Risk factors for foot-and-mouth disease in Zambia 1981 to 2012

A thesis presented

in partial fullfilment of the requirements

for the degree of Master of Veterinary Studies

at Massey University

Raymond Hamoonga

Institute of Veterinary, Animal and Biomedical Sciences

Massey University

Palmerston North, New Zealand

2014

(Submitted 31 March 2014)

Institute of Veterinary, Animal and Biomedical Sciences

Massey University

Palmerston North, New Zealand

2014

Dedicated to my best friend and wife Belinda for putting up with me during this period, and to Nathaniel and Hannah - I love and cherish you all.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Mark Stevenson. This work would not have eventuated without your support and guidance. It has been an absolute honour to pick your masterful brain. Apart from mentoring me to be an epidemiologist, your exemplary life of humility, patience and hard work, will forever impact my life.

I would like to extend my sincere gratitude to the lecturers both present and past, fellow students and staff at the EpiCentre, Massey University, for helping me during the course of my studies. My profound appreciation goes to Tim Carpenter, Cord Heuer, Naomi Cogger, Daan Vink, Jackie Benschop, Sara Rosanowski, Eric Neumann, Roger Morris, Simon Verschaffelt, Wendy Maharey and Christine Cunningham. I would also like to thank Nigel French and Deborah Prattley for teaching me how to write scientifically during the early stages of my studies. I am forever indebted to you for teaching me how to read and think critically.

I would also like to thank my co-authors on the paper presented in Chapter 3 Alberto Allepuz, Tim Carpenter and Yona Sinkala. I am grateful to Barry Robinson at the Department of Agriculture, Fisheries and Forestry, Australia and to Martin Hazelton at the Institute of Fundamental Sciences, Massey University for helpful comments and suggestions.

My studies were fully supported by the New Zealand Aid programme (NZAID). I am grateful to the New Zealand Government for providing financial assistance and to the people of New Zealand who have assisted and welcomed myself and my family during each step of our journey. I am grateful to Silvia Hooker, Leuaina Vaai-Hatier, Jamie Hooper, Susan Flyn and NZAID staff for ensuring my stay and studies were smooth.

Special thanks go my friends from all across the globe whom I've come to know during my stay in New Zealand. I wish you the very best in life as you pursue your dreams. I

would like to thank the Zambian Ministry of Agriculture and Livestock (MAL) for providing the data that forms the basis of the analyses presented in Chapter 3. I would like to specifically thank Dr. Yona Sinkala for geo-coding point locations of foot-and-mouth disease outbreaks, Dr. Joseph Mubanga for encouraging me to pursue post-graduate training and the Department of Human Resources of the Zambian Ministry of Agriculture and Livestock for facilitating my leave of absence to allow me to pursue my studies.

Finally, I would like to thank my family. To My mother Tabeth, my entire childhood I woke up to the sound of your prayers for me. I will always endeavour to make you proud. Thank you to my siblings Rae and Racheal and your lovely spouses for cheering me on. May God richly bless you all.

Nomenclature

CI	Confidence interval
CSO	Central Statistical Office, Zambia
DVLD	Department of Veterinary and Livestock Development, Zambia
FAO	Food and Agriculture Organization of the United Nations
FMD	Foot-and-mouth disease
GIS	Geographic Information System
GPS	Global Positioning System
MAL	Ministry of Agriculture and Livestock, Zambia
NZAID	New Zealand Agency for International Development
OIE	World Organization for Animal Health
OR	Odds ratio
PAR	Population at risk
ppm	Point pattern modelling
TB	Tuberculosis
SAT	Southern African Type
USD	United States dollar

Contents

Acknowledgements **v**

Nomenclature **vii**

1 Introduction **1**

2 Spatial point pattern analyses: A review **5**

2.1 Introduction 5

2.1.1 General concepts 6

2.2 Visualising spatial point patterns 8

2.3 Describing spatial point patterns 10

2.3.1 First-order properties 10

2.3.2 Second-order properties 17

2.4 Explaining spatial point patterns 17

3 Risk factors for foot-and-mouth disease in Zambia, 1981-2012 **21**

3.1 Introduction 21

3.2 Materials and methods 23

3.3 Results 26

3.4 Discussion 35

3.5 Conclusion 36

Bibliography **37**

List of Figures

1.1 Countries in which FMD was reported to the OIE between 1990 and 2002 3

2.1 John Snow's map of Soho showing the location of deaths during the 1854 outbreak of cholera in London 9

2.2 Map of England and Wales showing the distribution of 25 km^2 quadrats containing at least one dairy farm hypothesised to be at risk of acquiring multidrug-resistant *Salmonella* Newport 13

2.3 Steinen diagram showing the point location of FMD serotype O-positive communes in Viet Nam in 2007 14

2.4 Explanation of kernel smoothing of a spatial point process 15

2.5 Prevalence of BSE-positive holdings across Great Britain on 30 June 1997 16

3.1 Image plot of showing the intensity of cattle-containing wards in Zambia 29

3.2 Frequency histogram showing the number of index FMD-outbreak wards detected per year, Zambia 1981 through 2012 30

3.3 Map of Zambia showing the location of international border crossings, major roads, maximum ward-level wetness index and median ward elevation 31

3.4 Map of Zambia showing the location of major bodies of water 32

3.5 Rhohat plot showing FMD-outbreak ward intensity as a function of distance from the nearest railroad 33

List of Tables

3.1 Estimated regression coefficients and their standard errors from a point process model of risk factors for FMD in Zambia, 1981 through 2012 . 34

CHAPTER 1

Introduction

Foot-and-mouth disease (FMD) is a highly infectious disease of cloven hoofed animals (Anonymous, 2012). It is caused by an Aphthovirus of the family Picornaviridae. Since the first written description of FMD by Fracastorius in 1514 (Fracastoro, 1546), FMD outbreaks have occurred in all regions of the world in which livestock are maintained (Grubman and Baxt, 2004) except for New Zealand. Apart from Australia and North America, FMD is enzootic on all continents (Figure 1.1).

The occurrence of FMD in any country automatically excludes that country's participation in international trade of livestock and livestock products. The World Organisation for Animal Health (OIE) provides guidelines that each country must adhere to in order to be considered free of FMD (Anonymous, 2013*b*). Countries can be certified free of FMD by either complete eradication of the disease or by demarcating FMD free zones within the country. In countries such as Zambia in which FMD is endemic, attainment and maintaining FMD free zones is a necessary first step towards attaining FMD free status and gaining access to lucrative international livestock markets.

Establishment of FMD free zones in Zambia and indeed any country requires a knowledge of FMD *hotspots*, that is geographical areas where incursions of FMD are likely to occur on a regular basis. The choice of areas within a country to designate as FMD free will largely depend on the ability of the state veterinary service to either contain the spread of disease should an incursion occur or to create defined areas into which infectious livestock or infected material cannot gain entry. In addition, there is a need to identify risk factors that are associated with FMD incursions in order to better target disease surveillance measures. A detailed knowledge of factors shown to increase the risk of

FMD incursions provides useful information when defining the extent of FMD-free zones and for identifying the location of FMD hotspots.

This thesis is comprised of two chapters. Despite spatial point pattern analysis having been applied in scientific fields such as ecology for many years, its application in medicine is still in its infancy. The intent of the material presented in Chapter 2 is to provide a broad overview of the techniques that can be used for the analysis of spatial point patterns and a description of how these techniques can be applied in a veterinary epidemiological context. Chapter 3 applies many of the theoretical concepts described in Chapter 2 in an analysis of risk factors for FMD in Zambia for the period 1981 to 2012. For countries where FMD is endemic details of the location and timing of outbreaks, gathered over a number of years, provides a useful starting point for development of risk-based approaches for FMD surveillance. Knowledge of the location of FMD hotspots and the geographic features that render physical locations more likely to be FMD hotspots mean that measures can be deployed in these areas to either reduce the likelihood of incursion or reduce the likelihood of spread of disease if and when an incursion occurs. Ultimately this approach should reduce the number of infected livestock enterprises, minimising productivity losses in susceptible populations.

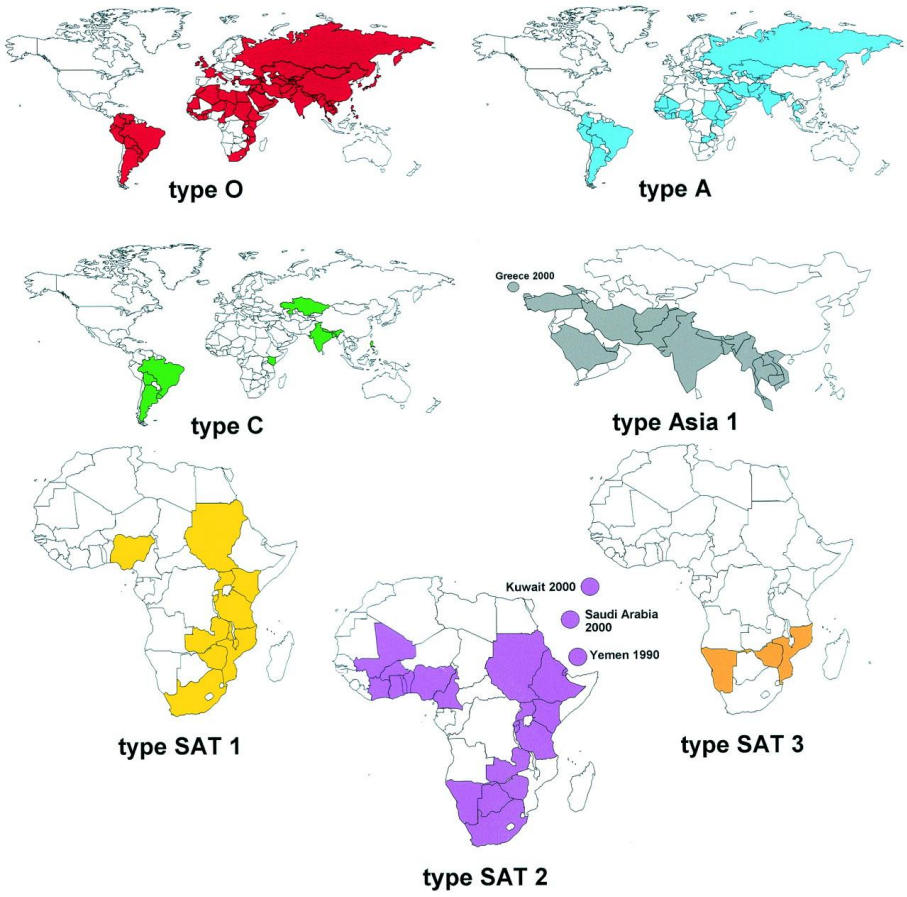


Figure 1.1: Countries in which FMD was reported to the OIE between 1990 and 2002. Data compiled by Knowles (URL: `http://www.wrlfmd.org/maps/fmd_maps.htm`).

CHAPTER 2

Spatial point pattern analyses: A review

2.1 Introduction

The gathering of disease event data by animal health authorities has almost always included documentation of the location of the event as well as the date of occurrence of the event. With continued advancements in mobile communications technology and more widespread use of global positioning system (GPS) enabled smart phones the process of recording precise details of events in space and time has become relatively easy. Furthermore, a wireless Internet connection capability on smart phones means that information that has been captured in the field can be seamlessly uploaded to disease event databases, eliminating the need for personnel to manually transfer event records from one device to another.

