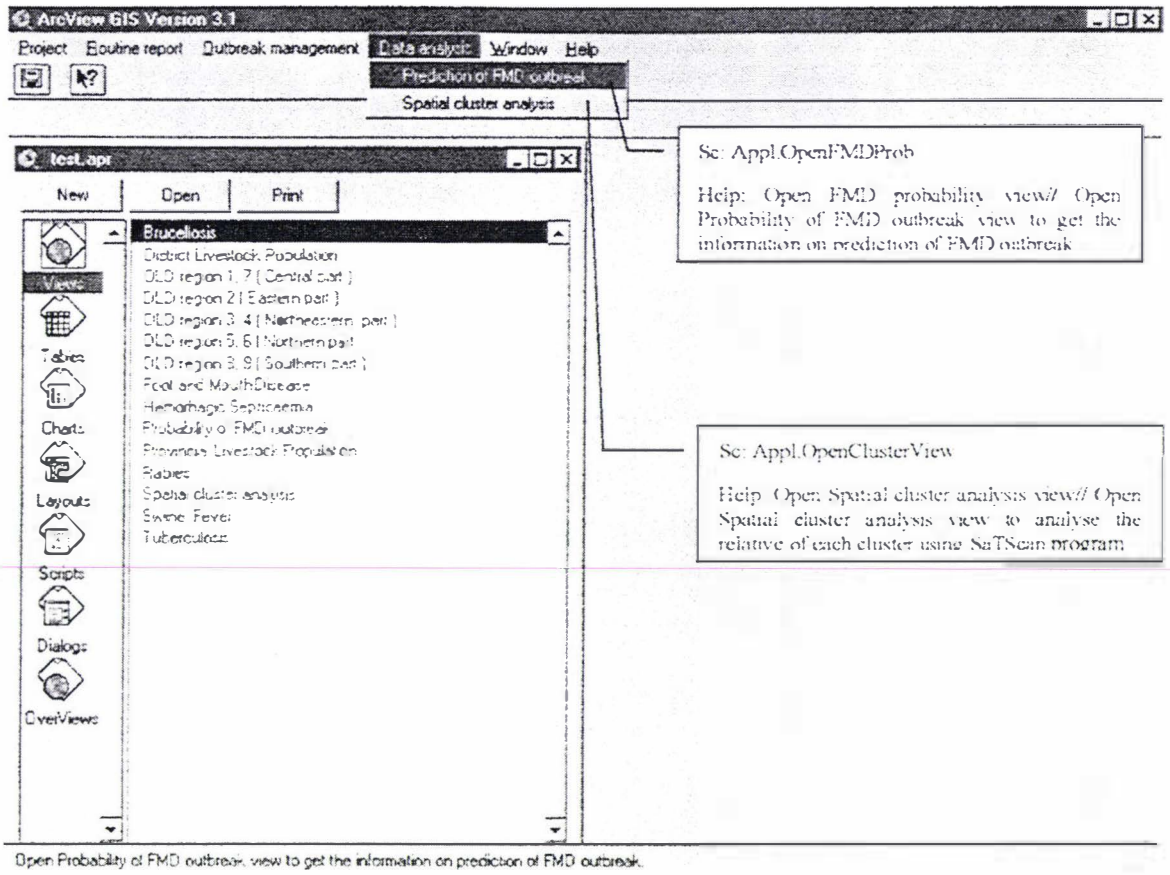
```
Progress=100*(counter/AmphoeBitMap.Count)
If (av.SetStatus(Progress).not) then
break
end
AmphoeoffRec = AmphoeTheme.GetFtab.ReturnValue(AmphoeoffField,amp)-2
AmphoeList = AmphoeList +nl+ "Amphoe: "+
AmphoeTheme.GetFtab.ReturnValueString(NameField,amp) + ", "
AmphoeList = AmphoeList + AmphoeTheme.GetFtab.ReturnValueString(CwatField,amp) + nl
Amphoeoff = AmphoeoffTheme.GetFTab.ReturnValue( AmphoeoffShapeField, AmphoeoffRec)
TheDistance = Units.convert(OBTambon.Distance(Amphoeoff),
av.GetActiveDoc.GetDisplay.GetUnits,
#UNITS_LINEAR_KILOMETERS)
AmphoeList = AmphoeList + " Distance from Amphoeoff to Tambon: "+TheDistance.AsString
++"km"+nl
AmphoePoly = AmphoeTheme.GetFTab.ReturnValue(AmphoeShapeField, Amp)
TheTheme.GetFTab.SetSelection(TS)
TheTheme.SelectByPolygon(AmphoePoly, #VTAB_SELTYPE_AND)
AmpTmbBM = TheTheme.GetFTab.GetSelection
AmphoeList = AmphoeList + " Number of Tambons in Amphoe: "+AmpTmbBM.Count.AsString+nl
CattleSum = 0
BuffaloSum = 0
for each rec in AmpTmbBM
theValue1 = theTheme.GetfTab.ReturnValueNumber( CattleField, rec )
theValue2 = theTheme.GetFTab.ReturnValueNumber( BuffaloField, rec )
if ( not ( theValue1.IsNull ) ) then
CattleSum = theValue1 + CattleSum
end
if ( not ( theValue2.IsNull ) ) then
BuffaloSum = theValue2 + BuffaloSum
end
end
AmphoeList = AmphoeList + " Total Livestock: "+(BuffaloSum+CattleSum).AsString+nl
end
av.ClearStatus
av.ClearMsg
TheTheme.GetFTab.SetSelection(TS)
'Calculate Road Block locations
RoadShapeField = RoadTheme.GetFTab.FindField("Shape")
PolyList = {theCircle.GetShape.AsMultiPoint.AsList}
CirclePoly = PolyLine.Make(PolyList)
BlockString = ""
BlockCount = 0
RoadTypeField = RoadTheme.GetFTab.FindField("Trans_typ")
RoadTheme.SelectByPolygon(theCircle.GetShape.AsPolygon, #VTAB_SELTYPE_NEW)
For each Rd in Roadtheme.GetFTab.GetSelection
RoadList = RoadTheme.GetFTab.ReturnValue(RoadShapeField, Rd).AsList
RoadType = RoadTheme.GetFTab.ReturnValue(RoadTypeField,Rd)
For Each RL in RoadList
For each Pt in 0..((Rl.count)-2)
LineSeg = Line.Make(RL.Get(Pt), RL.Get(Pt+1))
