

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.

**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<sup>®</sup>, the database management software Microsoft Access 97<sup>®</sup> and the spatial cluster analysis software SaTScan<sup>®</sup> version 2.1.3.

## II

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

<b>ABSTRACT</b>	<b>I</b>
<b>ACKNOWLEDGEMENTS</b>	<b>III</b>
<b>TABLE OF CONTENTS</b>	<b>IV</b>
<b>LIST OF TABLES</b>	<b>VIII</b>
<b>LIST OF FIGURES</b>	<b>X</b>
<b>CHAPTER 1. GENERAL INTRODUCTION</b>	<b>1</b>
Spatial Data Analysis	1
<i>Spatial data visualisation</i>	1
<i>Spatial data exploration</i>	2
<i>Spatial data modelling</i>	2
Application of Geographical Reporting and Analysis Systems in Animal Disease Control	2
Reference List	4
<b>CHAPTER 2. SPATIAL AND TEMPORAL ANALYSIS FOR DETECTION OF CLUSTERING OF DISEASE OCCURRENCE</b>	<b>7</b>
Introduction	7
<i>Disease clustering</i>	7
<i>Types of space -time patterns</i>	7
Units of analysis	8
Clustering in time	8
Clustering in space	9
Space-time clustering	11
Materials and Methods	12
<i>Methods for detection of clustering in time</i>	13
Scan statistic	14
Larsen's method	14
Empty cells method	14
<i>Methods for detection of clustering in space</i>	15
Moran's I with adjustment for population density (Ipop)	15
Spatial scan statistic	16
<i>Methods for detection of space-time clustering</i>	16
Knox's test	16
Mantel method	17
Space-time scan statistic	18

k-nearest neighbour method (k-NN) .....	18
Results .....	19
<i>Clustering in time</i> .....	19
<i>Clustering in space</i> .....	33
<i>Clustering in space-time</i> .....	37
Discussion .....	41
<i>Temporal clustering</i> .....	41
<i>Spatial clustering</i> .....	42
<i>Space-time clustering</i> .....	43
<i>Miscellaneous issues</i> .....	45
Conclusion .....	46
Reference List .....	47
<b>CHAPTER 3.                    DETECTION OF SPATIAL CLUSTERING OF</b>	
<b>ENZOOTIC BOVINE LEUCOSIS IN NEW ZEALAND.....</b>	<b>51</b>
Abstract .....	51
Introduction .....	51
Material and Methods .....	52
<i>Classification of EBL status</i> .....	52
<i>Data</i> .....	53
<i>Analytical methods</i> .....	55
Kernel smoothing .....	55
Ratio of kernels .....	56
Statistical spatial cluster analysis .....	57
<i>Software</i> .....	57
Results .....	60
<i>Visual analysis of the spatial pattern of herd density and EBL prevalence</i> .....	60
<i>Statistical analysis of spatial clustering of EBL positive herds</i> .....	60
Discussion .....	69
<i>Visualization of herd density and EBL prevalence using kernel density estimation</i> .....	69
<i>Statistical of spatial clustering of EBL positive herds</i> .....	70
Reference List .....	71
<b>CHAPTER 4.                    SPATIAL LOGISTIC REGRESSION AND</b>	
<b>CLASSIFICATION -TREE MODELS FOR THE DEVELOPMENT OF RISK</b>	
<b>MAPS OF FOOT-AND-MOUTH-DISEASE OUTBREAK OCCURRENCE IN</b>	
<b>THAILAND                    .....</b>	<b>72</b>
Abstract .....	72

Introduction .....	73
<i>Modelling of spatial data</i> .....	73
<i>General concepts of spatial data</i> .....	73
<i>Logistic regression modelling of spatial data</i> .....	75
<i>Classification tree modelling</i> .....	76
<i>Study aims</i> .....	77
Material and Methods .....	77
<i>Study area</i> .....	77
<i>Data layers</i> .....	77
Geographical data .....	77
FMD outbreak location data.....	78
<i>Statistical analysis</i> .....	81
Assessment of spatial clustering.....	81
Logistic regression analysis.....	81
Assessment of autocorrelation in the residuals .....	82
Classification tree analysis .....	83
Risk maps .....	84
Software.....	84
Results .....	85
<i>Univariate analysis</i> .....	85
<i>Logistic regression analysis</i> .....	86
<i>Classification tree analysis</i> .....	92
<i>ROC analysis</i> .....	96
<i>Risk maps</i> .....	97
Discussion.....	100
Conclusion .....	106
Reference List .....	107
<b>CHAPTER 5.                    DEVELOPMENT OF A SIMPLE GEOGRAPHICAL DISEASE REPORTING AND ANALYSIS SYSTEM FOR THAILAND.....</b>	<b>110</b>
Introduction .....	110
Review of GIS Components.....	111
Administrative Structure for Animal Disease Control and Eradication in Thailand	112
System Requirements.....	113
<i>Software</i> .....	113
GIS software.....	113