Before the advent of this technology location details were typically recorded in text format (e.g. the name of an affected farm or locality); with a GPS-enabled smart phone location details can be directly recorded as point locations, typically in longitude-latitude format. In turn this has increased both the speed and the accuracy of disease event data collection. So, while data collection has become substantially easier in recent years what is not so clearly defined is the suite of analytical methods that can be used by epidemiologists to 'make sense' of accumulated information. While analytical methods to display, describe and explain disease events aggregated to the area level are well described (Bailey and Gatrell, 1995; Pfeiffer et al., 2008) analogous methods for point data are not so clearly defined, nor have they been demonstrated to be robust for inference and decision making across a range of real-world data sets in a veterinary context. We propose that this is not so much due to a problem with the point process analytical techniques themselves,

but more due to a lack of accessibility of these techniques to the community of veterinary epidemiologists in general. With these comments, the purpose of this chapter is to provide an overview of techniques suitable for the analysis of animal health point pattern data sets. Our intention is to document a system for classifying available methods, to briefly describe them, list their strengths and their weaknesses and to finally provide examples of where they have been used in a veterinary context.

2.1.1 General concepts

In point pattern analysis, interest lies in determining how an observed set of point locations relate to each other. That is, we want to know if there is some factor underlying (i.e. explaining) the observed distribution of points and/or is there interaction among points. In general terms the observed distribution of point events is comprised of two components: (a) broad-scale trend, and (b) interaction among individual points.

Broad-scale trend arises from a dependence of the event of interest on one or more explanatory variables, for example physical features of the environment such as roads, soil type, temperature or humidity (Baddeley and Turner, 2005; Illian et al., 2007). Broad-scale trend means that there is systematic variation in the intensity of points that varies depending on proximity to a given explanatory variable. Aggregation of disease events around a hazard (for example, a nuclear power plant or an industrial incinerator) in the absence of other factors produces a marked broad-scale trend: a higher intensity of disease events closer to the hazard with a decreasing intensity as distance from the hazard increases. Colloquially the term 'cluster' is used to describe this pattern of disease events.

Spatial interaction, on the other hand, refers to the degree of attraction or repulsion among points measured at a given spatial scale (Pfeiffer et al., 2008). In epidemiology, interaction is more likely to occur with infectious diseases than it is for non-infectious diseases. This is because the occurrence of an infectious disease in one physical location is likely to influence the presence of disease in susceptible individuals located nearby. Quantifying the amount of spatial interaction among point (e.g. farm) locations provides important information that can be used to inform development of disease control strategies such as setting the diameter of foot-and-mouth disease (FMD) ring vaccination radii, establishing buffer zones around infected and detected farms and setting boundaries for animal move-

ment controls. Unlike the physical clusters described above (where the location of disease events are clearly aggregated) spatial interaction (that is, 'clustering') can still be present even when it is not immediately apparent on visual inspection of a point map. To clarify the distinction between the two we propose that the term *physical clustering* is used to describe the situation where there is an aggregation of point events in defined physical locations (i.e. broad-scale trend is present) and *interaction clustering* is used to describe the situation where there is spatial interaction amongst point events.

Despite spatial point pattern analysis becoming increasingly prominent in the early 1960s with applications confined to the fields of plant ecology, forestry, geology and astronomy (King, 1962; Gatrell et al., 1996), robust statistical methods for the analysis of spatial point patterns have only become widely available in recent years (Baddeley and Turner, 2005; Turner, 2009). In the late 1950s and early 1960s spatial point pattern analysis primarily involved distance-based techniques (Haggett et al., 1977; Gatrell et al., 1996) which involved measuring the distance between all pairs of point events. Area-based methods involved aggregating point events into a regular lattice of areas applied over a region of study and comparing the frequency of events in each cell with that expected under the assumption of complete spatial randomness.

Prior to 2005 the widespread usage of spatial point process analytical techniques was hampered in part by a lack of widely available and easy-to-use software (Gatrell et al., 1996). At the time of writing, a number of fully featured easy-to-use commercial and open-source Geographic Information Systems (GIS) are available allowing users to visualise, describe and explain the spatial distribution of point events with relative ease. Fully featured, open source GIS packages include Quantum GIS (QGIS Development Team, 2009) and GRASS (GRASS Development Team, 2012). A particular feature of the Quantum GIS system is its ability to integrate with other software packages such as R (R Development Core Team, 2014) and PostgreSQL (The PostgreSQL Global Development Group, 2014). An attractive feature of R is that it is the software development platform used by most spatial statisticians which means that newly developed methods are typically made publically available as contributed R packages many years before they are implemented within main-stream commercial software. Examples of R packages for spatial analysis include spatstat (Baddeley and Turner, 2005), sp (Pebesma and Bivand, 2005), splancs (Rowlingson and Diggle, 1993) and sparr (Davies et al., 2011).

2.2 Visualising spatial point patterns

The most frequently method used to visualise point data is to simply plot the Cartesian coordinates of each event as a point location within a defined area of study. When using this approach a balance needs to be struck between displaying the spatial distribution of event locations and providing a visual display that allows the data to be meaningfully interpreted. For this reason, point maps are suitable only for visualising small datasets. Large number of point data within a small region can be condensed into single points by converting them to marked points. For instance, instead of displaying the location of all settlement-level outbreaks of a disease that occurred within a ward individually, wards can be classified as either disease-positive or disease-negative and mapping the point location of disease-positive wards (Hamoonga et al., 2014). John Snow's work investigating the outbreak of cholera that occurred in London in the 1850s (Snow, 1855) shows that point maps are a simple yet extremely effective tool for visualising disease event locations and developing hypotheses about factors influencing the distribution of disease. Using a point map showing the location of cholera deaths around Broad Street in Soho (Figure 2.1) Snow hypothesised that water from a single communal water pump was the source of infection.

Easy access to GIS packages (including those that retrieve spatial data from the Internet on the fly, such as Google Earth) and the widespread availability of detailed spatial data means that it is relatively easy to visualise point data interactively and in finer detail. R packages such as plotKML (Hengl et al., 2014) provide a means for overlaying point patterns directly onto Google Earth maps. This allows visualisation of point patterns to be interactive. Point locations of disease events are superimposed over a real-time representation of the physical landscape, facilitating hypothesis generation about disease occurrence. Stevenson et al. (2005) superimposed the location of *Varroa destructor*-positive honey bee apiaries on a Google Earth image of the Auckland (New Zealand) metropolitan region, identifying a high density of positive apiaries in close proximity to Auckland International Airport. This supported the hypothesis that the airport was the portal of entry of *Varroa destructor* into New Zealand.

Figure 2.1: John Snow's map of Soho showing the location of deaths (•) during the 1854 outbreak of cholera in London. The location of water pumps are shown as crosses (×). Reproduced from (Gilbert, 1958)

2.3 Describing spatial point patterns

2.3.1 First-order properties

In spatial epidemiology the term 'first order' is used to describe broad-scale trend in the spatial pattern of the outcome of interest (Pfeiffer et al., 2008). To investigate spatial trend in a point pattern dataset a useful starting point is to estimate the intensity of the point pattern. Broadly speaking, the intensity of a point pattern is the expected number of points per unit area of the study region (Diggle, 2003). Intensity may be constant (homogenous) or it may vary from one location to another (inhomogeneous).

Several methods can be used to describe the first-order properties and/or intensity of a point-pattern. These methods include: quadrat counts, Stienen diagrams, and kernel smoothing. These methods are briefly discussed below.

Quadrat methods are area-based methods that involve dividing the study region into sub-regions of equal size called quadrats (Stoyan and Stoyan, 1994). The point intensity is then obtained by dividing the number of events in each quadrat by the quadrat area. A disadvantage of this technique is the loss of information that occurs from aggregating the data. Furthermore, the choice of quadrat size influences the number of point events per quadrat and in situations where the quadrat size is too small there will be many empty quadrats, making interpretation difficult. Figure 2.2 shows the distribution of 25 km^2 quadrats containing at least one dairy farm hypothesized to be at risk of acquiring multidrug-resistant *Salmonella* Newport (Durr and Gatrell, 2004). By visualising the spatial distribution of quadrats in this example, we cannot make a distinction between a 25 km^2 quadrat containing 4 farms and one that contains only 1 farm. Intensity of the event of interest within each quadrat is lost (though, legitimately the authors could have chosen to show it). There is little information we can get to determine the intensity of the event of interest within each quadrat. This together with empty quadrats render quadrat methods to be of little use in veterinary epidemiology.

In a Stienen diagram, the distance for each point location and its nearest neighbour is computed. A circle is then drawn around each point location with the radius of the circle equal to half the distance to its nearest neighbour (Stienen, 1980; Gelfand, 2010). The diameter of the resulting circles are equal to the nearest neighbour distances and the cir-

cles never overlap by construction. Stienen diagrams are useful for detecting trends in the proximity between point events. Aggregations of circles of small diameter indicate that the distances between disease-positive locations are small. Circles of large diameter indicate disease-positive locations that are isolated (i.e. the distance to the nearest disease-positive neighbour is large). Figure 2.3 shows a Stienen diagram of Vietnamese communes positive for FMD serotype O in 2007. The aggregations of small diameter circles in the north of the country and on the eastern border of the central area of the country are consistent with the situation where there is commune-to-commune spread of disease. The larger diameter circles are interpreted as communes where disease occurred but was quickly stamped out.

Kernel smoothing is a non-parametric method used to estimate the intensity of one-dimensional and two-dimensional data (Pfeiffer et al., 2008). In some aspects, kernel estimates are like histograms in that they sort sample values (point intensity) into classes and display their frequencies as a colour-coded raster map. Kernel estimates are useful for displaying the intensity of point locations because they make no assumptions about the distributional form of the data (Stoyan and Stoyan, 1994). This is particularly important because point patterns do not typically follow parametric distributions. Kernel smoothing provides an indication of local spatial variation in point intensity. The technique involves placement of a regular grid of cells over the region of interest, placement of a kernel of radius τ over each point and then estimating point intensity as the sum of the kernel estimates that fall within each grid cell (Figure 2.4) (Diggle, 2003; Gatrell et al., 1996). The radius of the kernel τ is usually referred to as the bandwidth or smoothing parameter, and has a direct bearing on how the point density is displayed. When τ is small the kernel smoothed surface will be similar to the spatial point pattern on which it is based. Conversely, when τ is large a greater amount of smoothing will be applied, obscuring detail. While statistical parameters such as normal optimal smoothing, cross-validation, and plug-in methods (Hogg, 1979; Rudemo, 1982; Bowman, 1984; Sheather and Jones, 1991) provide some objectivity for selecting a suitable bandwidth, *a priori* information about the spatial point pattern being described should also be used to guide decision making.

When interpreting smoothed plots of the density of counts of disease, we need to account for the spatial distribution of the underlying population at risk (Pfeiffer et al., 2008). Large numbers of cases present at a given location may either represent a true elevation in

incidence or, more simply may reflect the spatial distribution of the underlying population at risk with larger numbers of cases of disease in areas where population counts are highest (Kelsall and Diggle, 1995*a*,*b*; Lawson, 2006). One way to account for variation in the spatial distribution of the population at risk is to develop two kernel smoothed surfaces: the first for disease-positive locations and the second for disease-positive and disease-negative locations. An estimate of the spatial variation in disease frequency corrected for the spatial distribution of the population at risk is made by dividing the kernel smoothed surface of the disease-positive locations by the kernel smoothed surface of the disease-positive and disease-negative locations (Pfeiffer et al., 2008).

