

THESIS

Integrating GIS approaches with geographic profiling as a novel conservation tool

Sally Faulkner (MRes, BSc)

Supervisors: Dr Steven Le Comber, Dr Trenton Garner.

Thesis Submitted to Queen Mary University of London for the degree of Doctor of
Philosophy September 2018.

Funded by National Environment Research Council and Queen Mary University of
London.

I, Sally Faulkner, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

To Dad.

Abstract

Geographic profiling (GP) was originally developed to solve the problem of information overload when dealing with cases of serial crime. In criminology, the model uses spatial data relating to the locations of connected crimes to prioritise the search for the criminal's anchor point (usually a home or workplace), and is extremely successful in this field. Previous work has shown how the same approach can be adapted to biological data, but to date the model has assumed a spatially homogenous landscape, and has made no attempt to integrate more complex spatial information (eg, altitude, land use). It is this issue that I address here. In addition, I show for the first time how the model can be applied to conservation data and – taking the model back to its origins in criminology – to wildlife crime. In Chapter 2, I use the Dirichlet Process Mixture (DPM) model of geographic profiling to locate sleep trees for tarsiers in dense jungle in Indonesia, using as input the locations at which calls were recorded, demonstrating how the model can be applied to locating the nests, dens or roosts of other elusive animals and potentially improving estimates of population size, with important implications for management of both species and habitats. In Chapter 3, I show how spatial information in the form of citizen science could be used to improve a study of invasive mink in the Hebrides. In Chapter 4, I turn to the issue of 'commuter crime' in a study of poaching in Savé Valley Conservancy (SVC) in Zimbabwe, in which although poaching occurs inside SVC the majority of poachers live outside, showing how the model can be adjusted to reflect a simple binary classification of the landscape (inside or outside SVC). Finally, in Chapter 5, I combine more complex land use information (estimates of farm density) with the GP model to improve predictions of human-wildlife conflict.

Acknowledgments

Firstly, I would like to express sincere gratitude to my primary supervisor and good friend Steve. Not only did he offer a lot of patience, and guidance but his enthusiasm instilled a genuine sense of fun into my research and despite his relentless inability to spell whiskey correctly, I genuinely enjoyed the whole PhD process under his guidance. I would also like to thank my secondary supervisor Trent who offered advice, support and encouragement – often in the form of chocolate and trips to the pub. I could not imagine a better mentors and friends to get me through the last four years.

There is a long list of collaborators who have trusted me to analyse their data and I thank them all. Thanks to Bob Verity who has remained incredibly interested in the project, and has remained an integral part of the team and to Michael Stevens who despite not being a biologist was prepared to listen to me talk about tigers and tarsiers. I hope you enjoy the process as much as I did. Stephanie Romañach and Peter Lindsey who allowed me access to some poaching data which became the basis for one of my chapters and produced my favourite paper to date. It was this analysis that has cemented my passion for conservation. Sugoto Roy and David Roberts passed on some ‘throw away’ data to me and I was able to produce a very significant fundamental chapter. Thanks to everyone at the Operation Wallacea, Indonesian site, it was here that my path started as an independent researcher.

I must also thank Mo, who has provided some welcome laughter and unending support from start to finish and who has made me the happiest I have ever been.

Finally, the greatest support comes from my parents for their love and encouragement, and without whom I never would have had so many opportunities. My mum who has supported me the whole way, and who has always been interested and proud at every step. My dad who knew what I was capable of long before I knew it myself, I have never seen him prouder than when I told him I was doing a PhD. Sadly, despite his best efforts he did not quite make it to see me at this final stage but I have thought of him often. He remained enthusiastic about my research to the very end.

A truism: Every new level of understanding reached only reveals to us new questions of which we were unaware of before.

Contents

	Page
Statement of originality	2
Abstract	4
Acknowledgements	5
Contents	7
List of abbreviations	12
Figures	13
Tables	14
Publications and presentations arising from this thesis	15
Chapter 1: General introduction	
Statement of intent	18
1.1 Abstract	18
1.2 Spatial targeting of conservation efforts	19
1.3 Introduction to geographic profiling	20
History within criminology	21
Development of Rossmo function	22
Bayesian development	23
1.4 Applications in biology	24
GP in biology	24
Animal foraging	24
Invasive species biology	26
Epidemiology	27

A Dirichlet Process Mixture model of geographic profiling	28
1.5 Alternative spatial approaches in biology	30
Species distribution models	30
Maximum entropy models	31
Further developments	33
1.6 Summary	34
Chapter 2: Using geographic profiling to locate elusive nocturnal animals: a case study with spectral tarsiers	
2.1 Abstract	35
2.2 Introduction	36
2.3 Methods	43
Study area	44
Field data collection	44
Geographic profiling	45
Model implementation	46
Model evaluation	46
2.4 Results	47
Field study	47
Preliminary analysis	47
Full analysis	59
2.5 Discussion	51
2.6 Conclusion	54
Chapter 3: Using geographic profiling to compare the value of sightings vs trap data in a biological invasion	

3.1 Abstract	55
3.2 Introduction	55
3.3 Methods	60
The DPM model	60
Fitting σ from point pattern data	61
Model implementation	62
Testing the model	63
Comparing geoprofiles	63
Data	63
Sightings and trap data	64
Simulations	64
3.4 Results	65
Simulations	65
Profiling the trap and sightings data	65
3.5 Discussion	69
3.6 Conclusion	71
3.7 Appendix	73
Chapter 4: A spatial approach to combatting wildlife crime	
4.1 Abstract	75
4.2 Introduction	76
4.3 Methods	78
Ethics	78
General approach	79
Study area	79
Data	81

Geographic profiling: the DPM model	82
Model implementation	83
Model evaluation	84
Simulations	85
Spatial data	86
4.4 Results	86
Simulations	86
Spatial data	88
4.5 Discussion	91
4.6 Conclusion	94
Chapter 5: Human-wildlife conflict in Sumatra: using geographic profiling to identify areas of human-tiger conflict	
5.1 Abstract	95
5.2 Introduction	96
5.3 Methods	99
General approach	99
Study area/data collection	99
Geographic profiling: the DPM model	100
Model implementation	101
Model performance	102
Spatial information	102
5.4 Results	103
Can the locations where tigers have attacked humans or livestock predict where tigers are themselves hunted?	103

Can we use GP to correctly predict the locations of the livestock attacks from the locations where tigers have been seen?	106
5.5 Discussion	109
5.6 Conclusion	111
Chapter 6: Summary of research and main findings	
6.1 Abstract	113
6.2 Main findings	113
6.3 Are these geospatial layers important in criminology?	116
6.4 Implications and recommendations for policy and practice	117
6.5 Future research	118
Integrating SDM models	119
User-friendly application	119
Genetics	120
6.5 Concluding statement	120
References	122

List of abbreviations

CGT	Criminal geographic targeting
DPM model	Dirichlet process mixture model
GIS	Geographic information system
GP	Geographic profiling
MCMC	Markov Chain Monte Carlo
NERC	National Environment Research Council
QMUL	Queen Mary University of London
SVC	Savé Valley Conservancy

List of figures

	Page
Figure 2.1	Tarsier sleep tree and a typical habitat within the study site 41, 42
Figure 2.2	Buton Island. 44
Figure 2.3	Geoprofile showing the results of the preliminary tarsier analysis. 48
Figure 2.4	Geoprofile showing the results of the final tarsier analysis, including the raw probability scores. 50
Figure 2.5	Histogram showing the hitscores for all tarsier sleep trees. 51
Figure 3.1	Plots showing actual and fitted sigma values 67
Figure 3.2	Geoprofile showing results of mink analysis 68
Figure 4.1	Map of Savé Valley Conservancy 81
Figure 4.2	Boxplot showing simulated and Savé Valley data 87
Figure 4.3	Geoprofiles of Savé Valley results 89
Figure 4.4	Ring search strategy for Savé Valley results 90
Figure 5.1	Kerinci-Seblat National Park 100
Figure 5.2	Geoprofiles showing results of tiger attacking livestock data 104
Figure 5.3	Lorenz plot for tiger attacking livestock data 105
Figure 5.4	Geoprofiles showing results of tiger sightings data 107
Figure 5.5	Lorenz plot for tiger sightings data 108

List of tables

	Page
Table 2.1 Applications of geographic profiling	38
Table 3.1 Latitude and longitude of known mink dens, with hitscores	66

Publications and presentations arising from this thesis

Publications arising from thesis chapters

1. Faulkner SC, Stevenson MD, Verity R, Mustari AH, Semple S, Tosh DG, Le Comber SC. 2015. Using geographic profiling to locate elusive nocturnal animals: a case study with spectral tarsiers. *Journal of Zoology* **295**: 261-268.
2. Faulkner SC, Verity R, Roberts D, Roy SS, Robertson PA, Stevenson MD, Le Comber SC. 2017. Using geographic profiling to compare the value of sightings vs trap data in a biological invasion. *Diversity and Distributions* **23**: 104-112.
3. Faulkner SC, Stevens MCA, Romañach SS, Lindsey PA, Le Comber SC. 2018. A spatial approach to combatting wildlife crime. *Conservation Biology*. **32**: 685-693.
4. Faulkner SC, Struebig MJ, Stevens MCA, Linkie M, Martyr DJ, McKay JE, St. John FAV, Le Comber SC. (in prep). Human-wildlife conflict in Sumatra: using geographic profiling to identify areas of human-tiger conflict.

Other papers

5. Faulkner SC, Stevens MCA. 2017. Geographic profiling – Murder, maths, malaria and mammals. Chalkdust. **6**: 18-25.
6. Struebig MJ, Linkie M, Deere NJ, Martyr DJ, Milliyanawati B, Faulkner SC, Le Comber SC, Mangunjaya FM, Leader-Williams N, McKay JE, St. John FAV. (in press). Tolerating tigers: Do local beliefs offset human-wildlife conflicts? Nature Communications.

In prep

1. Faulkner SC, Makridou Z, Stevens MCA, Karamanlidis A, Le Comber SC. (In prep). Citizen science data and geographic profiling to prioritise conservation areas for brown bears in Greece.
2. Faulkner SC, Stevens MCA, Fraticelli C, Le Comber SC. (In prep). Using geographic profiling to identify the source of the ring necked parakeets in the UK.