if (CirclePoly.Intersects(LineSeg)) then
BlockCount = BlockCount+1
BlockString = BlockString + " " + BlockCount.AsString +
". "+LineSeg.ReturnCenter.GetX.AsString+" "+LineSeg.ReturnCenter.GetY.AsString
if (RoadType = 1) then
typestring = "Highway"
elseif (RoadType = 2) then
typestring = "Major Road"
elseif (RoadType = 3) then
typestring = "Secondary Road"
```

```
 elseif (RoadType = 4) then
 typestring = "Secondary Road"
 elseif (RoadType = 8) then
 typestring = "Railroads"
 elseif (RoadType = 9) then
 typestring = "Future roads"
 else
 typestring = "Minor Road"
 end
 BlockString = BlockString + " ("+typestring+")"+nl
 TheRoadBlock = GraphicShape.Make(LineSeg.ReturnCenter)
 TheRoadBlock.SetSymbol(RoadBlockMarker)
 TheView.GetGraphics.Add(TheRoadBlock)
 end
 end
 end
 end
 RBString = "Road blocks for livestock movement control"+nl+
 " Up to "+blockcount.asstring++"road blocks will be necessary"+nl+
 " UTM Map Reference"+nl+BlockString
 'Tidy the screen
 sel = RoadTheme.GetFTab.GetSelection
 sel.ClearAll
 RoadTheme.GetFTab.UpdateSelection
 sel = AmphoeTheme.GetFtab.GetSelection
 sel.ClearAll
 AmphoeTheme.GetFTab.UpdateSelection
 TheView.GetDisplay.Flush
 'Indicate date of report
 today = Date.Now
 today.SetFormat( "d/MMM/yyyy, hh.m AMPM" )
 'Display summary report and Tambon listing
 '---------------------
 MsgBox Report( OBTmbString ++
 ReportString ++
 AmphoeList ++nl+
 RBString + nl +
 "-----------------------------------------------"+nl+
 "Reported outbreak management date" ++ today.AsString, "Outbreak Management Report")
 ReportDoc = OBTmbString ++
 ReportString ++
 AmphoeList ++nl+
 RBString + nl +""+ nl+
 "Reported outbreak management date " ++ today.AsString
   Clipboard.The.Empty
   Clipboard.The.Add(ReportDoc)
   Clipboard.The.Update
    If (MsgBox.YesNo("This report was copied to the system clipboard.  To print them, open a text
 editor"++
   "(like notepad), paste in the contents of the clipboard, and print them from the text
 editor.","ReportDoc", True) )
   then
   system.execute("notepad.exe")
 exit
 end
 ReportTable.GetWin.Open
```