Figure 2.5 shows the prevalence of BSE-positive cattle holdings across Great Britain as at 30 June 1997 (Stevenson et al., 2000). Figure 2.5 shows an obvious first-order trend in the spatial distribution of BSE-positive cattle holdings with relatively high intensities in the south and low intensities in the north. In addition to first-order trend, local features in the data are clearly visible: high intensities of BSE in the south west and low intensities in the east.

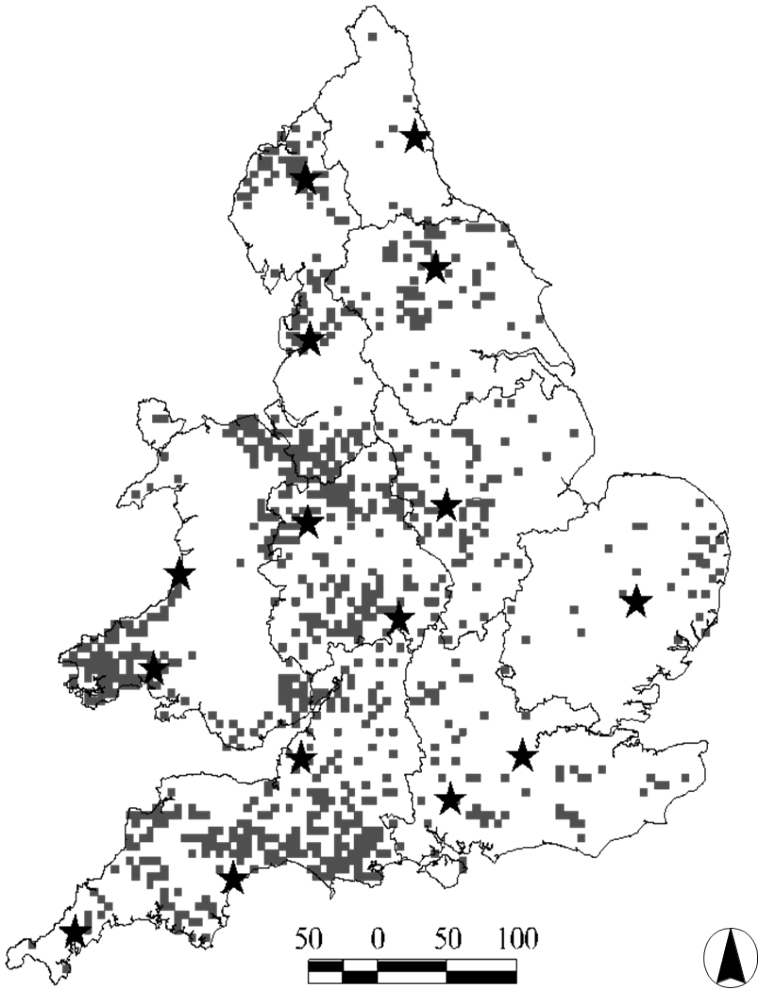


Figure 2.2: Map of England and Wales showing the distribution of 25 km^2 quadrats containing at least one dairy farm hypothesised to be at risk of acquiring multidrug-resistant *Salmonella* Newport. Reproduced from Durr and Gatrell (2004).

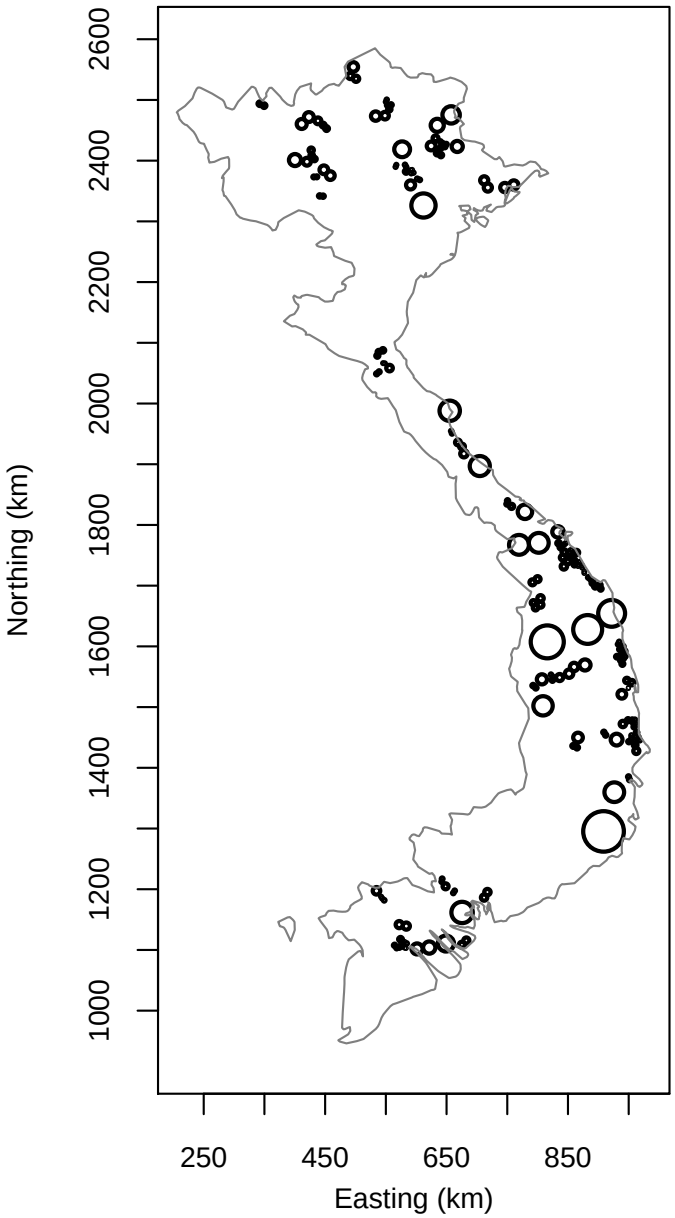


Figure 2.3: Steinen diagram showing the point location of FMD serotype O-positive communes in Viet Nam in 2007. Circles have been drawn around each point location with the diameter of each circle equal to the distance to the nearest FMD-positive commune

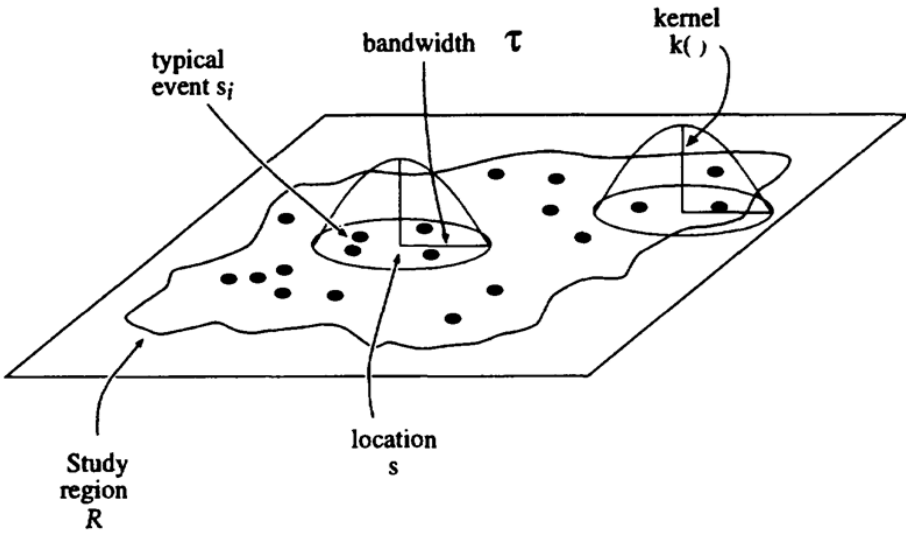


Figure 2.4: Explanation of kernel smoothing of a spatial point process. Kernel smoothing involves placement of a regular grid of cells over the region of interest and placement of a kernel of radius τ over each point. Intensity is estimated as the sum of the kernel estimates that fall within each grid cell. Reproduced from Bailey and Gatrell (1995).

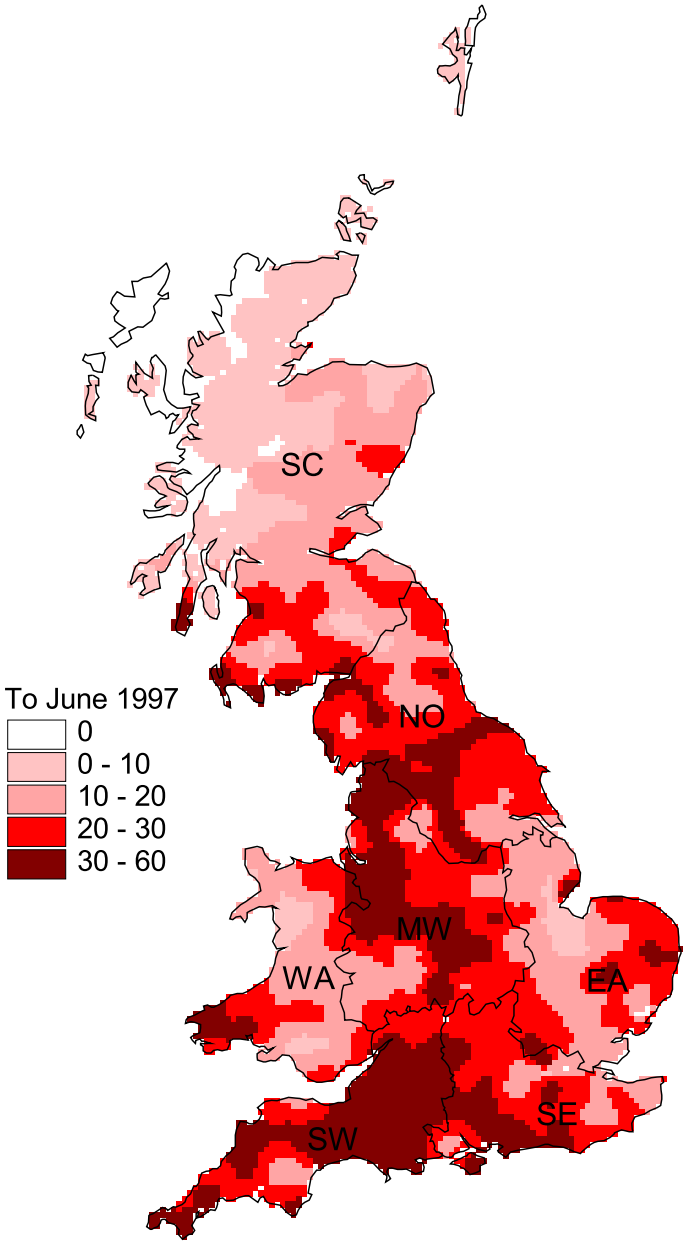


Figure 2.5: Prevalence of BSE-positive holdings across Great Britain (expressed as the number of BSE-positive holdings per 100 holdings per square kilometre) on 30 June 1997. EA Eastern, MW Mid and West, NO Northern, SC Scotland, SE South east, SW South west, WA Wales. Reproduced from Stevenson et al. (2000).