Major presentations

Oral

1. Geographic profiling in biology. Oral presentation at British Applied Mathematics Colloquium (BAMC), University of Surrey. April 2017.
2. Geographic profiling and wildlife crime. Oral presentation at Frontiers in Natural Environment Research conference, Imperial College London, August 2017.
3. Geographic profiling – Nazis, tarsiers, and mink. Oral presentation at ZSL student conference, ZSL, London. April 2016.

Poster

1. Using geographic profiling to compare the value of sightings vs trap data in a biological invasion. Poster presentation at Perspectives of environmental change, Kings College London, September 2016.
2. Using geographic profiling to compare the value of sightings vs trap data in a biological invasion. IoZ conference, ZSL, London. October 2016.

Chapter 1: General introduction

Statement of intent

This doctoral project applies geographic profiling, first developed in criminology, to conservation data. The project focuses on developing and extending this model to incorporate more complex spatial information to improve the model performance.

1.1 Abstract

In this chapter, I review the existing literature on geographic profiling, discussing its transition from criminology to biology and the differences in application associated with this transition. I discuss the different models available, from the Criminal Geographic Targeting (CGT) algorithm introduced by Rossmo (2000) via more explicitly Bayesian approaches such as those of O'Leary (2009, 2010*a*, 2010*b*) to the Dirichlet Process Mixture (DPM) model introduced by Verity et al. (2014). I go on to review the application of geographic profiling (GP) to biological data sets such as animal foraging, invasive species biology and epidemiology and discuss the similarities between criminological data sets and biological ones in the context of this model. I then go on to explain how GP can be a useful conservation tool and detail how this PhD deals with some of the weaknesses of the previous GP implementation. Finally, I summarise the research presented in this thesis.

1.2 Spatial targeting of conservation efforts

In conservation, as in many other fields, the need for action often outstrips the resources available. Thus, a fundamental challenge in conservation biology is prioritising areas for action (Ferrier et al. 2002). This knowledge is essential for developing cost-effective and efficient targeted interventions. Historically, conservation biologists and managers have taken a subjective approach to area prioritisation (Denoël & Ficetola 2015) but, due to the current biodiversity crisis, more quantitative approaches are needed.

Animals move in space (and time), and recognizing these patterns and having an understanding of the relationship between animals and their environment is important in many fields of biology including invasive species biology, epidemiology, population ecology and conservation (Trakhtenbrot et al. 2005; Bocedi et al. 2014). Interactions between the environment and animals or plants produce complex patterns and understanding how these patterns arise and their implications is essential when assessing their conservation status (Bowler & Benton 2005). Understanding the geographic distribution and habitat preferences of particular species is also important when assessing levels of threat associated with them (Jennings et al. 2013). This is also important when dealing with issues such as climate change when there may be a need to prioritise particular areas. This prioritisation can be based on different aspects of a species' biology; for example, understanding animal movement patterns, or the habitat preferences of particular species. Thus, we need to have a grasp of species' home ranges, migration patterns, how they interact with humans or their breeding sites, for example.

Research into animal movements and behavioural patterns has been revolutionised by technological advancements such as camera trapping, radio tracking, tagging and use of harmonic radar (Jonsen et al. 2003). Techniques used to model data such as these provide a powerful tool which can help conservation scientists or managers make effective use of sparse data in relation to conservation management decisions.

Recently, with the development of Geographic Information Systems (GIS), analyses such as habitat selection have been made easier and there are numerous sophisticated models that take into account the spatial dimensions of the data such as species distribution models and maximum entropy models. In this thesis, I show how geographic profiling (GP), a statistical tool originally developed in criminology, can be adapted and used successfully to deal with conservation data sets.

1.3 Introduction to geographic profiling

Geographic profiling is an investigative technique that applies crime theory to locations of criminal activity and predicts likely offender locations to prioritise suspect lists. While the majority of murders, arson attacks and rapes will inevitably be solved by the fact that there is a connection between the criminal and their victim, problems develop when trying to solve crimes when there is no apparent connection. Investigations of these ‘stranger crimes’ result in large numbers of suspects, leading to information overload (Rossmo 2000). Research from Simon Fraser University pioneered the development of the GP technique by examining the spatial patterns of serial crime and the hunting behaviour of serial criminals (Rossmo 2000). The GP model identifies probable locations where the offender is likely to have connections, such as work, home or family ties (Rossmo 1993). Importantly, this method does not

attempt to identify a single individual, but allows the prioritisation of the list of suspects according to the resulting probability surface.

History within criminology

Studies involving spatial movement of criminals have a long-standing history, and understanding the key developmental stages of the GP approach within criminology is useful in applying the model to ecological data sets since many of the problems – such as temporal and financial constraints – will be the same. In the 1980's, for example, police investigating the case of the Yorkshire Ripper accrued 268,000 suspect names and 5.4 million vehicle registration numbers (Doney 1990), making investigating each suspect time-consuming at best and impossible at worst.

The concept of environmental criminology, developed by Brantingham & Brantingham (1981), was an attempt to draw together the different disciplines associated with solving crime. This field takes into account much broader aspects of crime, such as choices the offenders make, movement and activity patterns and how these may influence criminality (Brantingham & Brantingham 1993; Felson 2000). Prior to this, there were many varying methodologies that dealt exclusively with the spatial element of criminology (Park et al. 1921; Shaw & McKay 1942). Environmental criminology places a focus on all the environmental factors associated with the crime (e.g. time, space, criminal, victim etc.). It first looks at how the crimes relate to geographic space and secondly incorporates knowledge of how criminals structure their actions to this space (Brantingham & Brantingham 1993; Felson 2000). Brantingham & Brantingham (1981) state that the five constituents required for a crime to be committed are space, time, law (deterrent), offender and victim, and as

such without any one of these a crime cannot occur. This novel approach meant that it focused on the circumstantial elements of the crime and not just the offender.

Capone & Nichols (1976) stated that criminals will attempt to maximise their gain and minimise their effort, and that therefore a geographic range could be applied to all criminal activity – a distance decay function. This assumed that crimes would be committed less frequently further from the offender’s home or anchor point. It was this idea that Rossmo used to develop his model of geographic profiling (Rossmo 2000). Analyses of spatial point patterns are based on one or other of two principles – those grounded on distance (nearest neighbour (Clark & Evans 1954)) and those focused on area (quadrat-based methods (Greig-Smith 1952)). When Rossmo (2000) developed his GP model for criminology he used the former. These three key components (environmental criminology, distance decay, point pattern analysis) along with improvements in mapping software and computer technology led the way for Rossmo to develop his geographic profiling model (Rossmo 2000).

Development of Rossmo function

The primary algorithm used by Rossmo’s geographic profiling method is the criminal geographic targeting algorithm (CGT). Here, Rossmo combined two approaches – Pareto’s distance decay function and a buffer zone (Rossmo 1993). The idea that there is a buffer zone – an area within a certain distance of their own home in which a criminal does not commit crimes – has been a controversial topic. Yet, when Rossmo (2000) looked at the frequency distributions of journey to crime distances there were patterns emerging that backed up this theory. This traditional method looks at the distance decay around each data point as an individual entity. It creates a ‘jeopardy

surface' over geographic space, in which the height of the surface is a measure of how likely it is that a source exists at a given point. However the CGT algorithm is not mathematically robust; it does not produce a probability surface because the integral over the surface does not equal one. As Verity et al. (2014) point out, within criminology there is no need for the model to do this, as it produces ranked priority scores rather than probability. It also adds together the independent surfaces created from independent observations and develops a single surface that represents the most likely source location, which is only strictly correct if it assumed that each point forms a separate 'cluster' (Rossmo 2000; Verity et al. 2014).

Bayesian development

O'Leary (2009, 2010a, 2010b) applied a Bayesian structure to the traditional CGT algorithm. A key limitation of this method, however, is the assumption that all data points have originated from one source, and this approach can produce highly misleading results if this assumption is violated. Additionally, although the idea of a single source may be relatively robust within criminology, it is likely to be much less so when the model is applied to biological data sets (Verity et al. 2014). This problem was solved when Verity et al. (2014) produced a new Bayesian approach based on a Dirichlet Process Mixture (DPM) model and Gibbs sampler that allowed multiple sources, even when the true number of sources was unknown; however, this development only arose after the basic CGT model had first been applied to biological data.

1.4 Applications in biology

GP in biology

Having an understanding of how animals move through an environment and how they spread out from a particular central area is interesting and valuable. Modeling such movement to be able to predict this spread is common, and research is becoming increasingly sophisticated and now includes niche-based models, kernel density models and species distribution models. However, there is much less research looking at this movement backwards in time, using current locations of organisms, populations or individual animals to predict where they came from (Le Comber & Stevenson 2012). This distinction becomes very relevant when looking at topics such as animal foraging (using the locations of feeding sites to identify nests, roosts or dens), invasive species biology (using the current distribution of an invasive species to identify potential areas of introduction) or epidemiology (using known locations of diseases to identify or infer source locations) (Le Comber et al. 2011; Le Comber & Stevenson 2012).

Animal foraging

This jump from predicting future patterns to estimating how current patterns occurred in an ecological setting was the basis for Le Comber et al. (2006). In this pioneering paper they applied the original CGT GP method to investigate bat foraging patterns. The study looked at two sympatric colonies of bats (*Pipistrellus pipistrellus* and *P. pygmaeus*) in North East Scotland. The fact that these species are living within the same geographic range made the authors query how the bats were minimising foraging range competition, and whether geographic profiling models could be used to predict roost sites from the distribution of known foraging sites. They were then able to follow

this up and see if there were any differences in foraging patterns between the two species. When the foraging sites were inputted to the model as being analogous to crime sites, it was able to successfully predict the location of the roost sites for both species. Furthermore, the fitted model parameters showed that there was a significant difference in foraging patterns between the species – *P. pipistrellus* was a more generalist forager whereas *P. pygmaeus* foraged specifically along water edges. This was the first time that methods founded in criminology had been used successfully in biology.