ReportDoc
This report was copied to the system clipboard. To print them, open a text editor (like notepad), paste in the contents of the clipboard, and print them from the text editor.
Yes
No

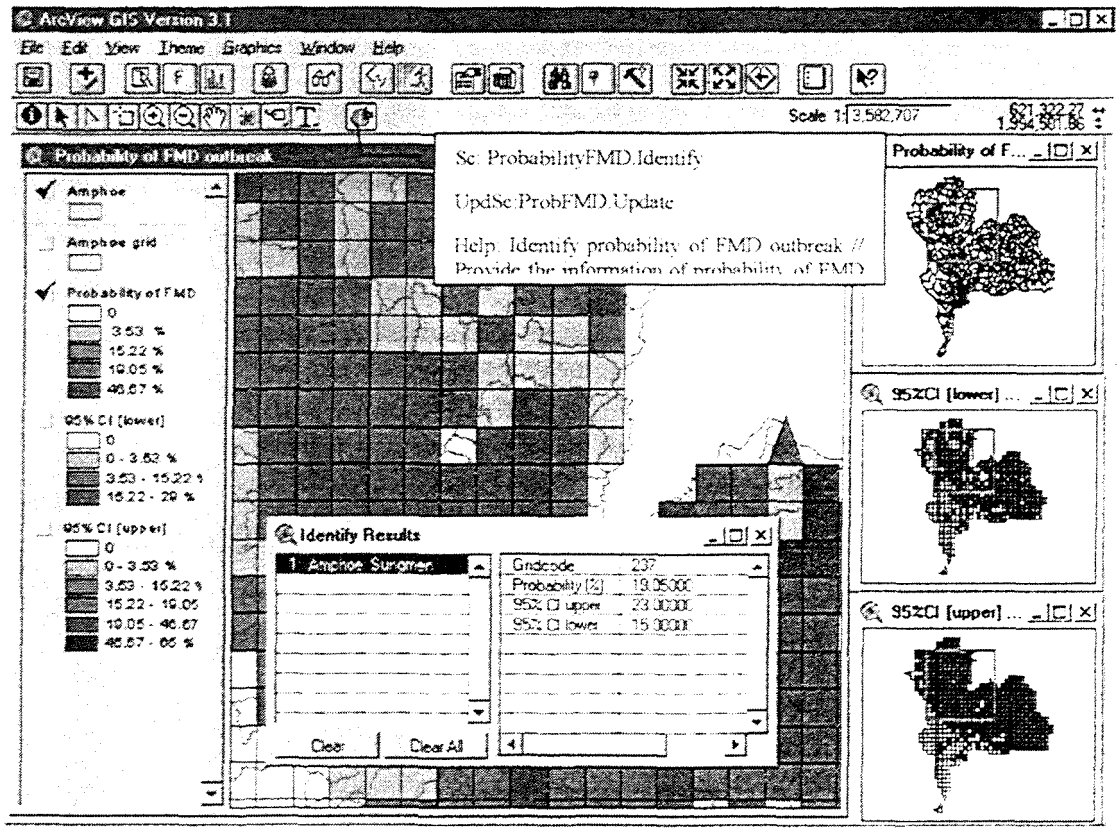
Interface 10 "Data analysis menu" let user open "prediction of FMD outbreak" view and "Spatial cluster analysis" view to analyse the disease data (associated scripts for each interface control in callout square).



Appl.OpenFMDProb

```
'This script opens view named Probability of FMD outbreak from the project menu. This script is
executed 'from menu bar on the project's user interface.
'Required:  Probability of FMD outbreak View Document from the defualt Project
theProject = av.GetActiveDoc
'Open theView named Probability of FMD outbreak
thisProject=av.GetProject
theView=thisProject.FindDoc("Probability of FMD outbreak")
theViewWindow=theView.GetWin
theViewWindow.open
```

Interface 11 Probability of FMD outbreak view comes with three overviews and works with two buttons which will be activated when this view is opened. "Identify probability of FMD outbreak" tool provides the information of Probability of FMD outbreak and 95% confidence interval from the result of decision tree model.