2.3.2 Second-order properties

The term 'second-order' is used to describe small-scale variation is a spatial point pattern typically arising from interactions of individual point locations with its neighbours (Pfeiffer et al., 2008). To assess one aspect of the second-order characteristics of a spatial point pattern, an exploratory technique is to take each event location in the study area and measure the distance to its nearest neighbouring event. This is called the nearest neighbour event-event distance. If there are n events in a study area, there will be n nearest neighbour event-event distances. For a specified distance h, G(h) is defined as the proportion of all nearest neighbour event-event distances that are less than or equal to h. A plot of G(h) as a function of h provides a graphical description of clustering in the process under investigation. Interaction clustering is said to be present if the curve shows a steep rise at small values of h.

The K-function (or reduced second moment measure) (Bartlett, 1964; Ripley, 1976, 1977) is a method used to quantify spatial dependence in spatial point pattern data sets. The K-function provides a summary measure of spatial dependence over distances beyond the distance to a point's nearest neighbour. This property makes it more desirable than the nearest-neighbour function described above. The K-function is defined as the expected number of further points within a distance h of an arbitrary point, divided by the overall density of points (Pfeiffer et al., 2008). For clustered patterns, case events are likely to be surrounded by other case events and for small vales of distance h, K(h) will be relatively large. Conversely, if cases are regularly spaced, each one is likely to be surrounded by empty space and, for small values of distance, K(h) will be small.

2.4 Explaining spatial point patterns

The preceding text has briefly outlined the steps involved when analysing a spatial point pattern data set. The first step involves visualising the spatial distribution of the recorded point events. This is then followed by analytical procedures designed to identify both the first-order and second-order trends in the data. A typical analysis would involve estimation of point intensity using kernel smoothing techniques and/or quadrat counts. Having described the intensity of point events, we may then wish to explore the level of interaction

among points using either nearest neighbour methods or the K-function. In a veterinary epidemiological context a typical interaction question would be: given the nature of this disease, is the observed level of interaction greater than the interaction expected if the point locations were distributed completely at random? If points are assigned a mark (for example, labels identifying them as disease positive or disease negative) we can segregate the points based on their mark and then separately explore interaction and intensity for each set to determine how disease status affects both the first- and second-order properties of the data. Equally important, it may be useful to use modelling techniques to account (i.e. control) for the presence of explanatory variables that may at least partly explain the observed spatial distribution of disease-positive locations. Knowing the relative contribution of a given explanatory variable on the intensity of disease-positive locations then provides a useful starting point for application of control measures. In areas where the explanatory variable is present (or present at high levels) steps should be taken to reduce the likelihood of a given location becoming disease positive (e.g. through enhanced biosecurity, use of prophylactic vaccinations and so forth).

An additional useful side effect of modelling is that it allows one to visualise the residual spatial variation in disease risk that remains once one has controlled for the presence of known explanatory variables. This idea of 'residual disease risk' is closely aligned with the concept of physical clustering, introduced in Section 2.1 and concisely defined by Elliot et al. (2000). Elliot et al. (2000) defines a disease process to be (physically) clustered if disease events remain aggregated once known explanatory variables have been accounted for. Returning to the example cited in Section 2.3.1 we may observe large numbers of cases of disease in a given area. This may represent a true elevation in disease frequency or more simply may be a function of the spatial distribution of the underlying population at risk with large numbers of cases of disease in areas where population counts are highest. In this situation Elliot et al.'s criteria would require us to account (control) for the spatial distribution of the population at risk before declaring a physical disease cluster to be present.

When working with non-spatial data it is common to use contingency tables (2×2 tables) to quantify the association between putative explanatory variables and a disease outcome (Dohoo et al., 2003). When working with spatial point pattern data sets the association between the intensity of a spatial point pattern and a hypothesised explanatory variable

can be assessed using the rhohat procedure (Baddeley et al., 2012) implemented within the spatstat package (Baddeley and Turner, 2005) in R. In brief, the rhohat procedure requires two input variables: (a) a pixel image (equivalent to a raster map) showing the intensity of the spatial point pattern of interest, and (b) a pixel image of the explanatory variable of interest (e.g. elevation). As output the procedure produces a line plot showing point pattern intensity (and a 95% confidence interval around that estimate) as a function of the numeric values of the explanatory variable pixel image. This procedure makes it relatively straightforward to determine firstly if there is an association between the intensity of a spatial point pattern and a hypothesised explanatory variable and secondly the nature of that association, if one exists. Using rhohat analyses as a method to 'screen' a set of putative explanatory variables, the logical next step would be to include each of the explanatory variables shown to be associated with point pattern intensity in a multivariable model.

Unlike models used in classical statistics, point process models (i.e. models of spatial point patterns) are specified in terms of their *conditional intensity* rather than their likelihood. The spatstat package(Baddeley and Turner, 2005) implemented in R (R Development Core Team, 2014) provides a number of functions for fitting point process models and for evaluating their fit. Point process models fitted by the spatstat package are expressed in terms of their Papangelou conditional intensity (Papangelou, 1974) function as follows:

$$log\lambda(u,x) = \phi^T b(u) + \theta^T S(u,x) \tag{2.1}$$

In Equation 2.1 the term $\phi^T b(u)$ represents the broad-scale trend component of the conditional intensity and $\theta^T S(u,x)$ represents the spatial interaction component. To model broad-scale trend linear combinations of explanatory variables may be included in addition to an offset term. In spatial epidemiology, it would be common to apply an offset to represent the spatial distribution of the underlying population at risk.

Selection of explanatory variables that best explain the point process under investigation can be carried out using a backward stepwise variable selection approach. Each of the explanatory variables that are deemed to be associated with the point process (based on the rhohat analyses described in Section 2.4) are entered into the model. Explanatory

variables that are not statistically significant (as estimated by a z test) are then removed from the model one at a time, beginning with the least significant, until the estimated regression coefficients for all of the remaining explanatory variables are significant at an alpha level of less than 0.05.

Once broad-scale trend (i.e. the first-order properties) of the spatial point pattern are accounted-for focus switches to accounting for the second-order properties of the data. This is carried out by inclusion of a spatial interaction term into the model, analogous to the 'family' argument in general linear regression equations (Baddeley and Turner, 2006). Interpoint interaction terms available in the spatstat package include: Poisson, Strauss, Strauss process with hardcore, pairwise soft core interaction, pairwise interaction step function potential, Lennard-Jones potential, Geyer's saturation process, and Ord's process. The Geyer interaction term is comprised of two parameters, a user-defined constant greater than zero c (the saturation threshold) within a spatial radius r of a given point. For a given analysis r is set to a value deemed to be an appropriate range of distance over which the effect of a point location in the study region might influence other locations. The value of r can then be tested against a series of candidate c values. On each occasion model diagnostics are carried out to determine whether or not the parameterised spatial interaction term has appropriately accounted-for the second order features of the data. The appropriateness of the spatial interaction term can be assessed using a quantile-quantile plot implemented in the qqplot.ppm procedure in spatstat.

Outputs from a point process model are regression coefficients for each of the parameterised explanatory variables with the exponent of the regression coefficient for a given explanatory variable interpreted as the relative change in spatial point pattern intensity corresponding to a one-unit change in the explanatory variable. Model fit can be assessed using the diagnose.ppm procedure implemented in spatstat. The diagnose.ppm function produces a series of diagnostic plots using the residuals from the fitted point process model. These plots can be used to assess goodness-of-fit, to identify outliers in the data, and to identify departures of the observed data from the fitted model.

CHAPTER 3

Risk factors for foot-and-mouth disease in Zambia, 1981-2012

Abstract – The aim of this study was to describe the spatial distribution of foot-and-mouth disease (FMD) outbreaks in Zambia for the period January 1981 to December 2012 and to quantify the association between geographical features (proximity to roads, national parks, wetland areas) and the spatial distribution of FMD using a Poisson point process model.

Details of FMD outbreaks retrieved from the Zambian Department of Veterinary and Livestock Development included the date of onset of clinical signs and the name of the ward in which the index case enterprise was located. A total of 62 FMD outbreaks occurred throughout the study period. Outbreaks occurred in the south of the Southern province along the border with Namibia and Botswana (n = 5), in the Western province (n = 2), in the Southern and Central provinces on the Kafue flood plains (n = 44), and in the north east of the country close to the border with Tanzania (n = 11). Increases in distance to the nearest major international border crossing, distance to the nearest major road, distance to the wetland area of the Kafue flood plain, wetness index and elevation were all associated with a decrease in FMD-outbreak ward intensity. Our analyses support the hypothesis that in drier areas of the country cattle are more likely to aggregate around communal drinking pools. Aggregation of cattle provides conditions suitable for FMD spread and detection.

Hamoonga R, Stevenson MA, Allepuz A, Carpenter TE, Sinkala Y (2014) Risk factors for foot-and-mouth disease in Zambia, 1981-2012. *Preventive Veterinary Medicine* 114: 64 – 71.

3.1 Introduction

The first reported outbreak of foot-and-mouth disease (FMD) in Zambia occurred in 1933 (Morris, 1934). With hindsight, the absence of FMD in Zambia prior to 1933 can be attributed to the rinderpest panzootic of 1896 that swept through the southern African region, killing large populations of cattle, buffaloes and other wild ungulates (Condy, 1979; Thomson, 1995). Between 1933 and 1983 FMD outbreaks in Zambia were reported on

29 occasions (Overby and Zyambo, 1983). Since 1983 outbreaks of FMD have continued to occur in Zambia on an almost annual basis.

FMD is a highly infectious disease of cloven-hoofed animals is caused by an Aphthovirus of the family Picornaviridae. It is characterised by high (up to 100%) morbidity in susceptible animal populations. As early as 1931 the impact of FMD on trade was apparent when Zambian animal health authorities placed restrictions on the importation of stock from Zimbabwe (then Rhodesia) due to ongoing outbreaks of FMD in that country (Smith, 1932). While the overall impact of FMD on livestock productivity in Zambia is believed to be relatively small, the primary source of loss arising from the on-going presence of disease is the inability to export meat or meat product due to trade restrictions. In a report commissioned by the World Bank in 2011 (Anonymous, 2011) it was estimated that if Zambia's beef and dairy populations were certified as FMD free their combined output had the potential to result in a seven-fold increase in export earnings from USD 230 million to USD 1.6 billion per year. If this situation were to eventuate it would mean that export of livestock and livestock product would constitute approximately 10% of the country's gross domestic product for 2011.

Between 1933 and 1981 identified outbreaks of FMD were confined to three areas of Zambia (Overby and Zyambo, 1983; Perry and Hedger, 1984): (1) in the southern part of the Southern province close to the border with Namibia and Botswana; (2) along the border of the Southern and Central provinces on the Kafue flood plains; and (3) along the border with Tanzania in the Northern province of the country. Risk factors for the spatial distribution of FMD outbreaks in sub-Saharan Africa include proximity to wildlife, proximity to road and rail networks, the presence of communal drinking places, livestock density and proximity to international border crossings (Perry and Hedger, 1984; Vosloo et al., 2009; Allepuz et al., 2013). The exact role or mechanism by which these factors influence and/or transmit FMD in susceptible populations is only partially understood (Dawe et al., 1994) and complex (Sutmoller et al., 2000). In addition, these risk factors have also been shown to behave differently in different environments (Perry and Hedger, 1984).