In 2009, another paper investigated this analogous relationship between human hunting behaviour and animal foraging behaviour. Raine et al. (2009) looked at whether this technique could be applicable to bumble-bee foraging. The social nature of bees and the fact that they are known to create a forage-free buffer zone around the entrance to their nest makes them ideal organisms for the application of GP techniques. The bees were observed entering a 1m enclosed arena through different entrances and then foraging on artificial flowers filled with sucrose solution. By using experimental observations of bees within an enclosed arena, Raine et al. (2009) were able to locate the entrance (the colony nest site) within which the bees entered the flight arena by applying the same algorithm as used in Le Comber et al. (2006). It was also shown that when flowers were present at different densities the bees undertook different foraging strategies. For example, when the flowers were present at higher densities the ‘buffer zone’ decreased around the nest entrance.

Martin et al. (2009) used GP methodology to look at the dispersal patterns of white sharks (*Carcharodon carcharias*) when hunting Cape fur seals (*Arctocephalus pusillus*

pusillus) off the coast of South Africa. They showed that the patterns of shark predation were non-random, that the sharks appeared to be preying at a well-defined anchor point 100m seaward from the seals' primary entry and exit point. Counter-intuitively this location did not have the greatest chance of encountering a seal prey, and it has been suggested that it actually represents a trade-off or balance between detection, competition and capture rates. The sharks may be using an optimal spatial arrangement, whereby they offset prey density against competition or success. Additionally, the researchers created probability surfaces for sharks of different age and size and were able to show that smaller sharks had a more dispersed search area than larger sharks, suggesting competitive exclusion, or that experience refined their hunting strategy (Martin et al. 2009).

Invasive species biology

Geographic profiling has also been applied to studies of invasive species, which are now considered to be the most significant factor (second only to habitat loss) affecting not only global biodiversity and ecological integrity (Vitousek & D'Antonio 1996; Wilcove et al. 1998; Blackburn et al. 2011; Pyšek et al. 2012; Robertson et al. 2015) but also having a huge global financial impact. Pimentel et al. (2001) reported a cost of \$97 billion dollars in damages between 1906 and 1991 as a result of invasive species. Knowing that these invasions are set to continue at an increasing rate as a result of anthropogenic influence (for example an increase in potential dispersal ranges), conservationists and wildlife governments across the world have identified this as top priority (Hulme 2007; Hulme 2008; Stevenson et al. 2012). Using historical data from the British Records Centre, Stevenson et al. (2012) were able to show that GP performed significantly better than analyses using measures of spatial central tendency, and a more complex single parameter kernel density model. GP

outperformed these other techniques for 52 out of 53 datasets for invasive species. The sites colonised by the species in question were considered to be the crime sites, and used to identify a number of potential sources. By knowing the location of the sources of the invasive species, dealing with it becomes a much more manageable problem. Even when there is only limited knowledge on an invasive species' population ecology, using this method means it can be eradicated early on in the invasion.

GP was also shown to be effective when locating the origin of the invasive species *Caulerpa racemosa* in the Mediterranean. Invasions of microalgae, such as *Caulerpa spp*, are one of the main threats to marine environments, due to their generalised spreading strategy, competitive nature and high rate of vegetative propagation (Papini et al. 2010; Papini et al. 2013). Papini et al. (2013) were able to test the GP model by calibrating it against known sources of invasion and distribution data of *C. taxifolia*. It was then possible to choose appropriate parameters to allow the model to run most effectively and enabled a better understanding of invasion routes of other microalgae. The results showed that the probable sources of invasion of *C. racemosa* were two locations, which correspond to ports, implying that the most likely method of invasion was related to fishing vessels.

Epidemiology

It is surprising considering the global scale of epidemic diseases that few studies have actually attempted to identify disease sources – with the exception of the classic study undertaken by Dr. John Snow or the London cholera outbreak in 1854 (Smith 2002). Although it might seem logical that when studying disease outbreaks, identifying the source of the outbreak is the first port of call, research in these fields of epidemiology

and invasive species biology tend to work forwards in time, spatially mapping the outbreaks and predicting where future outbreaks may occur. Very few studies look backwards in time and try to locate the sources of these problems.

Due to the clustering nature of many of these diseases, and therefore their suitability for GP, Le Comber et al (2011) re-examined the original cholera data set from 1854. They showed that when the disease sites were inputted as crime sites, the infamous Broad Street pump (the source as discovered by Snow) ranked first out of 17 water pumps and was found after searching just 0.2% of the geoprofile. A second part of the study examined an outbreak of malaria in Cairo, Egypt. 139 locations where malaria cases had been recorded were used as crime sites and 59 mosquitogenic water sources were inputted as possible sources. Seven of these sources tested positive in field studies for malaria carrying vectors (*Anopheles sergentii*), and the geoprofile identified six out of seven of these sources in the top 2% of the geoprofiles (Le Comber et al. 2011). Malarial larval habitats can be very transient and diverse in their nature and therefore a novel efficient technique to identify these locations and targeting interventions at these sources is of huge interest (Carter et al. 2000; Le Comber et al. 2011).

A Dirichlet Process Mixture model of geographic profiling

As noted above (Section 1.3), neither the CGT (Rossmo 2000) or O'Leary's simple Bayesian model (O'Leary 2009, 2010a, 2010b) is entirely suitable for the application of GP to biological data, the first because of its mathematical flaws and the second because of its assumption of a single source (Verity et al. 2014). The Dirichlet Process Mixture (DPM) model of geographic profiling used in this thesis solves both of these

problems, combining the original CGT model and the simple Bayesian model within a single framework.

While the O’Leary model assumes a single source and multiplies together the distributions around each crime, the CGT assumes as many sources as crimes (even though this is never explicitly said) and sums these distributions. In contrast, the DPM model takes the mathematically correct approach of multiplying together distributions around crimes assigned to one source, and summing across all sources (Verity et al. 2014).

Constructing a geoprofile can be broken down into two related tasks: allocating crimes to clusters and finding the sources of the clusters. Solving these two problems together is difficult, but each is simple if the answer to the other is known. That is, if one knows which crimes come from which sources, finding the sources is straightforward because they are most likely to be found at the spatial means of these clusters. Similarly, if one knows where the sources are, allocating crimes to clusters is easy because crimes are most likely to originate from the closest source. The solution is to alternate between these two problems in a process known as Gibbs sampling (Geman & Geman 1984). The Gibbs sampler begins by randomly assigning crimes to clusters and then – conditional on this clustering – estimates the locations of the sources. Then – conditional on these source locations – crimes are reassigned to clusters. These two steps are repeated thousands of times with standard Bayesian Markov chain Monte Carlo (MCMC) methods until the model converges on a posterior distribution of interest. Crucially, it is not necessary to decide on the number of clusters because at

each step there is a finite, positive probability that a crime comes from a previously unseen source.

1.5 Alternative spatial approaches in biology

Recently large amounts of geospatial data have become freely available (Mayfield et al. 2017). Governmental open data initiatives are one such example where census data, crime data and various other social and economic processes are made available at no cost in the form of a geographic information system (GIS) (Brundson & Comber 2015). This thesis looks to take advantage of these resources for conservation by integrating these data into the GP model. First, though, I will consider briefly other spatial models in biology.

Species distribution models

In the last few years there has been an increase in the development of species distribution models and there are now multiple methods available, each with their own specialities (Elith et al. 2010). The type of model used will largely depend on the type of data that are available. Many spatial approaches already incorporate varying levels of spatial complexity and can be used – either singly or in combination – to identify areas of conservation importance. For example, species distribution models (SDM) (sometimes referred to as ecological niche models, habitat suitability models or simply niche-based models (Elith & Leathwick 2009; Guisan et al. 2013) are increasingly becoming the main tool to predict how climate and habitat change can impact species and quantify the species-environment relationship (Guisan & Thuiller 2005; Peterson et al. 2011). SDM models emerged when statistical analyses of ecological habitat studies, such as generalised linear models and linear multiple regression, were

combined with GIS-based environmental layers (Ferrier et al. 2002; Elith & Leathwick 2009). SDMs are correlative models and they use environmental information to explain occurrence data (Elith & Graham 2009). They evaluate the relationship between species distributions and the environment and use this information to highlight the factors that will determine the chance of a species moving to a particular location (Howard et al. 2014). There has been a dramatic increase in the popularity of this approach, which is shown in the rapidly growing scientific literature on the potential application of these models to environmental planning (Guisan & Zimmerman 2000). Despite their potentially crucial role in addressing conservation problems there is little peer reviewed evidence of the practical utility of SDMs to real world conservation decisions (Guisan et al. 2013). However, when using an SDM the choice of environmental predictor is crucial and the output can vary depending on what variables are selected (Synes & Osborne 2011; Fourcade et al. 2017).

Maximum entropy models

There are several different statistical techniques that fall under the name of SDMs and that are used to predict species distributions and modeling complexity has increased over time with advancements in technology and statistics (Elith & Graham 2009). Maximum Entropy (MaxEnt) modeling is a method used for modeling the distribution of species based on presence data alone. MaxEnt was developed in 2004 and since then it has become hugely popular as a predictive method to map species distributions and is used by governmental and non-governmental organisations (Elith et al. 2010; Fourcade et al. 2014). It uses a machine-learning algorithm that applies maximum entropy theory to presence data and environmental/geographical variables to predict species distributions (Fourcade et al. 2014). Essentially the model has two parts. First,

there is a constraint component where we must define particular environmental constraints on the probability surface such as temperature, or altitude where the species could occur. Second, the model has an entropy component where the model assumes the probability is perfectly uniform in geographic space and moves away from this distribution until the constraint component precludes this. The model ultimately produces a probability surface with a probability value in each grid cell that expresses the suitability for a species at that point dependent on the environmental variables.

Because systematic biological survey data that includes both presence and absence data are often rare, using models that can run with restricted data is advantageous. Most methods for determining species distributions rely on occurrence data, which are naturally biased towards particular areas, and pseudo-absence data, which is drawn from a wider spatial area (Philips et al. 2009; Vaclavik & Meentemeyer 2009). However, both GP and MaxEnt can produce probability surfaces with presence-only data, and can therefore be used to analyse database records from herbariums and museum archives, as well as up-to-date survey data (Philips & Dudik 2008; Elith et al. 2011). Using presence-only data means that we do not have to worry about the unreliability of absence data which can be hard to obtain accurately – the species may be present but not detected, or absent for historical reasons even if the habitat is suitable (Hirzel et al. 2002).