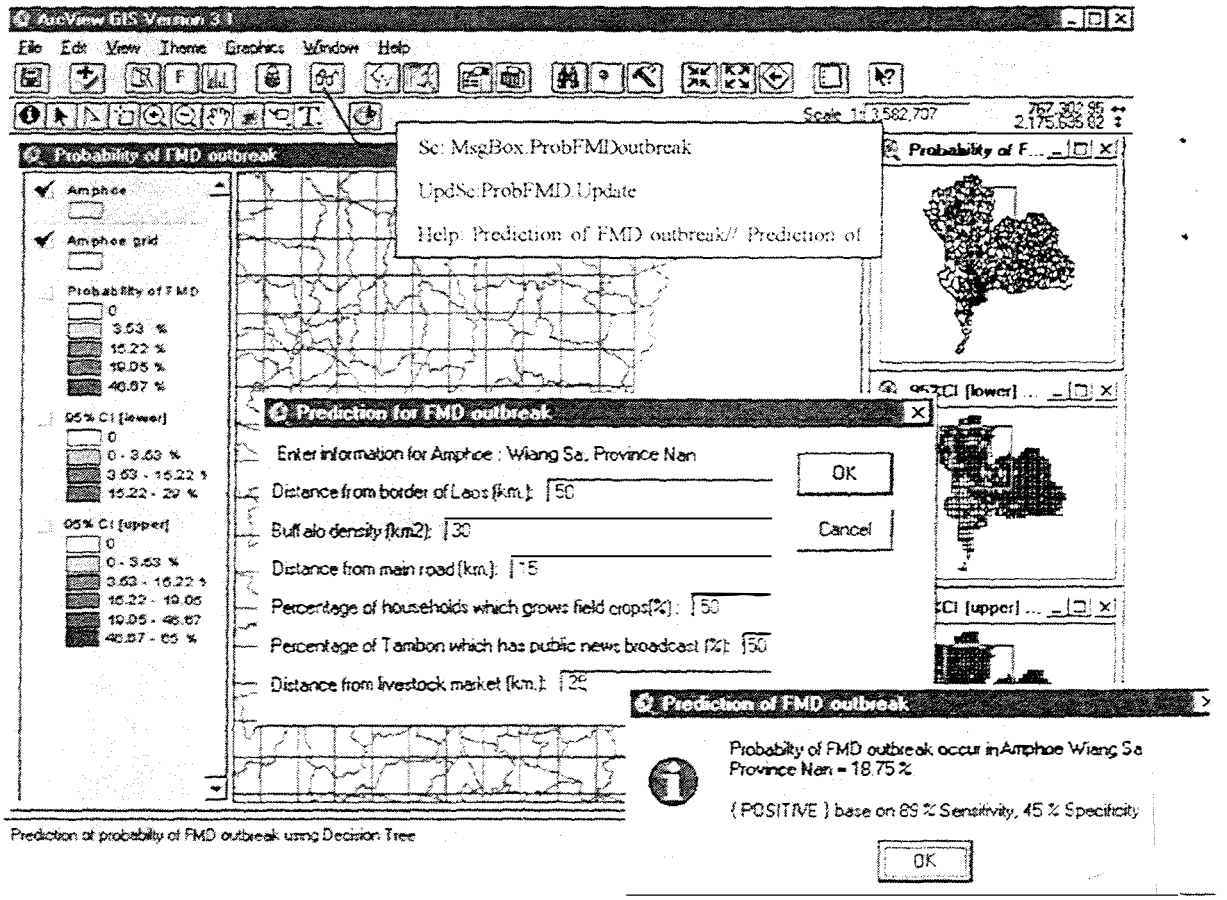


ProbabilityFMD.Identify

```
'Identify probability of FMD outbreak from the result of
'dicision tree no 5 using MultiInput.
'Detail for Tree 5 : Cost of false negative:false positive = 5:1
'                  : Priors type: Calculated from training sample
'                  : Adjusted Prior by cost: on
'                  : Target category: non outbreak = 0.5468, outbreak    = 0.4532
'                  : Resulting Tree: Total number of nodes = 13
'                                  : Total number of levels in tree = 3
'                                  : Total number of terminal nodes = 7
'                  : Resubstitution: Risk estimate = 0.35, : SE of risk estimate =0.02
'                  : Cross-validation: Risk estimate = 0.39, : SE of risk estimate =0.03
'                  : Sensitivity (95% CI): 0.89 (0.83-0.96)
'                  : Specificity (95% CI): 0.45 (0.41-0.49)
'Set up themes and field names
TheView = av.GetActiveDoc
ProbFMDTheme = TheView.FindTheme("Probability ofFMD outbreak")
ProbFld=ProbFMDTheme.GetFTab.FindField("Tree5")
Prob95UpFld=ProbFMDTheme.GetFTab.FindField("Tree5up")
Prob95LowFld=ProbFMDTheme.GetFTab.FindField("Tree5low")
```

```
AmphoeTheme = TheView.FindTheme("Amphoe")
AmphoeNameFld = AmphoeTheme.GetFTab.FindField("Amphoe_E")
ProvinceNameFld = AmphoeTheme.GetFTab.FindField("Province")
'Get selected Grid
p = theView.GetDisplay.ReturnUserPoint
ProbFMDTheme.SelectbyPoint(p,#VTAB_SELTYPE_NEW)
GS=ProbFMDTheme.GetFTab.GetSelection
AmphoeTheme.SelectByTheme(ProbFMDTheme,#FTAB_RELTYPE_CONTAINSTHECENTEROF,0,
#VTAB_SELTYPE_NEW)
AmphoeGridTheme = AmphoeTheme.GetFTab.GetSelection
AmphoeRec = AmphoeGridTheme.GetNextSet(-1)
AmphoeName = AmphoeTheme.GetFTab.ReturnValueString(AmphoeNameFld,AmphoeRec)
ProvinceName = AmphoeTheme.GetFTab.ReturnValueString(ProvinceNameFld,AmphoeRec)
Amp = AmphoeName.AsString
