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.

Statistical tools for spatio-temporal epidemiology, with application to veterinary diseases.

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Palmerston North, New Zealand.

Author:
Kate Richards
Supervisor:
Prof. Martin L. Hazelton (IFS)
Co-supervisor:
Prof. Mark Stevenson
(University of Melbourne)

December 14, 2015

Abstract

In epidemiology we are concerned with disease occurrence and its associated explanatory factors. Through analysis of the patterns in disease spread, in space and/or time, we are able to obtain information about possible risk factors and transmission mechanisms. The main focus in spatial epidemiology has been human health. However, economic costs and the concern about zoonoses has fuelled a growing field of veterinary epidemiology. Veterinary epidemiology has the added complication of the 'human effect'. For a disease to be recorded we require humans to detect and report the disease, and once reported human intervention is generally applied. This can lead to the true level of disease being under-represented with the loss of information impeding modelling and model predictions.

The reliability of statistical analyses depends on the quality of the underlying data. Anomalies could introduce significant bias and lead to inappropriate decision making. Residual analysis is often used to detect anomalous data, but with hierarchical models (common within epidemiology) the highly flexible representation of variation can mask outliers. We propose the use of exceedance probabilities as a tool for identifying and assessing anomalous data in spatio-temporal models for routinely collected areal disease count data. We illustrate this methodology through a case study on outbreaks of foot-and-mouth disease (FMD) in Viet Nam for the time period 2006 to 2008. The exceedance probabilities identify several provinces where the number of infected communes was unexpectedly low. These findings are particularly interesting as these provinces are located along major cattle movement pathways within Viet Nam.

With epidemic data, the primary interest is the understanding of the transmission of the disease and the effectiveness of intervention strategies. While epidemic curves provide an excellent representation of the temporal patterns, we propose the additional use of a new graphical tool, the 'cluster curve' to summarise the changes in spatial clustering through time. The cluster curve is based on the inhomogeneous K-function, and provides a means for summarizing the progression of clustering in infectious disease outbreak data taking into consideration spatial variation in the underlying population.

We look at the application of the cluster curve to two outbreaks of FMD in England (2001) and Japan (2010) and to the 2007 epidemic of Equine Influenza (EI) in Australia. By comparing our knowledge of the actual course of the outbreak with the insight provided by the cluster curve we are able to showcase the effectiveness of our tool. Throughout the progression of the outbreak several time windows obtained small sample sizes. Therefore, we also look at

the inclusion of significance indicators to definitively differentiate between true clustering and noise due to these small sample sizes.

The epidemic outbreaks studied all had intervention methods applied. The impact of intervention strategies was investigated through the simulation (via InterspreadPlus) of five intervention methods on outbreaks of FMD in two geographical regions. Using the cluster curve, we found that intervention methods that created buffer zones were found to have particular characteristics of spatial spread. We found that non-buffer methods were less effective in controlling local spread. This is most likely due to infection transmission prior to clinical signs. This kind of analysis demonstrates the practical importance of having effective tools for describing changes in the spatial patterns of disease during an epidemic outbreak.

Publications arising from this thesis

Richards, K.K., Hazelton, M.L., Stevenson, M.A., Lockhart, C.Y., Pinto, J. and Nguyen, L. (2014). Using exceedance probabilities to detect anomalies in routinely recorded animal health data, with particular reference to foot-and-mouth disease in Viet Nam. *Spatial and Spatio-temporal Epidemiology* **11**, 125-133.

The data in this thesis and the results we publish within contain privacy issues and are strictly confidential.

Acknowledgements

Firstly, I would like to give my greatest thanks to my primary supervisor Prof. Martin Hazelton. Without him pulling me into his office in my first month of being at Massey, Palmerston North and informing me about the honour track program, I would have never contemplated continuing my study after my undergraduate degree. I also can not thank him enough for his support throughout my postgraduate studies and being the best supervisor anyone could ever hope for.

Secondly, I would like to thank my co-supervisor Prof. Mark Stevenson for all his support with finding interesting datasets to base my PhD around and his work to get grants from the FAO to fund the primary years of my PhD.

I would like to send a large thank you to my family, who have been there though out this journey, with words of encouragement, being a sounding board and even a proof reader when required.

Also my greatest thanks and love goes out to my partner, Ian Donohue, who put up with years of long distance, and many months being ignored so I could be where I am today.

To all my friends I thankyou for all your support, specially Alice (PBF) and Rochelle for making my time away from home enjoyable.