For countries where FMD is endemic details of the location and timing of outbreaks, gathered over a number of years, can provide a useful starting point for development of risk-based approaches for FMD surveillance. Knowledge of the location of FMD 'hotspots'

(areas where outbreaks are known to occur on a regular basis) mean that measures can be deployed in these areas to either reduce the likelihood of incursion or reduce the likelihood of spread of disease should an incursion occur. Ultimately this approach should reduce the number of infected livestock enterprises minimising productivity losses in susceptible populations. With this background, the aim of this work was to describe the spatial distribution of FMD outbreaks in Zambia for the period 1981 to 2012 and to quantify the association between geographical features of the landscape (proximity to roads, national parks, wetland areas) and the spatial distribution of FMD outbreaks.

3.2 Materials and methods

Between 1981 and 2012 officials of the Zambian Department of Veterinary and Livestock Development (DVLD) recorded details of FMD outbreaks affecting cattle. Details of the index case (that is, the first enterprise identified as FMD-positive) included the date of onset of clinical signs and the name of the ward in which the index case enterprise was located. The total land area of Zambia is approximately 753,000 km^2. In 2011 the country was comprised of 9 primary administrative areas called provinces. Provinces were divided into secondary administrative areas called districts (n = 72) and districts divided into tertiary administrative areas called wards (n = 1421). The median land area of Zambian wards was 300 km^2 (Q1 95 km^2; Q3 665 km^2).

At the time of initial investigation of the index case for each outbreak samples of blood were collected into sterile blood collection tubes. Serotyping was carried out at the world reference laboratory for FMD at Pirbright (United Kingdom). Throughout the 31-year study period, four FMD serotypes were identified by laboratory diagnosis: the Southern African type (SAT) 1, 2, 3 and serotype O. For the analyses presented in this paper, the unit of interest was the ward and the outcome of interest was a laboratory confirmed diagnosis of FMD reported in a given ward, in a given year. All of the analyses presented in this paper are based on the index case details recorded for each FMD outbreak. The period of interest was from 1 January 1981 through 31 December 2012.

Our first task was to subset the population of Zambian wards to include only those that were known to contain cattle. The Gridded Livestock of the World (Robinson et al., 2007)

database for 2005, produced by the Food and Agriculture Organization of the United Nations, was used to develop a raster map covering the spatial extent of Zambia, comprised of 243 by 206 pixels (each of dimensions 5.5 km by 5.5 km) showing the number of cattle per square kilometer. This raster map was filtered to include only those pixels where cattle were present. The filtered map was then overlaid on a vector map of Zambian ward boundaries and those wards with cattle (n = 1114) were retained as the cattle-containing ward population at risk.

Variables thought to influence the spatial distribution of FMD in Zambia (referred to as explanatory variables in the remainder of the paper) were in two general classes: demographic and geographic. Demographic data included the raster map of cattle density, described earlier. Physical landscape explanatory variables included the point location of major international border crossings, national and secondary road networks, railway lines, national parks, wetlands and elevation. The road, railway line, national park and wetland maps were obtained as vector maps from the Central Statistical office of Zambia (Anonymous, 2013*a*). Elevation data were obtained from the DEM Explorer web site (Han et al., 2012) at a resolution of 30 metres. A raster map of topographical wetness index was developed for the entire land area of Zambia using the digital elevation model data. Wetness index provides a measure of water accumulation as a function of slope and catchment. Wetness index was calculated using a physically based, variable contributing area model of basin hydrology procedures (Beven and Kirkby, 1979) implemented in the Geographic Information System GRASS 6.4.3 (GRASS Development Team, 2012).

The spatial distribution of FMD-outbreak wards in Zambia for the period 1981 to 2012 (and indeed the spatial distribution of any infectious disease affecting livestock enterprises) can be classified into two components: broad-scale trend and dependence (or interaction) between individual points. To identify broad-scale trend the point location of the centroids of FMD-outbreak wards were overlaid on an intensity map of cattle-containing wards. The intensity map of cattle-containing wards was calculated using the spatstat package (Baddeley and Turner, 2005) implemented within R (R Development Core Team, 2014). This analysis was carried out using a regular grid of 200 $\times$ 200 cells superimposed over the extent of Zambia, with the standard deviation of the Gaussian kernel (that is, the bandwidth) fixed at 12,000 metres.

The presence of spatial dependence in the data was assessed by computing the inhomoge-

nous K-function (Ripley, 1976, 1977; Cressie, 1993; Baddeley et al., 2000; Diggle, 2003) for the FMD-outbreak wards and the cattle-containing Zambian ward population at risk. Spatial dependence was deemed to be present if (visually) there were relatively high values of K at given scales of distance for the FMD-outbreak wards, relative to the estimates of K computed for the cattle-containing ward population at risk (Bailey and Gatrell, 1995; Pfeiffer et al., 2008).

To identify which of the hypothesised explanatory variables were associated with broad scale spatial trend a kernel smoothed estimate of the intensity of FMD-outbreak wards was developed using a regular grid of 200 $\times$ 200 cells using a fixed bandwidth of 12,000 metres, similar to the approach used for the cattle-containing wards. Raster maps were then developed showing, for each pixel, the distance (in kilometres) to each of the explanatory variables in vector format: major international border crossings, major roads, railway lines, national parks, and wetland areas. The rhohat procedure (Baddeley et al., 2012) implemented in spatstat was then used to plot FMD-outbreak ward intensity as a function of distance to a given vector explanatory variable. Each of the hypothesised explanatory variables in raster format (cattle density, wetness index, and elevation) were summarised at the ward level and the rhohat procedure used to plot FMD-outbreak ward intensity as a function of the range of numeric values recorded for each raster feature.

Hypothesised explanatory variables where there was a clearly identifiable association with FMD-outbreak ward intensity were then selected as candidate explanatory variables in a point process regression model. The decision to include a given explanatory variable for modeling was based on biological plausibility and the presence of a clearly identifiable relationship between FMD-outbreak ward intensity and either the distance to a given explanatory variable (for those in vector format) or the value of the explanatory variable itself (for those in raster format).

Point process models fitted in spatstat are expressed in terms of the Papangelou conditional intensity function (Papangelou, 1974; van Lieshout, 2000) denoted by $\lambda(u, x)$. In general terms the intensity of FMD-outbreak wards $\lambda(u, x)$ is a loglinear function of parameters ϕ and θ:

$$log\lambda(u, x) = \phi^T b(u) + \theta^T S(u, x) \tag{3.1}$$

In Equation 3.1 the term $\phi^T b(u)$ represents the broad-scale trend component of the conditional intensity and $\theta^T S(u, x)$ represents the spatial interaction component. To model broad-scale trend we included an offset term representing the spatial distribution of cattle-containing ward-years at risk in addition to each of the explanatory variables identified as associated with FMD-outbreak ward intensity from the rhohat analyses, described earlier.

To select those explanatory variables that best explained FMD-outbreak ward intensity a backward stepwise approach was used. Each of the explanatory variables that were associated with FMD-outbreak ward intensity from the rhohat analyses were entered into the model. Explanatory variables that were not statistically significant (as estimated by a z test) were removed from the model one at a time, beginning with the least significant, until the estimated regression coefficients for all of the remaining explanatory variables were significant at an alpha level of less than 0.05.

The inhomogenous K function analyses failed to identify spatial dependence in the data up to a distance of 50 kilometres implying that inclusion of a spatial interaction term into the model was unnecessary. Regardless, a second model was developed that included a Geyer interaction term (Geyer, 1999). The Geyer interaction term is comprised of two parameters, a user-defined constant greater than zero c (the saturation threshold) within a spatial radius r of a given point. For these analyses we set r to 15 km and tested a series of candidate c values ranging from 1 to 10 using an approach similar to that described by Turner (2009). Beyond the interaction distance of 15 km (thought to be an appropriate range of distance over which the effect of a given FMD-positive location might influence other locations) no further point-to-point interactions were allowed and the model reduces to a Poisson model.

Outputs from the point process model were regression coefficients for each of the parameterised explanatory variables with the exponent of the regression coefficient for a given explanatory variable can be interpreted as representing the effect of a one-unit increase in the value of the explanatory variable on FMD-outbreak ward intensity

Model fit was assessed using the diagnose.ppm procedure implemented in spatstat. The diagnose.ppm produces a series of diagnostic plots using the residuals from the fitted point process model (Baddeley and Turner, 2005). These plots were used to assess goodness-of-fit, to identify outliers in the data, and to reveal departures from the fitted model. The appropriateness of the models with and without the Geyer interaction term was assessed

using a quantile-quantile (QQ) plot implemented in the qqplot.ppm procedure in spatstat (Baddeley and Turner, 2005). One hundred simulations were carried out to define the limits of the reference quantiles for the Q-Q plot.

3.3 Results

Figure 3.1 is an image plot showing the intensity of cattle-containing wards in Zambia. Superimposed on this plot are the point location of the n = 62 FMD-outbreak wards identified during the period January 1981 through December 2012. Of note in Figure 3.1 is the presence of outbreaks in the south of the Southern and Western provinces along the border with Namibia and Botswana (n = 5). Of note as well is the relatively large number of outbreaks in the Southern and Central provinces on the Kafue flood plains (n = 44) and along the border with Tanzania in the north east of the country (n = 11). Figure 3.2 is a frequency histogram showing the number of FMD-outbreak wards detected per year for the period 1981 through 2012. FMD outbreaks occurred in 19 of the 31 years that comprised the study period. The minimum number of ward-level outbreaks per year was one (1987, 1988, 1992, 1995, 2002, and 2007); the maximum number was 12 (2008).

Figure 3.3 is a series of maps showing the spatial distribution of selected explanatory variables included in the final point process model: location of major international border crossings, location of major roads, wetness index and elevation. Figure 3.4 shows the location of the major wetlands in Zambia, as recorded by Central Statistical Office of Zambia. For the purpose of modeling we classified each of the wetland polygons into one of four major groups, labeled A, B, C and D in Figure 3.4. Based on our descriptive analyses (Figure 3.1) there was a relatively strong spatial association between FMD-outbreak ward intensity and proximity to wetland area A. By grouping the wetlands into categories our intention was to quantify the strength of this association for each wetland area.

Figure 3.5a shows the distance map computed for the Zambian railroad network. The corresponding rhohat plot, showing FMD-outbreak ward intensity as a function of distance from the nearest railroad is shown in Figure 3.5b. In Figure 3.5b it is evident that FMD-outbreak ward intensity decreased as the distance of a ward from the nearest railroad increased. Similar plots (not necessarily demonstrating the same pattern of association)

were produced for each of the hypothesised explanatory variables described in this paper (data not presented).

Estimated regression coefficients and their standard errors for each of the explanatory variables included in the point process model without the Geyer interaction term are shown in Table 3.1. FMD-outbreak ward intensity increased with increases in the distance of a ward centroid from wetland areas B, C and D (as defined in Figure 3.4). FMD-outbreak ward intensity decreased with increases in the distance to the nearest major international border crossing, distance to the nearest major road, distance to wetland area A (as defined in Figure 3.4), maximum ward-level wetness index and median ward-level elevation. The results of the point process model show that FMD-outbreak ward intensity decreased in wetter areas of the country: unit increases in maximum ward-level wetness index decreased FMD-outbreak ward intensity by a factor of 0.88 (95% CI 0.80 to 0.97).