Pro=ProvinceName.AsString
if (ProbFMDTheme.CanFindByPoint) then
    keys = ProbFMDTheme.FindByPoint(p)
    for each key in keys
                found = TRUE
       idlabel = "Amphoe:"++Amp
       f=NIL
          if (ProbFMDTheme.CanLabel) then
           f=ProbFMDTheme.GetLabelField
          end
          if(f<>NIL)then
          s = ProbFMDTheme.ReturnValueString(f.GetName, key)
           idlabel = idlabel++s
          end
       ProbFMDTheme.Identify(key, idlabel)
    end
 end
```

Interface 12 "Prediction of FMD outbreak" button predicts the probability of foot and mouth outbreak for selected grid with the information put in the message box using the result of Decision tree analysis.



MsgBox.ProbFMDOutbreak

```
' Calculate probability ofFMD outbreak from the result of decision tree no 5 using MultiInput.
'Detail for Tree 5 : Cost of false negative:false positive = 5:1
'                  : Priors type: Calculated from training sample
'                  : Adjusted Prior by cost: on
'                  : Target category: non outbreak = 0.5468
                                      outbreak     = 0.4532
'                  : Resulting Tree: Total number of nodes = 13
'                                  : Total number of levels in tree = 3
'                                  : Total number of terminal nodes = 7
'                  : Resubstitution: Risk estimate = 0.35
'                                  : SE of risk estimate =0.02
'                  : Cross-validation: Risk estimate = 0.39
'                                  : SE of risk estimate =0.03
'                  : Sensitivity (95% CI): 0.89 (0.83-0.96)
'                  : Specificity (95% CI): 0.45 (0.41-0.49)
 'Set up themes and field names
 TheView = av.GetActiveDoc
GridTheme = TheView.FindTheme("Amphoe grid")
AmphoeTheme = TheView.FindTheme("Amphoe")
AmphoeNameFld = AmphoeTheme.GetFTab.FindField("Amphoe_E")
```