The fundamental differences between GP and MaxEnt is that MaxEnt produces a probability surface that can predict the most likely location of a species subject to particular conditions such as climate change, flooding or human disturbance – essentially asking where these species are moving to. This requires prior knowledge

of what environmental variables are favourable or unfavourable to the species in question. GP, on the other hand, asks where these species are coming from. Spatial information of the kind used by other SDMs can be incorporated but it is not required, making it more suitable when such data are unavailable. This difference – between where species will move in the future, and where they have come from in the past – may be thought of as starting with present-day distributions and running forwards in time (MaxEnt) or backwards in time (GP). This is an important distinction since there are cases where it is beneficial to look backwards in time – for example, using the current locations of invasive species to identify a source, identifying poachers from the locations where they have already poached, or identifying sources of infectious disease from the locations of people carrying the disease (Le Comber et al. 2011).

One major advantage of GP is its ability to produce a probability surface with very few data points but it can also cope well with very large datasets, and its Bayesian core means that more information can be added as it becomes available. In fact Rossmo (2000) states that it is possible to produce a geoprofile quickly and efficiently using as few as five data points. As such, in data sets of very rare species whereby it is difficult to obtain a lot of data, GP could potentially, still be a valuable tool.

Further developments

Despite its successful applications in biology and other fields, to date the work on GP in fields outside criminology is far from complete, and there are several areas which it does not address – for example, genetic data, temporal data and incorporating complex spatial information. It is the last of these – the spatial information – which this thesis addresses. Prior to this thesis, the application of GP to biology has assumed a spatially

homogenous landscape, making no attempt to integrate more complex spatial information (eg, altitude or land use). These habitat differences are likely to be particularly important in conservation – for example freshwater aquatic invertebrates will only be found in lakes, streams and ponds, while particular species of plants may only occur at specific altitudes.

1.6 Summary

In this thesis, I show for the first time how the GP model can be applied to conservation data and – taking the model back to its origins in criminology – to wildlife crime. Specifically I ask in Chapter 2 if GP can be used to aid conservation efforts. I use the GP model to locate sleep trees for tarsiers in dense jungle in Indonesia, using as input the locations at which calls were recorded, demonstrating how the model can be applied to locating the nests, dens or roosts of other elusive animals and potentially improving estimates of population size, with important implications for management of both species and habitats. In Chapter 3, I ask if GP is useful when responding rapidly is crucial and/or when data quality is poor? I show how spatial information obtained from citizen science data can be used to improve a study of invasive mink in the Hebrides. In Chapter 4, I turn to the issue of ‘commuter crime’ in a study of poaching in Savé Valley Conservancy (SVC) in Zimbabwe, in which although poaching occurs inside SVC the majority of poachers live outside, showing how the model can be adjusted to reflect a simple binary classification of the landscape (inside or outside SVC). Finally, in Chapter 5 I ask if GP can be improved by incorporated complex spatial information, I combine more complex land use information (point estimates of farm density) with the GP model to improve predictions of human-wildlife conflict.

Chapter 2: Using geographic profiling to locate elusive nocturnal animals: a case study with spectral tarsiers

Faulkner et al. (2015)

2.1 Abstract

Estimates of biodiversity, population size, population density and habitat use have important implications for management of both species and habitats, yet are based on census data that can be extremely difficult to collect. Traditional assessment techniques are often limited by time and money, and by the difficulties of working in certain habitats, and species become more difficult to find as population size decreases. Particular difficulties arise when studying elusive species with cryptic behaviours. Here, I show how geographic profiling – a statistical tool originally developed in criminology to prioritise large lists of suspects in cases of serial crime – can be used to address these problems. I ask whether GP can be used to locate sleeping sites of spectral tarsiers (*Tarsius tarsier*) in Sulawesi, SE Asia, using as input the positions at which tarsier vocalisations were recorded in the field. This novel application of GP is potentially of value as tarsiers are cryptic and nocturnal and can easily be overlooked in habitat assessments (for example, in dense rainforest). My results show that GP provides a useful tool for locating sleeping sites of this species, and indeed analysis of a preliminary data set during fieldwork strongly suggested the presence of a sleeping tree at a previously unknown location; two sleeping trees were subsequently found within five metres of the predicted site. I believe that GP can be successfully applied to locating the nests, dens or roosts of elusive animals such as tarsiers, potentially improving estimates of population size, with important implications for management of both species and habitats.

2.2 Introduction

The ability to accurately and reliably detect and locate individuals is vital for assessing animal population density and dynamics, habitat use and ranging patterns – crucial information that underpins the development of conservation management plans for both species and their habitat (Tyre et al. 2003). Traditional census and survey techniques are often limited by time constraints, prohibitive costs and the challenging logistics of certain habitats. Particular difficulties arise when studying elusive species with cryptic behaviours (MacKenzie et al. 2005; Neilson et al. 2013), or species with declining populations, since animals become more difficult to find as population size decreases (Chades et al. 2008; Neilson et al. 2013).

Researchers attempting to census such species typically encounter two main problems. First, the overall study area is often too large to explore exhaustively and inferences must therefore be made using a smaller sampling location, for example using a probabilistic sampling method (MacKenzie et al. 2005). Secondly, variation arises due to detectability, and research must take into account the likelihood that many individuals will go undetected even though they may be present (Chades et al. 2008; Neilson et al. 2013). In this study, I ask whether geographic profiling, a statistical method commonly used in criminology (Rossmo 2000) and increasingly being applied to biological data (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009; Le Comber et al. 2011; Le Comber & Stevenson, 2012; Stevenson et al. 2012; Papini et al. 2013; Verity et al. 2014), can be used to improve surveys of species that are rare and/or elusive.

In criminology, geographic profiling uses the spatial locations of a series of linked crimes to calculate the probability of offender residence for each location within the study area, producing a three-dimensional probability surface called a jeopardy surface; overlaid on a map of the study area, this is termed a geoprofile. Suspect sites are prioritised according to their height on the geoprofile (Rossmo 2000). Geographic profiling was designed to solve the problem of information overload when dealing with cases of serial crime, since such cases typically generate too many suspects rather than too few, and resources will often preclude a detailed investigation of them all: for example, police investigating the case of the Yorkshire Ripper accrued 268,000 suspect names and 5.4 million vehicle registration numbers (Doney 1990; Rossmo 2000). Similar constraints – lack of time, lack of resources – will obviously also often apply in conservation.

Geographic profiling has been extremely successful in criminology, and is routinely used by organisations including the Royal Canadian Mounted Police, the Bureau of Alcohol, Tobacco, Firearms and Explosives, the Los Angeles Police Department, the National Crime Agency in the UK and the United States Marine Corps (Rossmo 2012). This success has led to its application to biological data, notably animal foraging (where it can be used to find animal nests or roosts using the locations of foraging sites as input) (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009; Le Comber & Stevenson 2012), epidemiology (identifying disease sources from the addresses of infected individuals) (Le Comber et al. 2011; Le Comber & Stevenson 2012; Verity et al. 2014) and invasive species biology (using current locations to identify source populations) (Le Comber & Stevenson 2012; Stevenson et al. 2012) (Table 2.1). More recently, a version of the model based on a Dirichlet Process Mixture (DPM) model

that outperforms the Criminal Geographic Targeting (CGT) algorithm used in criminology has been developed and applied to biological data (Verity et al. 2014).

Field	Crime sites	Suspect sites	References
<i>Criminology</i>			
Murder, rape, burglary, arson	Spatial locations of linked crime sites (eg murders)	Areas likely to be associated with the offender (eg a home or workplace)	Rossmo 2000 and many others
<i>Biology</i>			
Animal foraging	Spatial locations of foraging sites	Possible nests, roosts, dens etc	Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009
Invasion biology	Spatial locations of current populations of invasive species	Areas associated with source populations	Stevenson et al. 2012; Papini et al. 2013
<i>Epidemiology</i>			
Malaria, cholera	Spatial locations of infected individuals	Vector breeding sites or other sources of infection (eg contaminated water)	Buscema et al. 2009; Le Comber et al. 2011; Verity et al. 2014

Table 2.1 Geographic profiling's application in criminology, biology and epidemiology

Here, I apply this model to a small, cryptic nocturnal primate species, the spectral tarsier, *Tarsius tarsier*. There are 10 extant species of tarsier, and eight of these are endemic to the Indonesian island of island of Sulawesi (Shekelle et al. 2008a; Shekelle et al. 2008b; Merker et al. 2005); however, there may well be many more species that have yet to be described (Shekelle et al. 2008a). Tarsiers are nocturnal, roosting in trees and crevices during the day, but sleeping sites vary depending on forest type. In most cases tarsiers will be found sleeping in the crevices of strangler figs. They are the

only extant obligatory carnivorous primates, feeding exclusively on live animals, predominantly insects. Their small size (100-150g), nocturnal lifestyle and ability to leap, almost silently, between branches in the forest make them difficult to locate in the field (MacKinnon & MacKinnon 1980). Understanding tarsier population dynamics and home ranges is critical to understanding their likely response to habitat change and destruction. However, many tarsier species are either critically endangered or listed by the International Union for Conservation of Nature (IUCN 2013) as data deficient, meaning that there are insufficient data for a proper IUCN assessment. Many tarsier species have been extensively studied, but much of this relates to aspects of their biology such as their phylogeny (for example, Merker et al. 2014 and Pozzi et al. 2014), and there is less information on their distribution or population size (IUCN 2013).

Lowland forest is thought to contain the highest diversity and abundance of animal and plant species and, with the exceptionally high biodiversity in SE Asian forests, in conjunction with carbon storage within the peatlands, deforestation of this area is of global concern. Despite this, deforestation is occurring at a rapid rate within Indonesia itself (Whitten 2000; Whitten et al. 2002; Miettinen et al. 2011*a*). Globally, Indonesia has exhibited the largest increase in forest loss between 2000 and 2012, with a low of 10,000km² lost per year to a high of over 20,000km². Further analyses of spatial resolution land maps have revealed an overall yearly decline of 1% of forest cover, with the main destruction occurring within plantations and secondary vegetation. (Miettinen et al. 2011*a,b*; Hansen et al. 2013). These recent studies show that deforestation is still occurring at a high rate in SE Asia, endangering many endemic forest species, such as the tarsier, and habitat destruction leading to the loss of suitable

habitat has led to the IUCN classifying two species of tarsier, the spectral tarsier *Tarsius tarsier* and Dian's tarsier *T. dentatus*, as vulnerable. This vulnerable status is primarily driven by habitat loss rather than population number (Shekelle et al. 2008*a,b*).