Database management software .....	114
Spatial cluster analysis software.....	114
<i>Hardware</i> .....	115
Data Requirements .....	115
<i>Spatial feature datasets</i> .....	115
<i>Attribute data</i> .....	116
Geographical attribute data .....	116
Animal health attribute data .....	116
Project Overview.....	117
<i>Routine report generator</i> .....	118
Livestock population .....	119
Disease occurrence .....	119
Outbreak management.....	121
<i>Data analysis</i> .....	122
Prediction of FMD outbreaks.....	122
Spatial cluster analysis .....	124
Discussion .....	125
Reference List .....	127
<b><i>GENERAL DISCUSSION AND CONCLUSIONS</i></b> .....	<b>130</b>
<b><i>APPENDIX</i></b> .....	<b>133</b>
<b><i>LIST OF INTERFACES AND ASSOCIATED SCRIPTS</i></b> .....	<b>134</b>

## List of Tables

Table 1 The four major components of an animal health information system.....	3
Table 2. Classification matrix of possible combinations of types of clustering of disease occurrence.....	8
Table 3. Results from analysis of temporal clustering using a single time series representing FMD outbreak occurrence <i>Type O</i> , <i>Type Asia1</i> , and <i>untyped</i> virus aggregated for the whole country. The results from Larsen's test for multiple time series analysis is presented by overall p-value .....	21
Table 4. Result of analysis of temporal clustering for the time series of individual provinces with Foot and Mouth Disease type O infection .....	22
Table 5. Results from analysis of temporal clustering of the time series of individual provinces with Foot and Mouth Disease type Asia1 infection.....	23
Table 6. Results from analysis of temporal clustering for the time series of individual provinces with Foot and Mouth Disease untyped virus infection.....	24
Table 7. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease type Asia1 infection.....	26
Table 8. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease type Asia1 infection.....	29
Table 9. Results from analysis of temporal clustering in the time series of individual districts with Foot and Mouth Disease untyped virus infection .....	31
Table 10. Results from analysis of spatial clustering using Moran's <i>I</i> adjusted for population density ( $I_{pop}$ ) using province as the unit of aggregation .....	33
Table 11. Results from analysis of spatial clustering with Moran's <i>I</i> adjusted for population density ( $I_{pop}$ ) using district as the unit of aggregation .....	33
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% .....	36
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 .....	38
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. ....	38
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.....	39

## List of Figures

Figure 1. Temporal pattern of FMD cases aggregated for the whole country by virus type between Jan 1995 and May 1997.....	21
Figure 2. Temporal pattern of FMD cases with virus type O in three provinces with unimodal clustering .....	25
Figure 3. Temporal pattern of FMD cases with virus type Asia1 in three provinces with unimodal clustering .....	25
Figure 4. Map of foot-and-mouth-disease cumulative incidence using province as the unit of aggregation.....	34
Figure 5. Map of foot-and-mouth-disease cumulative incidence using district as the unit of aggregation .....	35
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.....	54
Figure 7. Kernel estimation (from Bailey and Gatrell, 1995).....	56
Figure 8. Map of the location of the 5 study regions within New Zealand .....	59
Figure 9. Kernel density maps for all herds (left) and EBL positive herds (right) locations based on a bandwidth of 10 km .....	61
Figure 10. Kernel density maps for all herd (left) and EBL positive herd (right) locations based on a bandwidth of 15 km .....	62
Figure 11. Kernel density maps of all (left) and EBL positive (right) herd locations using a bandwidth of 20 km .....	63
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) .....	64
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) .....	65
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).....	66
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 .....	68
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.....	68
Figure 17. Map of administrative boundaries showing area included in this study .....	79
Figure 18. Raster maps for elevation and rainfall in the study area .....	81

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 .....	40
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).....	55
Table 18. Characteristics of the study areas used in the analysis. ....	57
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.....	67
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.....	67
Table 21. Characteristics of geographical and attribute data in the study area. ....	80
Table 22. Variables with statistically significant difference between outbreak and non-outbreak areas.....	85
Table 23. Odds ratios with 95% confidence limits (in brackets) of variables in logistic regression models <i>single cell (models 1, 2)</i> and <i>local region (model 2)</i> .....	88
Table 24: Odds ratios with 95% confidence limits (in brackets) for variables included in final logistic regression model <i>local region (model 2)</i> using proportion of cells with outbreaks in the local region (9 grid cells) as the response variable.....	88
Table 25. Summary table of deviance (G), Hosmer and Lemeshow's goodness of fit ( $\hat{C}$ ), and spatial autocorrelation of residuals (Moran's <i>I</i> ) for the four logistic models <i>single cell (models 1, 2)</i> and <i>local region (models 1, 2)</i> .....	89
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).....	93
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).....	94
Table 28. Node statistics for preferred classification tree <i>tree 5_1</i> .....	96
Table 29. Relationship between a GIS and a traditional animal health information system.....	111
Table 30. Spatial datasets used as part of this project.....	115

Figure 35. Map showing the output generated by SaTScan® and displayed using  
ArcView®. .... 124