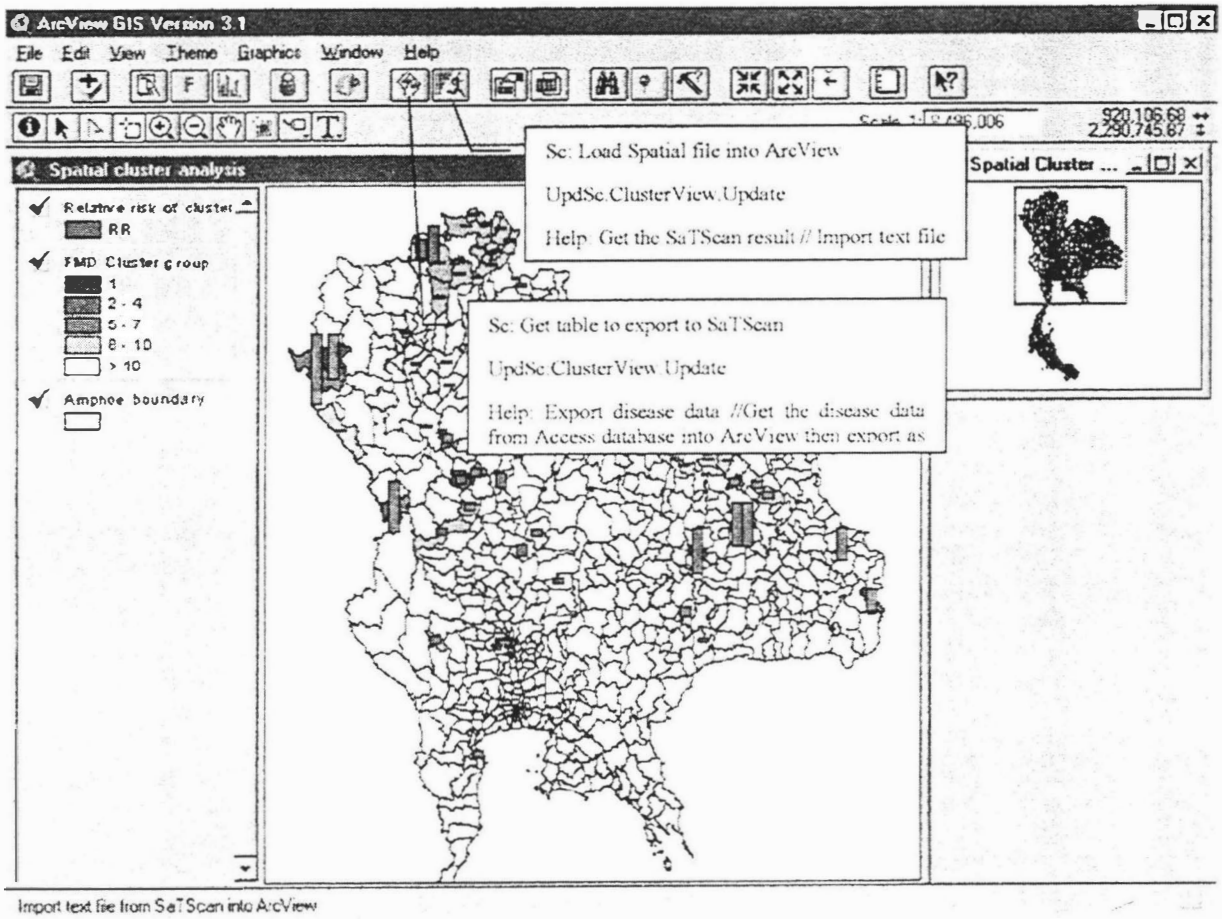
```
ProvinceNameFld = AmphoeTheme.GetFTab.FindField("Province")
'Get selected Tambon
GS = GridTheme.GetFTab.GetSelection
if (GS.count = 0) then
MsgBox.Error ("Please select Amphoe grid where you want to know the prediction.","Prediction for
FMD outbreak")
exit
end
for each g in GS
AmphoeTheme.SelectByTheme(GridTheme,#FTAB_RELTYPE_CONTAINSTHECENTEROF,0,
#VTAB_SELTYPE_NEW)
AmphoeGridTheme=AmphoeTheme.GetFTab.GetSelection
AmphoeRec= AmphoeGridTheme.GetNextSet(-1)
AmphoeName=AmphoeTheme.GetFTab.ReturnValueString(AmphoeNameFld,AmphoeRec)
ProvinceName=AmphoeTheme.GetFTab.ReturnValueString(ProvinceNameFld,AmphoeRec)

  AnswerBox = MsgBox.MultiInput("Enter information for Amphoe :"++ AmphoeName.AsString+",
Province"++ProvinceName.AsString,
     "Prediction for FMD outbreak",
     { "Distance from border of Laos (km.):", "Buffalo density (km2):", "Distance from main road
(km.):",
      "Percentage of households which grows field crops(%) :",
      "Percentage of Tambon which has public news broadcast (%):", "Distance from livestock market
(km.):"},
     { "", "", "", "", "", "", ""})
   'Check to see if the user clicked cancel to end input...
  if (AnswerBox.count < 1) then
exit
  end
if (AnswerBox.Get(0).AsNumber > 268) then
   if (AnswerBox.Get(1).AsNumber > 1.5) then
      if (AnswerBox.Get(2).AsNumber > 6) then
        MsgBox.Info("Probabilty of FMD outbreak occur in
Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 0 %", "Prediction of
FMD outbreak")
        exit
      Else
        MsgBox.Info("Probabilty of FMD outbreak occur in
Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 15.22 %."+nl+Nl+
                "{ NEGATIVE } base on 89 % Sensitivity, 45 % Specificity"+nl+
                "{ POSITIVE } with 97 % Sensitivity, 38 % Specificity" ,"Prediction of FMD outbreak")
        exit
      end
   else
     MsgBox.Info("Probabilty of FMD outbreak occur in
Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 0 %", "Prediction of
FMD outbreak")
     exit
   end
end
if (AnswerBox.Get(3).AsNumber <= 6) then
   if (AnswerBox.Get(4).AsNumber <= 63.5) then
     MsgBox.Info("Probabilty of FMD outbreak occur in
Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 46.67 %."+nl+nl+
            "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
     exit
   else
     MsgBox.Info("Probabilty of FMD outbreak occur in
Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 19.05 %."+nl+nl+
            "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
```

```
      exit
    end
  end
  if (AnswerBox.Get(5).AsNumber <= 30) then
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 18.75 %."+nl+nl+
              "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
      exit
    else
  exit

      end
    else
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 0 %", "Prediction of
  FMD outbreak")
      exit
    end
  end
  if (AnswerBox.Get(3).AsNumber <= 6) then
    if (AnswerBox.Get(4).AsNumber <= 63.5) then
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 46.67 %."+nl+nl+
              "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
      exit
    else
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 19.05 %."+nl+nl+
              "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
      exit
    end
  end
  if (AnswerBox.Get(5).AsNumber <= 30) then
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 18.75 %."+nl+nl+
              "{ POSITIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD outbreak")
      exit
    else
      MsgBox.Info("Probabilty of FMD outbreak occur in
  Amphoe"++AmphoeName.AsString+nl+"Province"++ProvinceName.AsString+" = 3.53 %."+nl+nl+
                "{ NEGATIVE } base on 89 % Sensitivity, 45 % Specificity","Prediction of FMD
  outbreak")
      exit
    end
  end
```