To all the statistic postgraduates, this journey would of not have been the same without you. With a special thanks to all those who made the final stages of my PhD the most enjoyable time, even when stress levels were high.

Thanks must be given to Massey University and especially the Institute of Fundamental Sciences, for their support both financially and through the knowledge of all those involved.

I would also like to thank the FAO for the access to their data as well as there financial support. Also to the other Government agencies that have allowed me access to their data.

Finally, I would like to thank my heavenly father for his all the strength and comfort he provides.

List of Abbreviations

ANEMIS	Animal emergency management information systems
CPPP	Cluster Poisson point process
CSR	Complete spatial randomness
D	Depopulation
DED	Depopulation and pre-emptive culling (depopulation)
DIC	Deviance information criteria
DV	Depopulation and vaccination
EI	Equine influenza
EVI	Enhanced vegetation index
FAO	Food and Agriculture Organisation
FMD	Foot-and-mouth disease
HPP	Homogeneous Poisson point process
IP	Infected premise
IPP	Inhomogeneous Poisson point process
MCMC	Markov chain Monte Carlo
NC	No control
NMB	Non-movement ban
OIE	Office Internationale des Epizooties
PPP	Poisson point process
SIR	Standardized incidence ratio
SP	Suspect premise
SPR	Squared Pearson residual
V	Vaccination

Contents

1 Introduction **1**

2 Review of methods and models **6**

2.1 Data types 6

2.1.1 Case event data (point pattern) 6

2.1.2 Count data (areal) 8

2.2 Point process data 9

2.2.1 Spatial Point Process Models 10

2.2.2 Estimation of the intensity function for point processes 11

2.2.3 Second order properties 12

2.2.4 Methods of cluster detection 15

2.3 Areal data 20

2.3.1 Explanatory variables 21

2.3.2 Modelling 21

2.3.3 Model selection and assessment 24

2.3.4 Residual analysis 25

3 Disease epidemiology and datasets **28**

3.1 Foot and Mouth Disease 28
3.1.1 Control methods 29
3.1.2 Global impact 30
3.2 Viet Namese outbreak 30
3.2.1 Background 30
3.2.2 Description of data from 2006–2008 31
3.3 English outbreak 33
3.3.1 Background 33
3.3.2 Description of the data 34
3.3.3 Survey of models and analysis 35
3.4 Japan 37
3.4.1 Background 37
3.4.2 Description of the data 38
3.4.3 Survey of models and analysis 39
3.5 Interspread 40
3.5.1 Description of Interspread 40
3.5.2 Use of Interspread to simulate our FMD outbreaks 42

4 Exceedence probabilities for detecting anomalies in animal health data. 48

4.1 Introduction 48
4.2 Data 49
4.2.1 Viet Nam FMD endemic 49
4.2.2 Explanatory variables 51
4.3 Modelling the Data 53

4.3.1 Model Building 53
4.3.2 Preliminary Assessment of the Fitted Models 55
4.4 Exceedance Probabilities for Random Effects 58
4.4.1 Defining the Exceedance Probabilities 58
4.4.2 Application of Exceedance Probabilities to the Viet Nam Data 59
4.5 Discussion 60

5 Spatial clustering through time: an extension to the epidemic curve 63

5.1 Introduction 63
5.2 Models 64
5.2.1 Homogeneous Spatio-Temporal Models 65
5.2.2 Inhomogeneous Spatio-Temporal Models 65
5.2.3 Cluster Poisson point process 65
5.3 Implementation 68
5.3.1 Dataset specification 69
5.3.2 Epidemic curves 71
5.4 Cluster curve 78
5.4.1 Cluster curve smoothing 79
5.4.2 Implementation in R 81
5.4.3 Application of the Cluster Curve to simulated data 82
5.5 Integrated cluster curve 86
5.5.1 Application of the Integrated Cluster Curve to simulated data 87
5.6 Real world application 89
5.6.1 England 2001 FMD outbreak 89

5.6.2 Japan 2010 FMD outbreak . . . 93
5.7 Discussion . . . 98

6 Extensions to the cluster curve **99**

6.1 Introduction . . . 99
6.2 Highlighting significance in the cluster curve . . . 100
6.2.1 Methodology . . . 100
6.2.2 Application of cluster curve with significance marks to simulated data 101
6.2.3 Real World Application . . . 105
6.3 Shiny application of the cluster curve . . . 110
6.3.1 Implementation in R-studio . . . 110
6.3.2 Cluster curve online prototypes . . . 110
6.4 Discussion . . . 114