The cumulative residual sum plots produced by the diagnose.ppm procedure lay predominantly within the 2σ-limits. The exception was the easting and northing coordinates consistent with the location of wetland area A where there were aggregations of positive-sign residuals. This indicates the presence of FMD in this area of the country that was not completely explained by the explanatory variables included in the point process model shown in Table 3.1. The QQ plot of the residuals from the point process model lay within the bounds of the reference quantiles, indicating that the model presented in Table 3.1 provided an adequate representation of the interpoint interaction in the data.

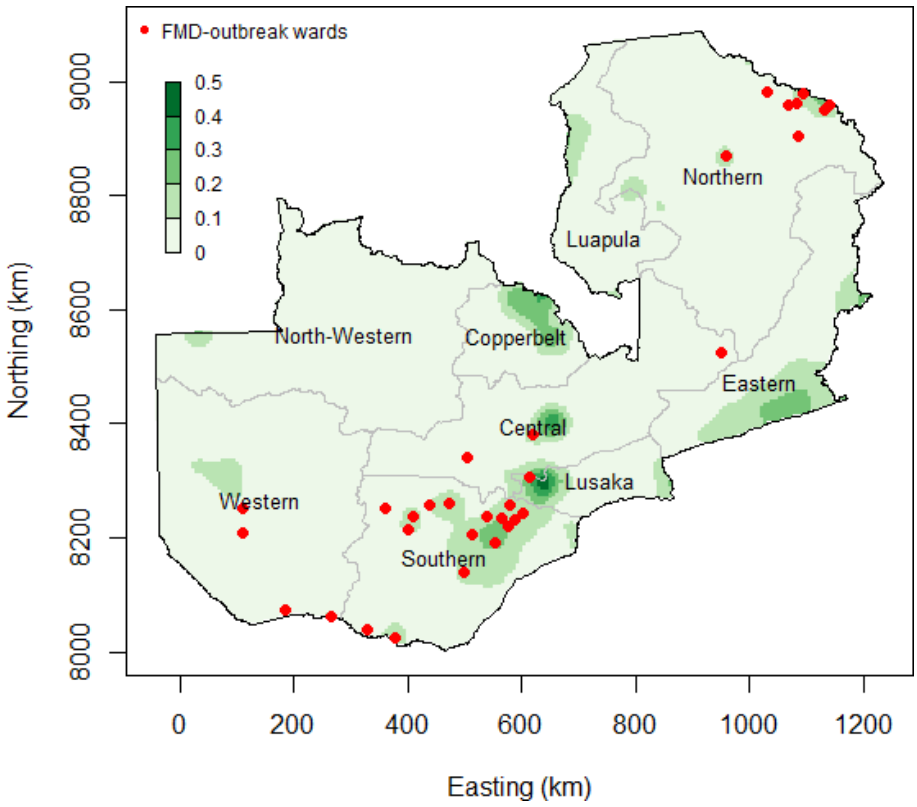


Figure 3.1: Image plot showing the intensity of cattle-containing wards (expressed as the number of cattle-containing wards per square kilometer). Superimposed on this plot are the 62 point locations of wards identified with FMD outbreaks during the period January 1981 through December 2012.

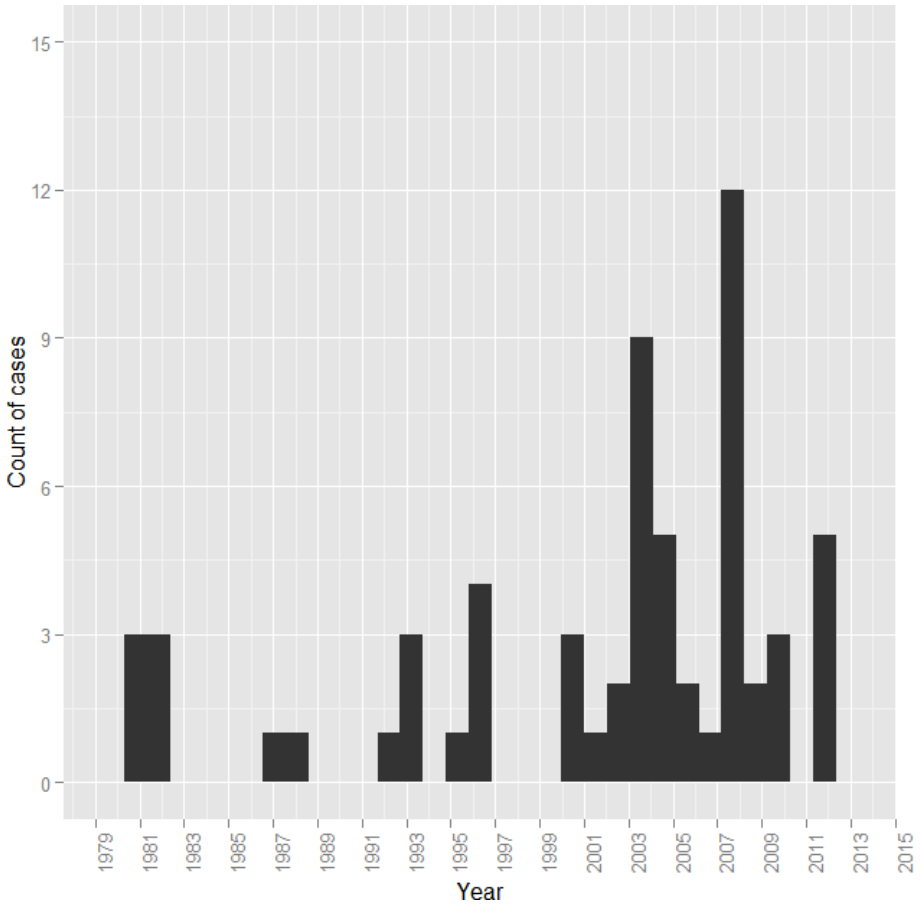


Figure 3.2: Foot-and-mouth disease in Zambia, January 1981 through December 2012. Frequency histogram showing the number of index FMD-outbreak wards detected per year.

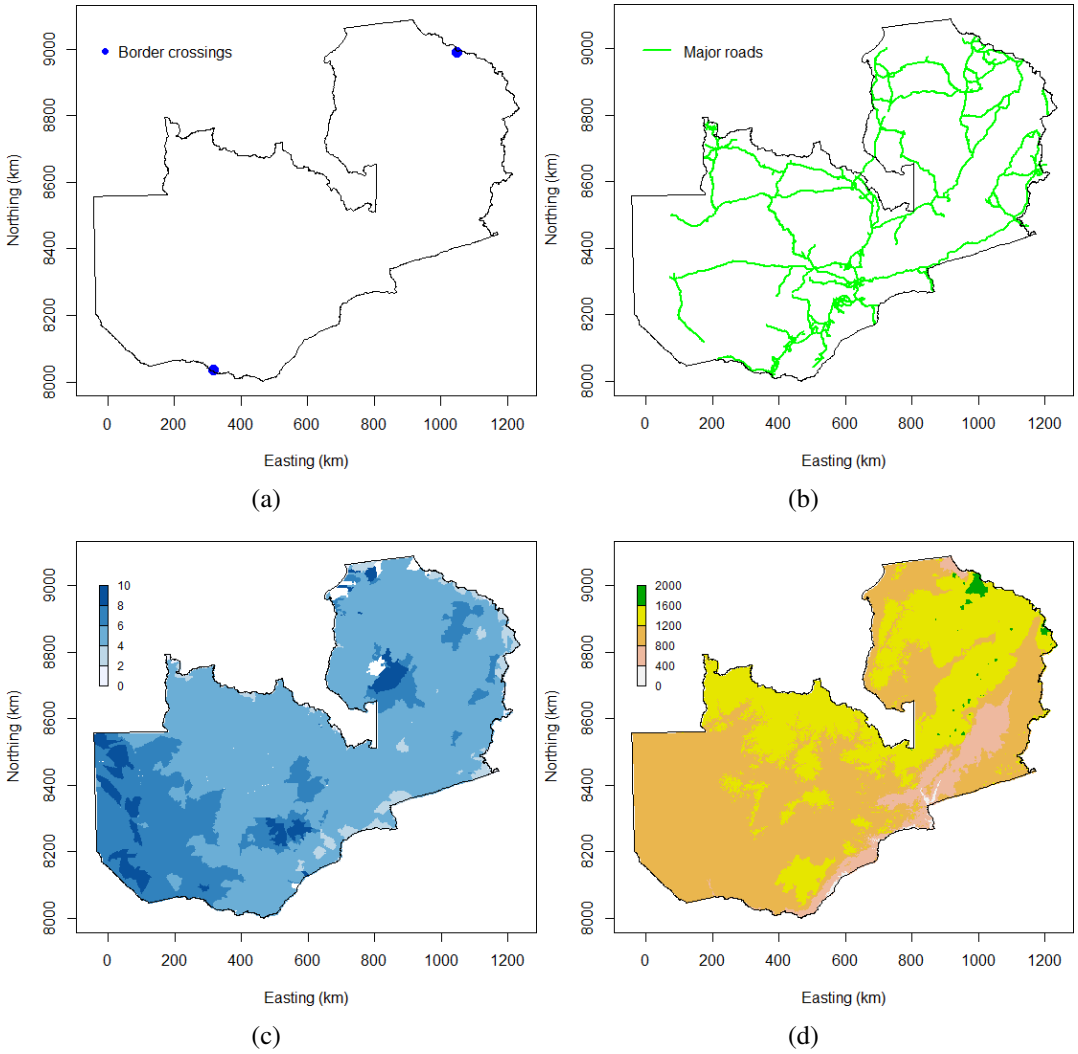


Figure 3.3: Foot-and-mouth disease in Zambia, January 1981 through December 2012. Maps of Zambia showing: (a) location of major international border crossings; (b) location of major roads; (c) maximum ward-level wetness index, as described in the text; and (d) median ward elevation (in metres).