ProbFMD.Update

```
theView = av.GetActiveDoc.GetName
TheProbFMDView="Probability of FMD outbreak"
SELF.SetEnabled(theView=TheProbFMDView)
```

Interface 13 Spatial cluster analysis comes with two buttons with will be activated when this view is opened. "Export disease data" button use to export disease data from ArcView to analyse in SaTScan program then get the result back to ArcView to perform the relative risk and cluster group with "Get the SaTScan result" button (associated scripts for each interface control in callout square).



ClusterView.Update

```
theView = av.GetActiveDoc.GetName
TheClusterView="Spatial cluster analysis"
SELF.SetEnabled(theView=TheClusterView)
```

Get Table to Export to SaTScan

```
' Retrive the data base of FMD
' Verify that there is a FMD table in this project or not
thisProject=av.GetProject
FMDcaseTable=thisProject.FindDoc("FMDcase")
if (FMDcaseTable=nil) then
 ' Add table from SQL connection
' If Oracle server is not available, then
' display a list of available databases to
' the user for selection.
mySQLConnection = SQLCon.Find ("oracle")
mySQLConnection = SQLCon.Find ("MS Access 97 Database")
mySQLConnection.Login("Tippawon/Took")
 astring="Select`FMD9597`.`AmphurIDtxt`, `FMD9597`.`Infected animal`, `FMD9597`.`spicies` from
FMD9597"
```

```
 theVTab=VTab.MakeSQL(mySQLConnection, astring)
 FMDcaseTable=Table.Make(theVTab)
 FMDcaseTable.SetName("FMDcase")
end
'Retrieve the basic information of data table
FMDcaseVTab=FMDcaseTable.GetVTab
 TableExport    = FMDcaseVTab.Export("C:/SatScan/GIS/FMDCase.cas".asFileName, DText, false)
 system.execute("C:\SatScan\FMDspt.bat")
```