7 Application to FMD intervention strategies **115**

7.1 Introduction . . . 115
7.2 Simulated datasets . . . 116
7.2.1 Scenarios . . . 116
7.2.2 Data generation . . . 118
7.3 Border counties, Great Britain . . . 118
7.3.1 Spatial-temporal plots . . . 119
7.3.2 Epidemic curves . . . 120
7.3.3 Cluster curve . . . 120
7.3.4 Integrated Cluster Curve . . . 122

7.4 Miyazaki, Japan 136
7.4.1 Spatial-temporal plots 137
7.4.2 Epidemic curves 137
7.4.3 Cluster curve 138
7.4.4 Integrated Cluster Curve 138
7.5 Discussion 151

8 Application of the Cluster curve to Equine influenza 153

8.1 Introduction 153
8.2 Disease epidemiology 154
8.3 Australian epidemic history 154
8.4 Literature review 155
8.5 Data 156
8.5.1 Infected premises data 156
8.5.2 Horse population data 157
8.6 Application of the Cluster curve 157
8.7 Discussion 164

9 Conclusion and discussion 165

List of Figures

2.1 Example of case event data. 7

2.2 Example of count data. 8

3.1 Spatial aggregation of Viet Nam 32

3.2 FMD in cattle in Viet Nam 33

3.3 Point process plot of farms in the UK epidemic of FMD 35

3.4 (a) Map of Japan showing prefectures with the Miyazaki prefecture outlined in red (Wikimedia.org, 2015). (b) Miyazaki prefecture showing regions with our spatial window outlined in red (blogspot.co.nz, 2010). 39

3.5 Miyazaki 2010 FMD outbreak spatial distribution 40

3.6 Overall simulation flowchart in InterspreadPlus Stevenson et al. (2013) 41

3.7 Map of Great Britian showing counties (left) and the four selected counties (right) (Baxter, 2014). 43

3.8 Border counties spatial window 43

3.9 Image plots of simulated FMD outbreaks in Border counties : (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling. Grey represents farms with no FMD and no intervention, blue where no FMD but vaccination applied, green where no FMD but depopulation occurred and red where FMD is present. 45

3.10 Miyazaki spatial window 46

3.11 Image plots of simulated FMD outbreaks in Miyazaki (Grey all farms, Red FMD positive farms) : (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling Grey represents farms with no FMD and no intervention, blue where no FMD but vaccination applied, green where no FMD but depopulation occurred and red where FMD is present. 47

4.1 Map of Viet Nam showing provincial (left) and commune boundaries (right). The location of Thai Nguyen province is indicated by the box on the provincial map. Commune boundaries for Thai Nguyen province are shown on the right. 50

4.2 Map of Viet Nam showing the point location of commune-level FMD outbreaks, for: (a) 2006, (b) 2007, and (c) 2008. 51

4.3 Image plots of Viet Nam showing: (a) elevation (expressed in metres) and (b) Enhanced Vegetation Index for January 2006. 52

4.4 Square Pearson residual plots: (a) 2006, (b) 2007, (c) 2008. 57

4.5 Exceedance probabilities 5% level 59

4.6 Exceedance probabilities with serotype 60

4.7 FAO movement pathways 61

5.1 Inhomogeneous intensity function 66

5.2 Illustration of a cluster point process 67

5.3 Parent/children process with temporal component 68

5.4 Study region for artificial datasets 69

5.5 Spatial-temporal distribution of homogeneous PPP dataset every 10 days, with red representing new cases on the day and grey previous cases. The larger the plotting size the more recent the infection. 73

5.6 Spatial-temporal distribution of homogeneous CPPP dataset every 10 days, with red representing new cases on the day and grey previous cases. The larger the plotting size the more recent the infection. 74

5.7 Spatial-temporal distribution of inhomogeneous PPP dataset every 10 days, with red representing new cases on the day and grey previous cases. The larger the plotting size the more recent the infection. 75

5.8 Spatial-temporal distribution of inhomogeneous CPPP dataset every 10 days, with red representing new cases on the day and grey previous cases. The larger the plotting size the more recent the infection. 76