One way to estimate tarsier numbers is via identification of their roost sites (Fig. 2.1a). In practice, though, this is extremely difficult, and is usually carried out by tracking the animals using their duet vocalisations – characteristic calls made prior to dawn each morning as pairs return to their sleeping sites (MacKinnon & MacKinnon 1980). However, the short overall duration of the duetting period, and rapid sound attenuation in the dense rainforest typical of this species' habitat, mean that actually locating individuals from their duets is at best difficult and time consuming, and at worst impossible. Other methods of estimating population numbers, such as camera trapping, are arguably less labour intensive but perhaps less effective (Foster & Harmsen 2012) and in any case also present logistical difficulties due to the dense habitat (Fig. 2.1b); camera trapping is also expensive.



Figure 2.1 (a) A sleeping tree;



Figure 2.1 (b) Typical habitat within the study site.

Here, I use the DPM implementation of the geographic profiling model (Verity et al. 2014) to identify likely sleeping sites, with locations of duet vocalisations representing ‘crime sites’ and the sleeping sites analogous to an offender’s ‘anchor point’. To test this model, I take a retrospective approach, using data on vocalisations and sleeping trees that were collected in a previous study. I suggest that this approach has the potential to provide a more efficient way to identify physical locations of elusive, cryptic species such as the spectral tarsier, and to improve estimates of population size, population density and home range size for such species.

2.3 Methods

Study area

This project was conducted on Buton Island, Southeast Sulawesi, Indonesia (longitude 123°12’ E-122° 33’E and latitude 5°44’ S-4° 21’ S; Fig. 2.2). Buton has a length of 100 km and is 42 km at its widest point, and an area of approx. 4,520 km² (Priston et al. 2011). The forest where sampling took place is split into two reserves, separated only by a small road: Kakanauwe Nature Reserve (810 ha) and Lambusango Wildlife reserve (28,500 ha) (Martin & Blackburn 2010). Buton Island experiences a tropical climate, with a pronounced dry season between August and October, and a wet season between October and April; there is variable rainfall between May and July each year (Whitten 2000; Whitten et al 2002; Gillespie et al 2004). The traditional economy is based on subsistence farming, arising from slash and burn agriculture (Priston et al. 2011). The main anthropogenic impacts on the Buton rainforest are forest clearance for land use such as subsistence farming, illegal logging, asphalt mining and rattan collection (Priston et al. 2011).



Figure 2.2 Buton Island. The area in the red square represents the two forested reserves (Kakanauwe and Lambusango) where the study took place.

Field data collection

I collected the data between May and August 2013. Surveys started before 05:00 each morning, and continued until approximately 06:00. Before dawn, between 05:00 and 06:15, mated pairs of tarsiers make shrill, sexually dimorphic vocalisations that last up to four minutes as they are returning to roost; these are repeated as the animals move closer to the sleeping sites, covering in total a period of around 20 minutes (Faulkner, pers. obs.). Each morning, after the first tarsier was heard, the observer remained static and noted the time of day. I recorded the GPS data on the location of the observer, and

estimated the distance/compass bearing of the vocalising animals from the observer. I then attempted to follow the tarsiers to the sleeping tree; where this was possible, its precise location was recorded. A sleeping site was only classed as such if a tarsier or tarsiers were seen entering a tree, vine tangle or rock crevice. A total of 25 sleeping trees were identified during the study, including two located using output from a preliminary GP analysis carried out during data collection; since these were not independent of the model they were excluded from the final analysis.

Geographic profiling

The DPM model of geographic profiling, which is fully described in Verity et al. (2014), provides a mathematically robust method of estimating source locations from the spatial locations of the observed crime sites. Conceptually the method can be split into two parts. First, the crime sites are partitioned into distinct clusters, with crimes that are close to one another being more likely to end up in the same cluster. It is assumed that all points within a cluster originated from the same source, while points in different clusters originated from different sources. Second, conditional on a particular partition of the data into clusters, the posterior distribution of the unknown source locations is calculated using a method analogous to that described by O'Leary (2010a). By alternating between these two steps using standard Markov Chain Monte Carlo (MCMC) methods it is possible to reconstruct the full posterior distribution of the source locations, integrated over all possible partitions of the data into groups.

In this case, I was interested in using the spatial locations of tarsier vocalisations as input to the DPM model, and using the resulting geoprofile to construct a prioritised search strategy when looking for sleep sites. By carrying out this analysis

retrospectively, on data in which the true sleep sites (or at least a subset of the true sleep sites) are known, I could measure the effectiveness of this search strategy compared with a number of alternative approaches.

Finally, I note that although the terminology used in geographic profiling (crime sites, offender residence, etc.) is not always appropriate for use with biological data, for consistency with the criminological literature we will here refer to tarsier sleeping tree sites as anchor points or source locations and vocalisation location sites as crime sites.

Model implementation

The DPM model was implemented in R (R Core Team, 2014) using the package Rgeoprofile as described by Verity et al. (2014); this package is available from from the authors on request. Model parameters were set to default values, with the exception of sigma, the distance (in lat/long degrees) representing one standard deviation of the bivariate normal distribution centred on the source. This was set to 0.0009, corresponding to a distance of 100m at the study site location. This is equivalent to stating that 99% of the ‘crime sites’ are expected to occur within a circle of radius 300m centred on the source. This is consistent with the (admittedly sparse) literature on the dispersal distances of tarsiers, which report them as having nightly path lengths of roughly 1 km (Neri-Arboleda et al. 2002; Merker 2006).

Model evaluation

Model output can be assessed using the hitscore, the proportion of the area covering the crimes (in this case, the positions at which tarsiers were recorded, plus a ‘guard rail’ of 5 % surrounding this) in which the anchor point/s (sleep tree/s) are located.

The hitscore is calculated by dividing the ranked score by the total search area. The smaller the hitscore, the more efficient the search strategy described by the model; a hitscore of 0.5 (50%) is what would be expected from a nonprioritised (i.e. random or uniform) search (Rossmo 2000).

2.4 Results

Field study

In total, 190 tarsier call locations were recorded across 45 mornings of sampling, within both the Kakanauwe and Lambusango forest reserves. A total of 15% of the vocalisations were recorded between 05:00 and 05:30, 47% between 05:30 and 05:45, 34% between 05:45 and 06:00 and 4% after 06:00). This follows previously published data on the timing of tarsier calls (Neitsch 1999; MacKinnon & Mackinnon 1980). I identified 25 sleeping sites, mostly large strangler fig trees (Fig. 2.1a), but dense vine tangles and dead tree trunks were also used.

Preliminary analysis

Although the analysis here is retrospective, I ran the model on an early partial data set while fieldwork was still in progress. This preliminary analysis strongly suggested the presence of a sleeping tree in the NW part of the study area; two sleeping trees were subsequently found within 5 m of the predicted location (Fig. 2.3). Since these two sleeping trees were not independent of the model, they were not used to test the model and were excluded from the full analysis reported below.

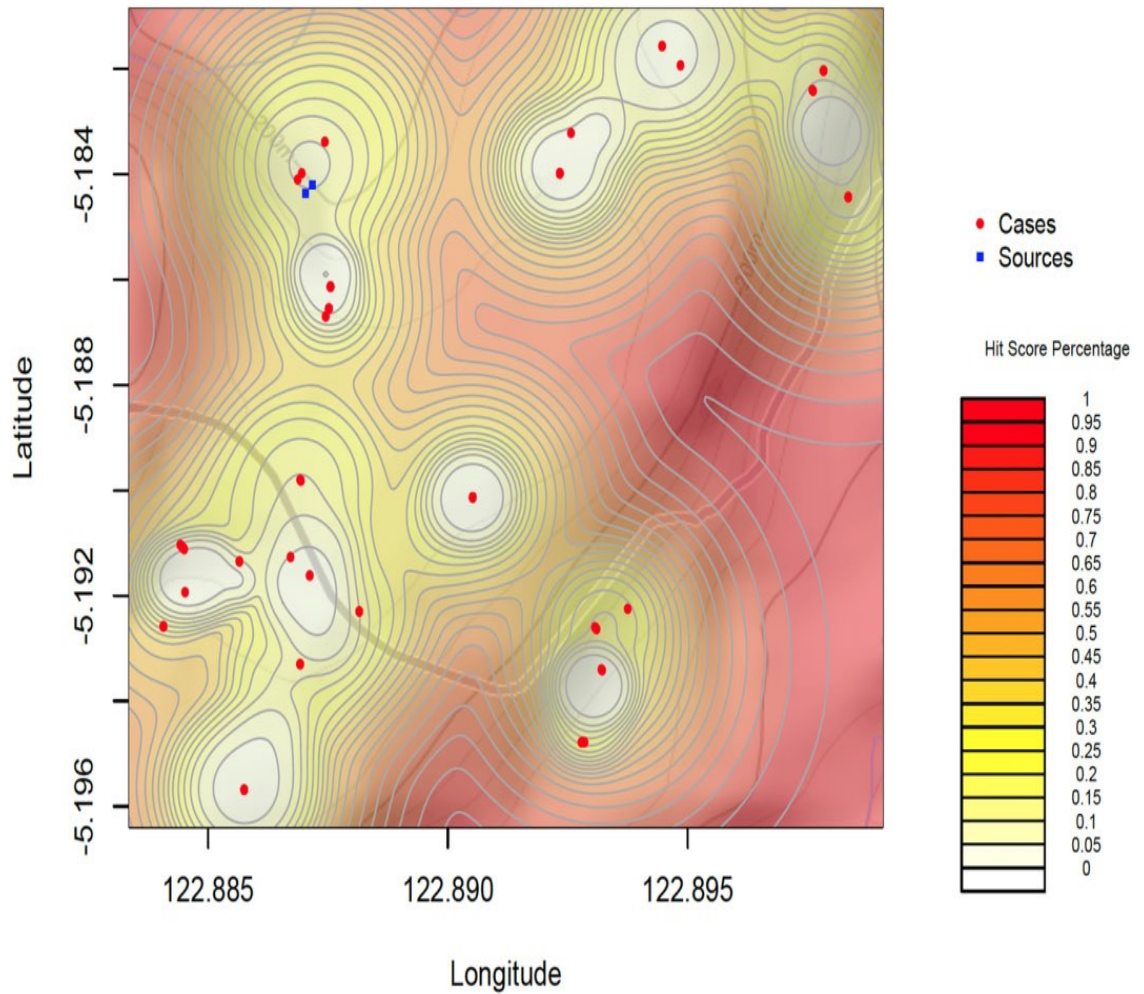


Figure 2.3 Geoprofile showing the results of the preliminary analysis. Red circles show the positions of the observer when tarsier vocalizations were recorded, and blue squares the two sleeping trees whose locations were successfully predicted in this analysis. Contours show bands of 5%, with lighter colours corresponding to higher parts of the geoprofile.