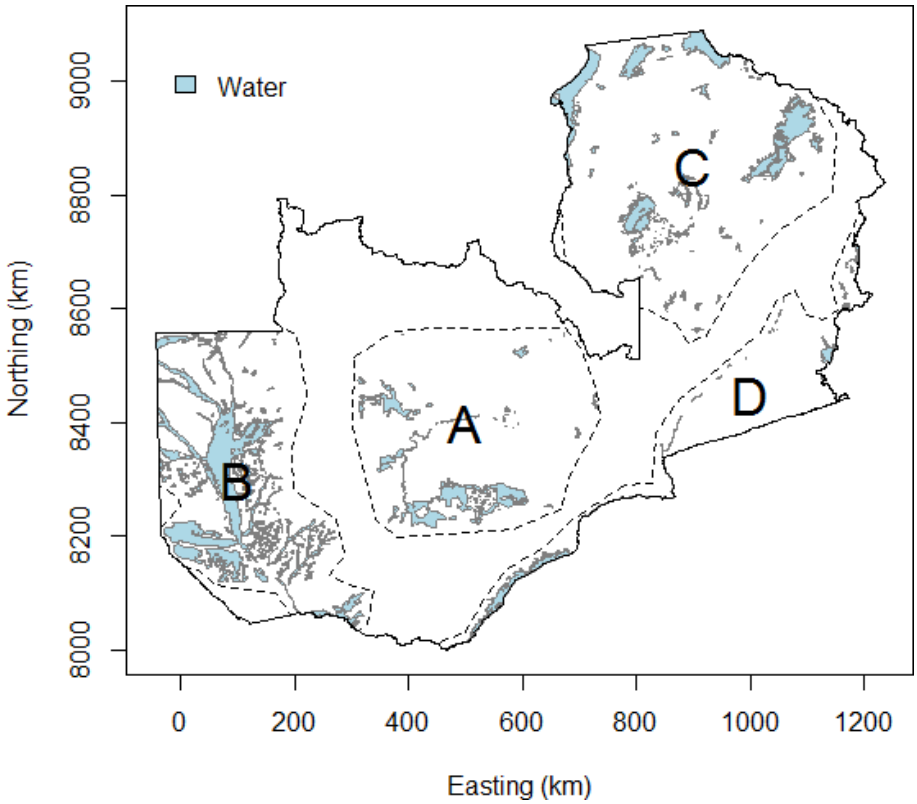
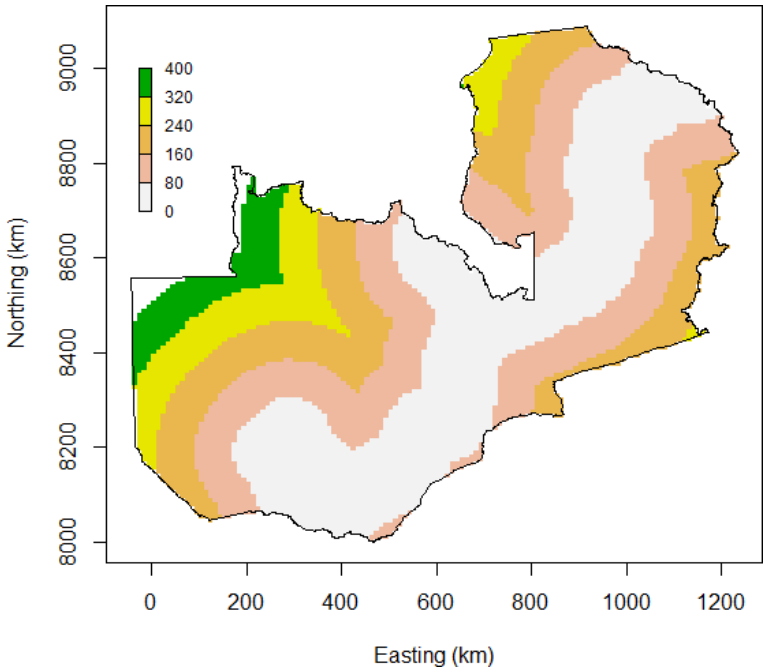
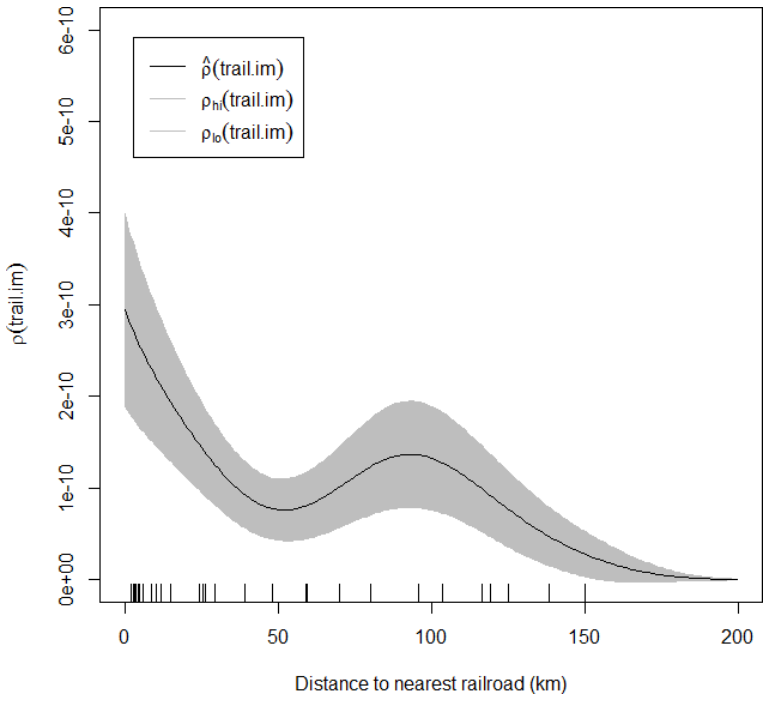


Figure 3.4: Foot-and-mouth disease in Zambia, January 1981 through December 2012. Map of Zambia showing the location of major bodies of water. For modeling these have been arbitrarily grouped into the four regions (A, B, C and D) delineated by the dashed lines shown in the above plot.



(a)



(b)

Figure 3.5: Foot-and-mouth disease in Zambia, January 1981 through December 2012: (a) image plot showing distance (in kilometres) from railroads; (b) rhohat plot showing FMD-outbreak ward intensity as a function of distance (in kilometres) from the nearest railroad.

Table 3.1: Estimated regression coefficients and their standard errors from a point process model showing the association between distance to the nearest major border crossing, distance from major roads, distance from wetlands, maximum ward-level wetness index and median ward-level elevation on the intensity of FMD-outbreak wards in Zambia, 1981 to 2012.

Explanatory variable	Coefficient (SE)	z test	Intensity (95% CI)
Intercept	-34.7601 (4.894)		
Border crossing (25 km increases)	-0.4019 (0.0574)	< 0.01	0.67 (0.60 to 0.75)
Distance from roads (25 km)	-1.1093 (0.2712)	< 0.01	0.33 (0.19 to 0.56)
Distance from wetland 'A' (25 km)	-0.8064 (0.1186)	< 0.01	0.45 (0.35 to 0.56) [a]
Distance from wetland 'B' (25 km)	0.8335 (0.1550)	< 0.01	2.30 (1.70 to 3.12)
Distance from wetland 'C' (25 km)	0.7400 (0.1616)	< 0.01	2.10 (1.53 to 2.88)
Distance from wetland 'D' (25 km)	0.3311 (0.0738)	< 0.01	1.39 (1.20 to 1.61)
Maximum wetness index (unit increases)	-0.1228 (0.0492)	< 0.05	0.88 (0.80 to 0.97)
Median elevation (100 m increments)	-0.1904 (0.0942)	< 0.05	0.83 (0.69 to 0.99)

[a] Interpretation: 25 km increases in the distance from the boundaries of water area 'A' (as shown in Figure 3.4) decreased the intensity of FMD-outbreak wards by a factor of 0.45 (95% CI 0.35 to 0.56).
SE: standard error.
CI: confidence interval.

3.4 Discussion

The development of both targeted control strategies and effective surveillance systems for FMD occurrence is dependent on knowledge of 'hotspots'; areas where outbreaks of disease are known to occur on a regular basis. In Zambia, the system of livestock husbandry system is extensive and traditional pastoralists manage greater than 80% of the total cattle population (Anonymous, 2011). There is little control over livestock movement as animals are moved from one location to another in search of suitable grazing areas. Spatial modeling is useful in this situation because it provides the opportunity to quantify the association (albeit at a crude level when applied at the whole-country level) between physical landscape features and the starting point (i.e. the index case location) of individual FMD outbreaks. In turn, this allows animal health authorities to better target disease surveillance and control activities.

Our analyses show that the distribution of FMD outbreaks in Zambia for the period 1981 to 2012 was almost identical to the distribution of disease outbreaks documented prior to 1981 (Overby and Zyambo, 1983; Perry and Hedger, 1984). This consistency provides indirect evidence of an absence of selection or misclassification bias in the 1981 to 2012 data. While under-reporting of FMD outbreaks in endemic countries is common (Sumption et al., 2008) the consistent appearance of FMD in the three regions shown in Figure 3.1, in the absence of a formal disease event data collection and management system, provides a reasonable level of confidence that there was little systematic error in disease reporting. Furthermore, the spatial persistency of the identified outbreak areas implies that risk factors for FMD incursions in Zambia have remained relatively constant over time.

We found that distance to the nearest major international border crossing, distance to the nearest major road, wetness index and elevation were all associated with FMD-outbreak ward intensity. While outbreaks of FMD are known to occur in both rainy and dry seasons, dry season outbreaks have been reported to account for a larger proportion (up to 70%) of outbreaks in Zambia in a given year (Perry and Hedger, 1984). The increased risk of FMD occurrence in drier areas of the country (that is, wards with a lower maximum water index) can be attributed to animals in these areas being more likely to move in search of water and aggregate at communal drinking pools. Wetlands, such as the Kafue

flood plains (located in the north of the Southern province, labeled 'A' in Figure 3.4) are a popular aggregation point for large numbers of cattle during winter grazing (Overby and Zyambo, 1983; Perry and Hedger, 1984; Muma et al., 2011). Our analyses support the hypothesis that in drier areas of the country cattle are more likely to aggregate around communal drinking pools. Aggregation of cattle then provides conditions suitable for FMD spread.

Although proximity to railways showed a clearly identifiable association with FMD outbreaks (Figure 3.5b), after adjusting for the effect of the other explanatory variables included in the model, proximity to railways was not significantly associated with FMD-outbreak ward intensity. This demonstrates that proximity to railways was confounded by one or more of the other explanatory variables included in the point process model. While railways can facilitate disease transmission by allowing people or animals carrying infection from distant areas to come into contact with naive populations (Muuka et al., 2013) the presence of disease in the population of domestic livestock in close proximity to railways could also be simply attributed to the fact that relatively large numbers of the rural human population (and therefore the livestock population at risk) live in close proximity to rail networks. This observation represents an important lesson for the spatial epidemiologist. The presence of an association between a single geographical feature and disease intensity needs to be interpreted with caution because of the fact that geographical features (particularly if they are man-made) tend to be spatially correlated with other, potentially more direct, determinants of disease. The point process modeling approach presented in this study provided an effective means for dealing with this problem.

3.5 Conclusion

Our analyses show that the spatial distribution of FMD outbreaks in Zambia for the period 1981 to 2012 followed a similar pattern to that of the outbreaks recorded between 1933 and 1981. The intensity of FMD-outbreak wards was associated with distance to the nearest major international border crossing, distance to the nearest major road, wetness index and elevation. The increased risk of FMD occurrence in drier areas of the country can be attributed to animals in these areas being more likely to move in search of water and aggregate at communal drinking pools. Our analyses support the hypothesis that in

drier areas of the country cattle are more likely to aggregate around communal drinking pools. Aggregation of cattle is conducive for FMD spread and detection.

Bibliography

Allepuz, A., Stevenson, M., Kivaria, F., Berkvens, D., Casal, J. and Picado, A. (2013), 'Risk Factors for Foot-and-Mouth Disease in Tanzania, 2001-2006', *Transboundary and Emerging Diseases* . DOI: 10.1111/tbed.12087.

Anonymous (2011), Zambia — What Would it Take for Zambias Beef and Dairy Industries to Achieve Their Potential?, Technical report, World Bank. Available as: `http://documents.worldbank.org/curated/en/home`. Accessed 19 March 2014.