5.9 Epidemic curves of homogeneous datasets 77

5.10 Epidemic curve of inhomogeneous datasets 77

5.11 Cluster curve creation example 80

5.12 Raw cluster curve examples 81

5.13 Cluster curve slider example 82

5.14 Cluster curve for spatially homogeneous datasets 84

5.15 Cluster curve for spatially inhomogeneous datasets 85

5.16 Integrated cluster curve creation example 87

5.17 Integrated cluster curve for spatially homogeneous datasets 88

5.18 Integrated cluster curve for spatially inhomogeneous datasets 89

5.19 Spatial distribution of England 2001 FMD cases 90

5.20 Farm density in Northern England 90

5.21 Epidemic curve of Northern England 2001 91

5.22 Cluster curve for FMD outbreak in England 2001 2km to 10km radius 92

5.23 Integrated clustering curve for FMD outbreak in England 2001 0:5km and 0:10km radius 93

5.24 Spatial distribution of Japan 2010 FMD outbreak 94

5.25 Farm density in Miyazaki, Japan 95

5.26 Epidemic curve of Miyazaki, Japan 2010 95

5.27 Clustering curve for FMD outbreak in Miyazaki, Japan 2010 2km to 10km radius 96

5.28 Integrated clustering curve for FMD outbreak in Miyazaki, Japan 2010 0:5km and 0:10km radius 97

6.1 Cluster curve inhomogeneous PPP 'bump' example 100

6.2 Inhomogeneous K-function example 101

6.3 Inhomogeneous K-function envelope example 102

6.4 Cluster curve with significant dots for spatially homogeneous CPPP dataset . 103

6.5 Cluster curve with significant dots for spatially inhomogeneous CPPP dataset 104

6.6 Cluster curve with significant dots for FMD outbreak in England 2001 2km to 10km 106

6.7 Epidemic curve and cluster curve for 2km radius of the English FMD outbreak 107

6.8 Cluster curve with significant dots for FMD outbreak in Miyazaki, Japan 2010 2km to 10km 108

6.9 Epidemic curve and cluster curve for 2km radius of the Japanese FMD outbreak 109

6.10 Cluster curve example 111

6.11 Cluster curve prototype 113

7.1 Vaccination buffer ring example 117

7.2 Pre-emptive culling example 117

7.3 Border counties farm density 119

7.4 Image plots of simulated FMD outbreaks in Border counties: (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling. Grey represents farms with no FMD and no intervention, blue where no FMD but vaccination applied, green where no FMD but depopulation occurred and red where FMD is present. 123

7.5 Spatial-temporal 10 day window plots of simulated No control FMD outbreak in Border counties: (a) Day 10 (b) Day 20 (c) Day 30 (d) Day 40 (e) Day 50 (f) Day 60. Red represents the farms diagnosed with the presence of FMD has been diagnosed on the day, Grey represents the previously diagnosed farms. The plotting size represents the temporal scale, with the larger the point the more recently it was diagnosed. 124

7.6 Spatial-temporal 10 day window plots of simulated depopulated FMD outbreak in Border counties: (a) Day 10 (b) Day 20 (c) Day 30 (d) Day 31. Red represents the farms diagnosed with the presence of FMD has been diagnosed on the day, Grey represents the previously diagnosed farms. The plotting size represents the temporal scale, with the larger the point the more recently it was diagnosed. 125

7.7 Spatial-temporal 10 day window plots of simulated vaccinated FMD outbreak in Border counties: (a) Day 10 (b) Day 20 (c) Day 26. Red represents the farms diagnosed with the presence of FMD has been diagnosed on the day, Grey represents the previously diagnosed farms. The plotting size represents the temporal scale, with the larger the point the more recently it was diagnosed.126

7.8 Spatial-temporal 10 day window plots of simulated depopulated and vaccinated FMD outbreak in Border counties: (a) Day 10 (b) Day 20 (c) Day 22. Red represents the farms diagnosed with the presence of FMD has been diagnosed on the day, Grey represents the previously diagnosed farms. The plotting size represents the temporal scale, with the larger the point the more recently it was diagnosed. 127

7.9 Spatial-temporal 10 day window plots of simulated depopulated and pre-emptive culled FMD outbreak in Border counties: (a) Day 10 (b) Day 20 (c) Day 24. Red represents the farms diagnosed with the presence of FMD has been diagnosed on the day, Grey represents the previously diagnosed farms. The plotting size represents the temporal scale, with the larger the point the more recently it was diagnosed. 128

7.10 Epidemic curve of Border counties intervention datasets 129

7.11 Cluster curve for all intervention strategies at a radius of 2km on a comparable scale. 129

7.12 Cluster curve for the Border counties FMD simulated outbreak with no control methods 130

7.13 Cluster curve for the Border counties FMD simulated outbreak with depopulation control methods 131

7.14 Cluster curve for the Border counties FMD simulated outbreak with vaccination control methods 132

7.15 Cluster curve for the Border counties FMD simulated outbreak with depopulation and vaccination control methods 133

7.16 Cluster curve for the Border counties FMD simulated outbreak with depopulation and pre-emptive culling control methods 134