FMDspt.bat

```
'batch file use to execute the SaTScan program and manage format of text file from these two programs
c:
cd \
cd satscan
tail +2 GIS\FMDcase.cas | tr "," "\011" > GIS\FMDcase.cas
satscan.exe FMDspt.prm
cp header.txt GIS\FMDcase.txt
gawk -f awk.scr GIS\FMDSPT.GIS >> GIS\FMDcase.txt
exit
```

Load spatial file into ArcView

```
TxtFile= ("C:/SatScan/GIS/FMDCase.txt").asFileName
TempVTab=VTab.Make(TxtFile,false,False)
TempTable=Table.Make(TempVTab)
TempTable.SetName("TempTable")
 TempVTab.Export ( "C:/SatScan/GIS/FMDSpt.dbf".asFileName, dBASE, FALSE )
 'output file, newfile.dbf, and all records
TempTableDoc = av.GetProject.FindDoc("TempTable")
av.GetProject.RemoveDoc(TempTableDoc)
  PreviousTable = av.GetProject.FindDoc("SpatialFMD")
If (PreviousTable <> nil) then
av.GetProject.RemoveDoc(PreviousTable)
end
DbfFile=("C:/SatScan/GIS/FMDSpt.dbf").asFileName
ForWrite=true
SkipFirst=False
SptVTab=VTab.Make(DbfFile,ForWrite,SkipFirst)
SptTable=Table.Make (SptVTab)
SptTable.SetName("SpatialFMD")
AmpTxtField=Field.Make("AmphoeIDTxt",#FIELD_CHAR,8,0)
SptVTab.AddFields({AmpTxtField})
AmpIDField=SptVTab.FindField("AmphoeID")
For each r in SptVTab
SptVTab.SetValueString(AmpTxtField,r,SptVTab.ReturnValueNumber(AmpIDField,r).AsString)
end
SptVTab.RemoveFields({AmpIdField})
SptTable.StopEditing
'Retrieve the basic information for joining the result
'of spatial analysis file into FMD theme.
ClusterView = av.GetProject.FindDoc("Spatial Cluster analysis")
RRTheme = ClusterView.FindTheme("Relative risk of cluster")
RRVTab= RRTheme.GetFTab
RRVTab.UnjoinAll
ClusGrpTheme = ClusterView.FindTheme("FMD Cluster group")
ClusGrpVTab= ClusGrpTheme.GetFTab
ClusGrpVTab.UnJoinAll
SptFMD = av.GetProject.FindDoc("SpatialFMD")
```

```
SptFMDVTab = SptFMD.GetVTab
  'Join the theme with the SpatialFMD table data
 AmpherJoinField=SptFMDVTab.FindField("AmphoeIdTxt")
  for each oneVTab in {RRVTab,ClusGrpVTab}
   theJoinField = oneVTab.FindField("Amp_code")
   oneVTab.Join (theJoinField,SptFMDVTab,AmpherJoinField)
   end
   'Set each theme's legend.
'Use  chart for the first legend and
'graduate for the second
RRLegend=RRTheme.GetLegend
RRLegendFile="E:\Maps\Legend\RRChart.avl".AsFileName
RRLegend.Load(RRLegendFile,#LEGEND_LOADTYPE_ALL)
'Apply the legend changes
RRTheme.InvalidateLegend
RRTheme.SetVisible(True)
RRTheme.SetLegendVisible(True)
ClusterLegend= ClusGrpTheme.GetLegend
ClusterLegendFile="E:\Maps\Legend\clusterGrp.avl".AsFileName
ClusterLegend.Load(ClusterLegendFile,#LEGEND_LOADTYPE_ALL)
'Apply the legend changes
ClusGrpTheme.InvalidateLegend
ClusGrpTheme.SetVisible(True)
ClusGrpTheme.SetLegendVisible(True)
aViewDisplay=ClusterView.GetWin
```