Full analysis

The mean hitscore for the 23 sleeping trees identified during field work (excluding an additional two sleeping trees identified in the preliminary analysis) was 0.16 (sd = 0.154); the median hitscore was 0.15. The model found 10 of the sleeping trees after searching less than 10% of the total map area of 3.4 km², and 16 after searching less than 20% (Fig. 2.4, Fig. 2.5).

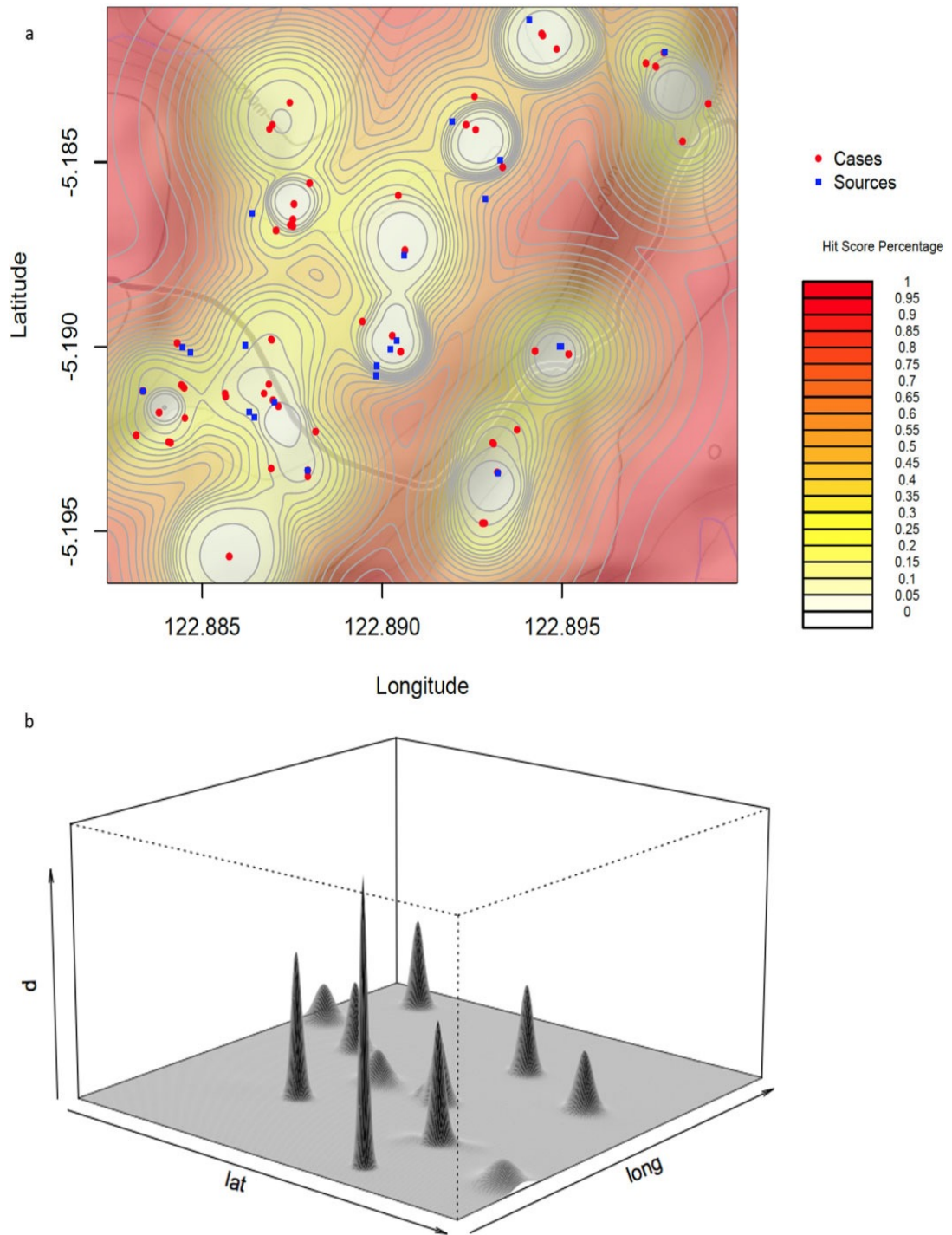


Figure 2.4 (a) Geoprofile showing the results of the final analysis. As before, red circles show the positions of the observer when tarsier vocalizations were recorded, and blue squares the sleeping trees. Contours show 5% increments. (b) The raw probability scores (P) underlying the geoprofile in (a).

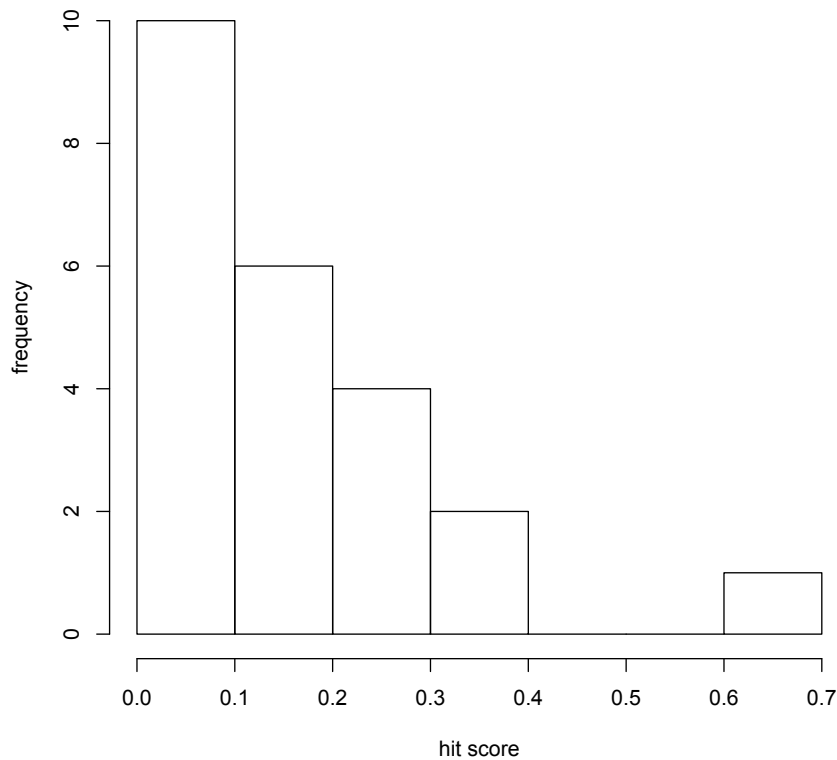


Figure 2.5 Histogram showing the hitscores for all 23 sleeping trees in the final analysis.

2.5 Discussion

These results show that the DPM model of geographic profiling described in Verity et al. (2014) can successfully be used to prioritise the search for tarsier sleeping sites. I suggest that the DPM model could be used to help locate nests, dens or roosts of other rare and hard-to-study animals, potentially improving estimates of population size in these cases, and allowing important management decisions at both the species and habitat level to be based on better data. In fact, although the analysis here is retrospective, the DPM model of geographic profiling also proved useful in practice, with a preliminary analysis successfully predicting the locations of two sleep trees that were subsequently confirmed in the field. While I would be cautious about inferring too much from such a limited example, I note that this is the first time that GP has been

used in biology as a prelude to further investigation and at the very least demonstrates the potential of this approach.

The ability to locate tarsier roost sites quickly and efficiently is of critical importance if conservation strategies are to be put into place to help conserve this species. Although there is limited destruction occurring on Buton itself in comparison to other areas, and although many tarsier species, including the spectral tarsier, often do prefer secondary forest, possibly due to increased insect abundance arising as a consequence of disturbance, as has been suggested in other contexts (Zhang et al. 2014) and locomotory aids from smaller trees (Merker et al. 2005), there will be a limit to how much disturbance this species can tolerate before population numbers begin to decline. The use of GP to better target behavioural studies and gain more accurate information on the species' ecology could thus be of considerable practical use.

Large cavity-bearing trees provide important ecological functions such as nesting and denning resources for many species (Lindenmayer et al. 2012). The overall decline in cavity-bearing trees such as the strangler fig is an emerging conservation concern, due to the importance of these trees to a variety of cavity-dwelling wildlife including tarsiers, other mammals, birds and invertebrates (Struebig et al. 2013). Forests are not recruiting these trees fast enough due to the extended period required for their replacement once they are destroyed (Lindenmayer et al. 2012). A study by Cockle et al. (2010) showed that logging was reducing the basal area of primary forests by half and therefore reducing the number of trees with suitable cavities for nesting animals. However, there are limited studies on the availability of these cavity-bearing trees in Southeast Asian rainforests (Cockle et al. 2010; Struebig et al. 2013).

There are numerous advantages to using GP rather than other more simple measures of spatial tendency such as spatial mean, spatial median and centre of minimum distance. Chief among these is the fact that the DPM model can deal with multiple sources (eg, multiple roost sites in this study), as opposed to providing a single point estimate for a single source or roost site. In criminology, Rossmo & Velarde (2008) made direct comparisons between measures of spatial tendencies and the CGT (Criminal Geographic Targeting) algorithm used in geographic profiling and showed that GP is three times more effective at detecting the anchor points of serial offenders, while in ecology, in a study looking at source populations of invasive species, Stevenson et al. (2012) found that the CGT algorithm outperformed other measures of spatial central tendency (spatial mean, spatial median, centre of minimum distance and a kernel density model) in 52 of the 53 species studied.