7.17 Integrated cluster curve for simulated FMD outbreaks in Border counties over a radius of 0:5km : (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling. . . . 135

7.18 Farm density in Miyazaki, Japan 136

7.19 Image plots of simulated FMD outbreaks in Miyazaki (Grey all farms, Red FMD positive farms) : (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling. Grey represents farms with no FMD and no intervention, blue where no FMD but vaccination applied, green where no FMD but depopulation occurred and red where FMD is present. 140

7.20 Spatial-temporal 10 day window plots of simulated No control FMD outbreak in Miyazaki : (a) Day 10 (b) Day 20 (c) Day 30 (d) Day 40 (e) Day 50 (f) Day 60. Red represents the FMD positive farms recorded on the day, grey represents the previously diagnosed farms. The plotting size representing the temporal scale, with the larger the point the more recently it was diagnosed. 141

7.21 Spatial-temporal 10 day window plots of simulated depopulated FMD outbreak in Miyazaki : (a) Day 10 (b) Day 20 (c) Day 30 (d) Day 40 (e) Day 50 (f) Day 60. Red represents the FMD positive farms recorded on the day, grey represents the previously diagnosed farms. The plotting size representing the temporal scale, with the larger the point the more recently it was diagnosed. 142

7.22 Spatial-temporal 10 day window plots of simulated vaccinated FMD outbreak in Miyazaki : (a) Day 10 (b) Day 20 (c) Day 30 (d) Day 37. Red represents the FMD positive farms recorded on the day, grey represents the previously diagnosed farms. The plotting size representing the temporal scale, with the larger the point the more recently it was diagnosed. 143

7.23 Spatial-temporal 10 day window plots of simulated depopulated and vaccinated FMD outbreak in Miyazaki : (a) Day 10 (b) Day 20 (c) Day 30 (e) Day 36. Red represents the FMD positive farms recorded on the day, grey represents the previously diagnosed farms. The plotting size representing the temporal scale, with the larger the point the more recently it was diagnosed. 144

7.24 Spatial-temporal 10 day window plots of simulated depopulated and pre-emptive culled FMD outbreak in Miyazaki : (a) Day 10 (b) Day 19. Red represents the FMD positive farms recorded on the day, grey represents the previously diagnosed farms. The plotting size representing the temporal scale, with the larger the point the more recently it was diagnosed. 145

7.25 Epidemic curve of Miyazaki intervention datasets 145

7.26 Cluster curve for all intervention strategies at a radius of 1km for Miyazaki FMD outbreaks on a comparable scale. redoing 146

7.27 Cluster curve for the Miyazaki FMD simulated outbreak with no control methods147

7.28 Cluster curve for the Miyazaki FMD simulated outbreak with depopulation control methods 148

7.29 Cluster curve for the Miyazaki FMD simulated outbreak with vaccination control methods 148

7.30 Cluster curve for the Miyazaki FMD simulated outbreak with depopulation and vaccination control methods 149

7.31 Cluster curve for the Miyazaki FMD simulated outbreak with depopulation and pre-emptive culling control methods 149

7.32 Integrated cluster curve for simulated FMD outbreaks in Miyazaki over a radius of 0:5km : (a) No intervention (b) depopulation (c) vaccination (d) depopulation and vaccination (e) depopulation and pre-emptive culling. 150

7.33 Histogram of the distribution of pairwise distances between farm locations : (a)Border counties, Great Britain (b) Miyazaki, Japan. 152

8.1 Map showing the spatial plot of EI cases. Grey indicates the location of all susceptible farms within NSW and red all susceptible, infected and resolved EI farms 158

8.2 Estimated density for horse farms 158

8.3 Spatial plots of the outbreak of EI, Australia 2007, for 20 day windows. Red represents new cases recorded on the day and grey the previous cases. The plotting size indicates how recently the case occurred, with large symbols indicating very recent events and small symbols earlier ones. 160

8.4 Spatial plots of the outbreak of EI, Australia 2007, for 20 day windows. Red represents new cases recorded on the day and grey the previous cases. The plotting size indicates how recently the case occurred, with large symbols indicating very recent events and small symbols earlier ones. 161

8.5 Epidemic curve of EI Australia 2007 162

8.6 Cluster curve for EI outbreak at 1:5km at 1km intervals 162

8.7 Epidemic curve and cluster curve for 2km radius 163