Anonymous (2012), 'Foot and mouth disease. Terrestrial Animal Health Code. Chapter 2.1.5'. Available as: `http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.01.05_FMD.pdf`. Accessed 19 March 2014.

Anonymous (2013*a*), *Digital maps of Zambia*, Central Statistical Office, Zambia, Lusaka, Zambia. Available as: `http://www.zamstats.gov.zm`. Accessed 19 March 2014.

Anonymous (2013*b*), 'Foot and mouth disease. Terrestrial Animal Health Code. Chapter 8.6'. Available as: `http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/2010/chapitre_1.8.6.pdf`. Accessed 19 March 2014.

Baddeley, A., Chang, Y.-M., Song, Y. and Turner, R. (2012), 'Nonparametric estimation of the dependence of a spatial point process on spatial covariates', *Statistics and Its Interface* **5**, 221–236.

Baddeley, A. J., Moller, J. and Waagepetersen, R. (2000), 'Non- and semi-parametric estimation of interaction in inhomogeneous point patterns', *Statistica Neerlandica* **54**, 329–350.

Baddeley, A. and Turner, R. (2005), 'spatstat: An R package for analyzing spatial point patterns', *Journal of Statistical Software* **12**, 1–42.

Baddeley, A. and Turner, R. (2006), *Modelling Spatial Point Patterns in R*, Vol. 185 of *Lecture Notes in Statistics*, Springer, New York, USA, pp. 23–74.

Bailey, T. and Gatrell, A. (1995), *Interactive Spatial Data Analysis*, Longman Scientific and Technical, London, UK.

Bartlett, M. (1964), 'The spectral analysis of two-dimensional point processes', *Biometrika* **51**, 299–311.

Beven, K. J. and Kirkby, M. J. (1979), 'A physically based, variable contributing area model of basin hydrology', *Hydrological Sciences Bulletin* **24**, 43–69.

Bowman, A. (1984), 'An alternative method of cross-validation for the smoothing of density estimates', *Biometrika* **71**, 353–360.

Condy, J. B. (1979), 'A history of foot-and-mouth disease in Rhodesia', *Rhodesian Veterinary Journal* **10**, 2–10.

Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley & Sons, London, UK.

Davies, T. M., Hazelton, M. L. and Marshall, J. C. (2011), 'sparr: Analyzing spatial relative risk using fixed and adaptive kernel density estimation in R', *Journal of Statistical Software* **39**, 1–14.

Dawe, P. S., Sorensen, K., Ferris, N. P., Barnett, I. T. R., Armstrong, R. M. and Knowles, N. J. (1994), 'Experimental transmission of foot-and-mouth disease virus from carrier African buffalo (*Syncerus caffer*) to cattle in Zimbabwe', *Veterinary Record* **134**, 211–215.

Diggle, P. (2003), *Statistical Analysis of Spatial Point Patterns*, Arnold, London, UK.

Dohoo, I. R., Martin, W. and Stryhn, H. (2003), *Veterinary Epidemiologic Research*, AVC Incorporated, Charlottetown, Prince Edward Island, Canada.

Durr, P. and Gatrell, A. (2004), *GIS and Spatial Analysis in Veterinary Science*, CABI, London, UK.

Elliot, P., Wakefield, J. C., Best, N. G. and Briggs, D. J. (2000), *Spatial Epidemiology: Methods and Applications*, Oxford University Press, London, UK.

Fracastoro, G. (1546), *De sympathia et antipathia rerum liber unus. De contagione et contagiosis morbis et curatione libri. III. - Venetiis, (Heredes-Lucae-Antonii Juntae) 1546*, Apud Heredes Lucaeantonij Iuntae Florentini.

Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. (1996), 'Spatial point pattern analysis and its application in geographical epidemiology', *Transactions of the Institute of British Geographers* **21**, 256–274.

Gelfand, A. E. (2010), *Handbook of Spatial Statistics*, CRC Press, Boca Raton, USA.

Geyer, C. (1999), *Likelihood inference for spatial point processes*, Springer, New York, USA, pp. 79–140.

Gilbert, E. W. (1958), 'Pioneer maps of health and disease in England', *Geographical Journal* **124**(2), 172–183.

GRASS Development Team (2012), *Geographic Resources Analysis Support System (GRASS GIS) Software*, Open Source Geospatial Foundation. Available as: `http://grass.osgeo.org/`. Accessed 19 March 2014.

Grubman, M. J. and Baxt, B. (2004), 'Foot-and-mouth disease', *Clinical Microbiology Reviews* **17**, 465–493.

Haggett, P., Cliff, A. D. and Frey, A. (1977), *Locational Analysis in Human Geography*, Edward Arnold, London, UK.

Hamoonga, R., Stevenson, M. A., Allepuz, A., Carpenter, T. E. and Sinkala, Y. (2014), 'Risk factors for foot-and-mouth disease in Zambia, 19812012', *Preventive Veterinary Medicine* **114**(1), 64–71.

Han, W., Di, L., Zhao, P. and Shao, Y. (2012), 'DEM Explorer: An online interoperable DEM data sharing and analysis system', *Environmental Modelling & Software* **38**, 101–107.

Hengl, T., Roudier, P., Beaudette, D., Pebesma, E. and Blaschek, M. (2014), *plotKML: Visualization of spatial and spatio-temporal objects in Google Earth*. Available as: `http://cran.at.r-project.org/web/packages/plotKML/index.html`. Accessed 19 March 2014.

Hogg, R. V. (1979), 'Statistical robustness — One view of its use in applications today', *American Statistician* **33**, 108–115.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2007), *Statistical Analysis & Modelling of Spatial Point Patterns*, Wiley-Interscience, London, UK.

Kelsall, J. E. and Diggle, P. J. (1995*a*), 'Kernel estimation of relative risk', *Bernoulli* **1**, 3–16.

Kelsall, J. E. and Diggle, P. J. (1995*b*), 'Non-parametric estimation of spatial variation in relative risk', *Statistics In Medicine* **14**, 2335–2342.

King, L. J. (1962), 'A quantitative expression of the pattern of urban settlements in selected areas of the United States', *Tijdschrift Voor Economische En Sociale Geografie* **53**, 1–7.

Lawson, A. (2006), *Statistical Methods in Spatial Epidemiology*, John Wiley & Sons, London, UK.

Morris, J. P. A. (1934), *Annual Report of the Department of Animal Health for the Year 1933*, Government Printer, Northern Rhodesia.

Muma, J. B., Munyeme, M., Matope, G., Siamudaala, V. M., Munang'andu, H. M., Matandiko, W., Godfroid, J., Skjerve, E. and Tryland, M. (2011), 'Brucella seroprevalence of the Kafue lechwe (*Kobus leche kafuensis*) and Black lechwe (*Kobus leche smithemani*): Exposure associated to contact with cattle', *Preventive Veterinary Medicine* **100**, 256–260.

Muuka, G., Songolo, N., Kabilika, S., Hang'ombe, B., Nalubamba, K. and Muma, J. (2013), 'Challenges of controlling contagious bovine pleuropneumonia in sub-Saharan Africa', *Tropical Animal Health & Production* **45**, 9–15.

Overby, E. and Zyambo, G. C. N. (1983), 'Foot-and-mouth disease outbreaks in Zambia', *Revue Scientifique et Technique, Office International des Epizooties* **2**, 189–197.

Papangelou, F. (1974), 'The conditional intensity of general point processes and an application to line processes', *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete* **28**, 207–226.

Pebesma, E. J. and Bivand, R. S. (2005), 'Classes and methods for spatial data in R', *R News* **5**, 9–13.

Perry, B. D. and Hedger, R. S. (1984), 'History and epidemiology of foot-and-mouth disease in Zambia: A review', *Tropical Animal Health and Production* **16**, 107–114.

Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D. and Clements, A. (2008), *Spatial Analysis in Epidemiology*, Oxford University Press, London, UK.

QGIS Development Team (2009), *QGIS Geographic Information System*, Open Source Geospatial Foundation.
URL: *http://qgis.osgeo.org*

R Development Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B. (1976), 'The second-order analysis of stationary point processes', *Journal of Applied Probability* **13**, 255–266.

Ripley, B. (1977), 'Modelling spatial patterns', *Journal of the Royal Statistical Society Series B* **39**, 172–192.

Robinson, T. P., Franceschini, G. and Wint, W. (2007), 'The Food and Agriculture Organization's Gridded Livestock of the World', *Veterinaria Italiana* **43**, 745–751.

Rowlingson, B. and Diggle, P. (1993), 'Splancs: Spatial point pattern analysis code in S-Plus', *Computers & Geosciences* **19**, 627–655.

Rudemo, M. (1982), 'Empirical choice of histograms and kernel density estimators', *Scandinavian Journal of Statistics* **9**, 65–78.

Sheather, S. J. and Jones, M. C. (1991), 'A reliable data-based bandwidth selection method for kernel density-estimation', *Journal of the Royal Statistical Society Series B* **53**, 683–690.

Smith, J. (1932), *Annual Report of the Department of Animal Health for the Year 1931*, Government Printer, Northern Rhodesia.

Snow, J. (1855), *On the Mode of Communication of Cholera*, John Churchill, London.

Stevenson, M., Benard, H., Bolger, P. and Morris, R. (2005), 'Spatial epidemiology of the Asian honey bee mite (*Varroa destructor*) in the North Island of New Zealand', *Preventive Veterinary Medicine* **71**, 241–252.

Stevenson, M., Wilesmith, J., Ryan, J., Morris, R., Lawson, A., Pfeiffer, D. and Lin, D. (2000), 'Descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in Great Britain to June 1997', *Veterinary Record* **147**, 379–384.

Stienen, H. (1980), 'The sectioning of randomly dispersed particles, a computer-simulation', *Mikroskopie* **37**, 74–78.

Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes, and Point Fields: Methods of Geometrical Statistics*, John Wiley & Sons, London, UK.

Sumption, K., Rweyemamu, M. and Wint, W. (2008), 'Incidence and distribution of foot-and-mouth disease in Asia, Africa and South America: Combining expert opinion, official disease information and livestock populations to assist risk assessment', *Transboundary and Emerging Diseases* **55**, 5–13.

Sutmoller, P., Thomson, G. R., Hargreaves, S. K., Foogin, C. M. and Anderson, E. C. (2000), 'The foot-and-mouth disease risk posed by African buffalo within wildlife conservancies to the cattle industry of Zimbabwe', *Preventive Veterinary Medicine* **44**, 43–60.

The PostgreSQL Global Development Group (2014), *PostgreSQL 9.3.3*. Available as: `http://www.postgresql.org/`. Accessed 19 March 2014.

Thomson, G. R. (1995), 'Overview of foot-and-mouth-disease in southern Africa', *Revue Scientifique Et Technique De L Office International Des Epizooties* **14**, 503–520.

Turner, R. (2009), 'Point patterns of forest fire locations', *Environmental and Ecological Statistics* **16**, 197–223.

van Lieshout, M. (2000), *Markov Point Processes and Their Applications*, Imperial College Press, London, UK.

Vosloo, W., Thompson, P. N., Botha, B., Bengis, R. G. and Thomson, G. R. (2009), 'Longitudinal study to investigate the role of impala (*Aepyceros melampus*) in foot-and-mouth disease maintenance in the Kruger national park, South Africa', *Transboundary and Emerging Diseases* **56**, 18–30.