Although the results from this study are highly promising, geographic profiling is still in the initial stages of application to the field of biology. For example, not all species will offer a conspicuous vocalisation like the elusive tarsier; however, other measurable parameters abound. These include sightings, scat evidence or GPS data from radio collars, for example, all of which could potentially be used as ‘crime sites’ to help locate the animals’ anchor points (nests, dens or roosts). However, it is striking how well even this simple model performs: using other dispersal distributions (eg Cauchy distributions) or incorporating more complex priors based on habitat information could all add power to the model.

2.6 Conclusion

In conclusion, this study suggests that GP could be useful when searching for elusive and rare species such as tarsiers, increasing search efficiency, improving estimates of population numbers and range size and leading to more efficient use of scarce resources such as time, money and personnel.

Chapter 3: Using geographic profiling to compare the value of sightings vs trap data in a biological invasion

Faulkner et al. (2017)

3.1 Abstract

The development of conservation plans, including those dealing with invasive species, is underpinned by the need to obtain reliable and accurate data. However, in many cases responding rapidly is equally critical. These data were obtained from the Hebridean Mink Project, which was set up with the objective of removing mink from North Uist, Benbecula and South Uist. Here, I introduce an extension of the Dirichlet Process Mixture (DPM) model of geographic profiling that can be used to estimate source locations of invasions directly from spatial point pattern data without the need to specify dispersal parameters. I use this model to analyse a biological invasion of American mink (*Neovison vison*) in the Hebrides. The results suggest that sightings data – which are relatively easy and quick to acquire – can be used to capture much of the information about sources of invasive species that is obtained from the harder to acquire and more intensive trap data. These results have important implications for the development of conservation plans and, in this case, in the early stages of biological invasions, when interventions are most likely to be successful.

3.2 Introduction

In many areas of biology – for example wildlife conservation or invasion biology – the development of management strategies is hampered by the lack of reliable and accurate data (Tyre et al. 2003; Faulkner et al. 2015). However, in many cases responding rapidly is critical – perhaps more so than waiting to obtain high quality

data – since delaying action until further data is obtained can be catastrophic and can hinder opportunities to prevent species extinctions (Grantham et al. 2009; Martin et al. 2012). Thus, extracting the maximum amount of information from existing data is crucial, even when the data sources are less than ideal – for example, ad hoc sightings rather than systematic survey data. Traditional survey techniques can often be limited by time and financial constraints, along with logistical challenges that are presented by particular habitats (Faulkner et al. 2015). In addition detection of a species is not always possible until the species has reached a particular threshold, and this detection threshold will depend not only on how well established the species is but also on the monitoring method used (Myers et al. 2000; Harvey et al. 2009; Dejean et al. 2012).

One instance in which rapid action may be especially important is invasive species management. Invasive species are a major factor affecting global diversity and ecological integrity, with costs totalling millions of dollars per episode (Vitousek et al. 1996; Wilcove et al. 1998; Gurevitch & Padilla 2004; Didham et al. 2005; Blackburn et al. 2011; Pyšek et al. 2012; Robertson et al. 2015). The cost of the response to invasive species incursion is best predicted by the area over which the species has spread (Martins et al. 2006; Pluess et al. 2012; Robertson et al. 2015). Estimating this area at an early stage is therefore important to reduce costs. Damage due to invasive species is particularly prevalent within island ecosystems, where species have evolved in the absence of strong ecological competition (Nogales et al. 2006). In addition island environments are often characterised by high levels of endemism, or distinctive ecological processes that have a high potential to be disrupted by invasive mammals (Stone et al. 1994; Nogales et al. 2006). However, it is important to note that not all invasive species are deleterious, and that there are some positive

interactions that can occur which drive the diversity-invasibility relationship. These interactions are particularly important when dealing with issues such as climate change, which makes it even harder to identify invaders (Altieri et al. 2010).

The American mink (*Neovison vison*) is a widely distributed invasive carnivore now found within 28 European countries, and is one of five non-native mammals that have established feral populations in Europe (Bonesi & Palazon 2007; Roy et al. 2015). It can reach high population densities and is known to have a major negative impact on ground nesting birds and other native fauna (Roy 2011). Management efforts have proved challenging with most strategies concentrating on long-term control operations rather than eradications (Bonesi & Palazon 2007; Roy et al. 2015). Feral mink populations are also establishing in vulnerable island populations such as the Outer Hebrides in Scotland. Here they established following the release or escape from fur farms on the Isle of Lewis in the 1950s, after which they then spread south on to Harris and the Uists (Roy et al. 2015).

Early detection, rapid assessment and rapid response are crucial for preventing invasive species establishment and spread (National Invasive Species Council 2008). Even when there is only limited knowledge of an invasive species' population ecology, it can often be eradicated early on in the invasion. For this reason, financial investments are most beneficial during the early stages of the search for an invader (Holden et al. 2016). However, much of the invasive species literature is focused on costly population biology studies and determining the best control strategies under complete knowledge of invader dispersal and demography (Simberloff 2003; Epanchin-Niell & Hastings 2010; Holden et al. 2016). Given limited conservation budgets cost efficacy is important for achieving the best outcomes (Naidoo et al. 2006)

and therefore studies on dispersal processes may only be necessary when the window of opportunity for early detection and eradication is lost (Simberloff 2003). Consequently, methods that can be used quickly on limited data sets (for example citizen science data) can be important, potentially decreasing the financial strain of invasive species management

As in many similar examples, the mink invasion highlights the time and expense required to acquire high-quality trapping data. One tempting alternative is to use sightings data produced by members of the public as part of ‘citizen science’, since this is likely to be quicker, easier and less costly to collect. Citizen science is the process by which the public are involved in science as researchers (Keough & Blahna 2006; Conrad & Hilchey 2011) – although of course ‘citizen science’ in the form of the naturalist was established long before academic research became a form of employment. Advocates of citizen science in its modern sense claim a number of benefits (for example, the ‘democratisation’ of science, or the acquisition of new skills to non-professional volunteers) although fewer provide empirical evidence (Legg & Nagy 2006; Conrad & Hilchey 2011). More recently, citizen-based observation data have been viewed as an important tool for investigating spatial patterns of a wide range of taxa and natural processes (Sullivan et al. 2009; Snäll et al. 2011). However, a number of concerns have been raised about the quality of data collected via citizen science (Tyre et al. 2003). In this study, I use geographic profiling to compare the results obtained from analysis of (a) professional trap data, and (b) sightings data derived from citizen science.

Geographic profiling (GP) is a statistical technique originally developed in criminology to prioritise large lists of suspects in cases of serial crime such as murder, rape and arson by using the spatial locations associated with linked crimes to calculate the probability of offender residence for each point in the study area (Rossmo 2000); suspects are then selected for further investigation according to their height on the resultant probability surface. More recently this method has been applied to biological data, including animal foraging behaviour (finding animal nests or roosts using the locations of foraging sites as input) (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009; Le Comber & Stevenson 2012), epidemiology (identifying disease sources from the addresses of infected individuals) (Buscema et al. 2009; Le Comber et al. 2011), conservation (locating the sleeping sites of rare, nocturnal tarsiers using locations at which calls were detected) (Faulkner 2015) and invasive species biology (using current locations to identify source populations) (Stevenson et al. 2012). Using historical data from the Biological Records Centre, Stevenson et al. (2012) were able to show that GP performed significantly better both than analyses using measures of spatial central tendency, and a more complex single-parameter kernel density model. At the same time the model has been adapted to fit a Bayesian framework in the form of a Dirichlet Process Mixture (DPM) model which is designed to deal with situations with multiple sources when even the number of sources is unknown (Verity et al. 2014). The model can also be used to estimate dispersal parameters (as in Le Comber et al. 2006).

In this chapter, I ask whether GP can be applied to the type of data-poor situations that are likely to characterise the early stages of biological invasions when data quality may be low; for example, when sample sizes are small, when data are based on public

sightings data rather than systematic trap data, when locations are not precisely recorded, when identification is uncertain or when there is limited knowledge about the way in which the species in question might disperse across novel habitat. To deal with this last point, I develop (in collaboration with Robert Verity and Mark Stevenson) a novel mathematical extension to the DPM model that requires minimal input from the user about the likely dispersal of the invading species, testing this on both simulated and real-world data sets. Specifically, I ask: (i) does the model extension introduced here lead to correct posterior inference of dispersal parameters when tested with simulated data? (ii) can this novel implementation of the DPM model be used to locate known den sites associated with the invasion of American mink in the Hebrides, using either trap data or sightings data as ‘crime sites’?; (iii) how does the geoprofile derived from public sightings data compare to that derived from professional trap data?

3.3 Methods

Note that although the terminology used in GP – for example crime sites, offender residence – is not always appropriate for use with ecological and biological data, for consistency this paper will refer to mink sightings/captures as crime sites and mink dens/breeding grounds as source locations. No negative connotations should be inferred from this.

The DPM model

The model used here is an extension of the DPM model described by Verity et al. (2014). The original method works by breaking down the difficult problem of estimating multiple unknown source locations into two conceptually simpler sub-

problems. The first step is to estimate which “crimes” originated from which sources based on the locations of these sources – which are assumed known at this stage. Observations are more likely to be allocated to sources that are close by, leading to a clustering of observations into a finite number of groups, where members of the same group are assumed to have originated from the same source. The second step is to re-estimate source locations based on the clustering, which is now assumed known. The final model implementation uses Bayesian Markov chain Monte Carlo (MCMC) to alternate between these steps many thousands of times until the algorithm converges on the posterior distribution of interest, with convergence being measured via the Gelman-Rubin diagnostic statistic (Gelman et al. 2004). These posterior draws can be used to generate a probability surface of the unknown sources (i.e. a geoprofile) together with an estimate of the number of sources from which the observed data originated. Further details of the method can be found in Verity et al. (2014).

Fitting σ from point pattern data

Note: The two following paragraphs were written in collaboration with Dr. Robert Verity and Dr. Mark Stevenson.

In the original implementation of the model described in Verity et al. (2014) (and used in Faulkner et al. (2015)) the dispersal distance was defined by the fixed parameter σ , which had to be set by the user. σ represents the standard deviation of the bivariate normal distribution describing dispersal from a central source. Previously, this was in units of degrees of latitude/longitude; in the new version of the model presented here, I use km to avoid differences in distance between 1 degree latitude and 1 degree longitude. So, a σ value of 1 km will imply that 39% of dispersal events occur within

1 km of the source, 87% within 2 km and 99% within 3 km. In many cases, existing knowledge of the species in question will suffice to provide a reliable estimate of σ . However, in the case of invasions this may be more challenging, due to the uncertainty associated with a non-native species moving into a new and atypical habitat. In such situations the ability to estimate σ from point pattern data alone will clearly be advantageous, and it is this extension to the DPM model that I introduce here.

The underlying mathematics of this extension are described in Section 3.7, but in brief the method uses a Gibbs sampler to estimate the locations of the sources and the value of σ simultaneously. I assume an inverse-gamma prior on σ^2 , which has the advantage of being flexible enough to capture broad uncertainty in σ , while also being mathematically convenient due to being conjugate to the likelihood. The user is required to specify two parameters that control the mean and shape of this distribution. Generally, this approach is to use values such that the prior is extremely diffuse and heavy-tailed, and thus is relatively uninformative of σ .

Model implementation

The DPM model described here was implemented in R (R Core Team 2014) using a modification of the package Rgeoprofile as described by Verity et al. (2014); this package is available at <https://github.com/bobverity/Rgeoprofile>.

Testing the model

Model output can be assessed using the hitscore: the proportion of the area covering the crimes (here, as is common in criminology, the positions of the ‘crimes’, plus a ‘guard rail’ of 5% surrounding this). The hitscore is calculated by dividing the ranked score of each grid square within the total search area by the total number of grid squares. The smaller the hitscore, the more efficient the search strategy described by the model; a hitscore of 0.5 (50%) is what would be expected from a non-prioritized (i.e. random or systematic) search (Rossmo 2000).

Comparing geoprofiles

For both analyses, the study area was defined as the rectangular area encompassing all trap and sightings data, plus a ‘guard rail’ of 5%. Since correlations between geoprofiles produced from the two data sets would be artificially inflated by the low scores assigned to the sea in both cases (where there are neither sightings nor trap data), even if the two approaches produced quite different results for the areas of interest (the land), I took the conservative approach of calculating Pearson’s product moment correlation coefficients after restricting the analysis to the top 50% of the geoprofile in each case (largely corresponding to land, and shown in Fig. 3.2). This ensured that distributions were compared in terms of high priority regions.

Data

The data were obtained from the Hebridean Mink Project, which was set up with the objective of removing mink from North Uist, Benbecula and South Uist, whilst also trying to reduce the density from South Harris. Trapping started in 2001, and data were collected until 2005. The area used in the study included a mixture of saltwater and

freshwater habitats – 1,116km of coastline, 2,416km of loch shore and 189km of freshwater rivers and streams. Full details of the project can be found in Roy et al. (2011, 2015).

Sightings and trap data

The data consisted of 125 sightings identified by the public (in many cases, crofters) over four years (2002-2005) and 409 total captures over five years (2001-2005, 141,498 trap nights). These data covered both Uist and Harris. Locations for 15 known dens (sources) were used for assessing the performance of the model.

Simulations

To test the accuracy of the model, 1000 simulated data sets were created, with 1-5 sources (drawn from a uniform distribution between -4 and -3° longitude, and 56 and 57° latitude), with 5-10 crimes per source, again drawn from a uniform distribution. The actual value of σ was set independently for each simulation, drawing from a uniform distribution ranging from 1-10 km. The DPM model was run on each data set, with the prior expectation of σ chosen independently but in the same way, and using the same range of values; these values, combined with the area from which the sources were selected, meant that crimes from different sources often overlapped, providing a more challenging (and more realistic) test for the model. The shape parameter of the inverse-gamma prior on σ^2 was set to 2, corresponding to a weakly informative distribution. In each simulation the prior expectation of σ , the posterior mean value of σ , and the true value of σ were recorded.

3.4 Results

Simulations

Across 1000 replicates, the fitted value of σ was most strongly affected by the actual value, although the prior expectation of σ , the number of sources and the number of crimes also had significant effects; the fitted coefficient for the relationship between actual and fitted values was 0.92, with an intercept of -1.35 (ANOVA: prior expectation $F_{1,995} = 114.5$, $p < 0.0001$; sources $F_{1,995} = 15.8$, $p < 0.0001$, crimes $F_{1,995} = 123.2$, $p < 0.0001$; actual value $F_{1,995} = 2799.1$, $p < 0.0001$) (Fig. 3.1). Model error, expressed as the difference in km between the actual σ value and the fitted value, increased with number of sources, with a slight tendency to underestimate σ when there was a single source and to overestimate with multiple sources ($F_{4,995} = 21.5$, $p < 0.0001$) (Fig. 3.1c).

Profiling the trap and sighting data

The geoprofiles for the trap and sightings data are shown in Fig. 3.2 and the hitscores for known den locations in Table 3.1. Using the trap data, the mean hitscore for the 15 known den sites was 0.06 (sd 0.058) and the median was 0.04, with the model finding all of the known dens by searching just over 20% of the study area, 13 by searching 10% and four less than 1%. Using the sightings data the mean hitscore was 0.14 (sd 0.112; median 0.10), with the model finding eight of the dens by searching 10% of the total search area. Although the hitscores for the trap data were lower (paired t-test: $t = 3.1$, $df = 14$, $p = 0.007$), overall the geoprofiles derived from the trap data and the sightings were similar (Pearson's product moment correlation: $r = 0.42$, $p < 0.0001$), despite the discrepancy in the size of the two data sets (409 data points for the trap data versus 125 sightings).

Latitude	Longitude	Hitscore		Area Searched (km ²)	
		Trap data	Sightings data	Trap data	Sightings data
-6.8645701	57.794056	0.000	0.014	0.2	67.5
-7.1439991	57.652154	0.003	0.058	14.7	286.3
-7.0150114	57.758104	0.003	0.164	15.4	811.0
-7.1384491	57.657167	0.004	0.070	18.1	347.2
-6.8578971	57.789041	0.013	0.038	64.4	185.1
-7.1339153	57.663083	0.020	0.093	99.8	461.0
-7.1550095	57.665138	0.040	0.123	195.8	607.0
-7.3170788	57.177832	0.044	0.060	218.4	298.2
-7.1289061	57.663277	0.051	0.100	249.2	491.7
-7.0513014	57.655395	0.057	0.127	279.6	625.9
-7.289443	57.701967	0.064	0.265	317.4	1308.2
-7.3499298	57.271959	0.098	0.373	484.4	1838.9
-7.3498816	57.27197	0.098	0.373	484.4	1838.9
-7.1242865	57.672283	0.141	0.166	693.3	816.8
-7.289001	57.163076	0.203	0.082	1002.9	406.7

Table 3.1. Latitude and longitude coordinates of known dens, with hitscores of search area (km²) for the two analyses.

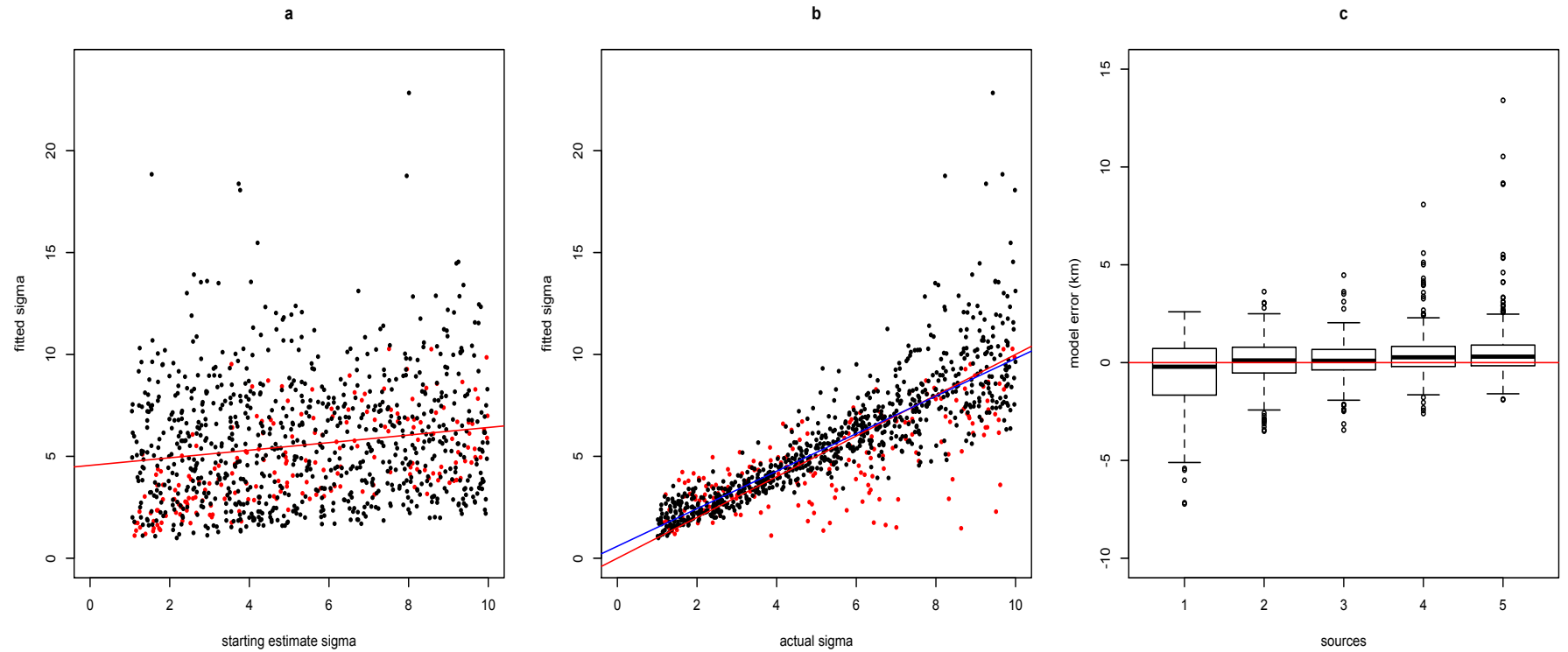


Figure 3.1. (a) Prior expectation and actual value of σ (in km), and (b) fitted sigma versus the actual sigma. (c) shows the error in σ for one to five sources; positive errors represent fitted values greater than the true value, and negative errors fitted values lower than the true value. In both (a) and (b), replicates from single sources are shown in red, and from multiple (2-5) sources in black, with fitted lines in red. The red line in (b) shows the 1:1 line and the blue line the fitted line.

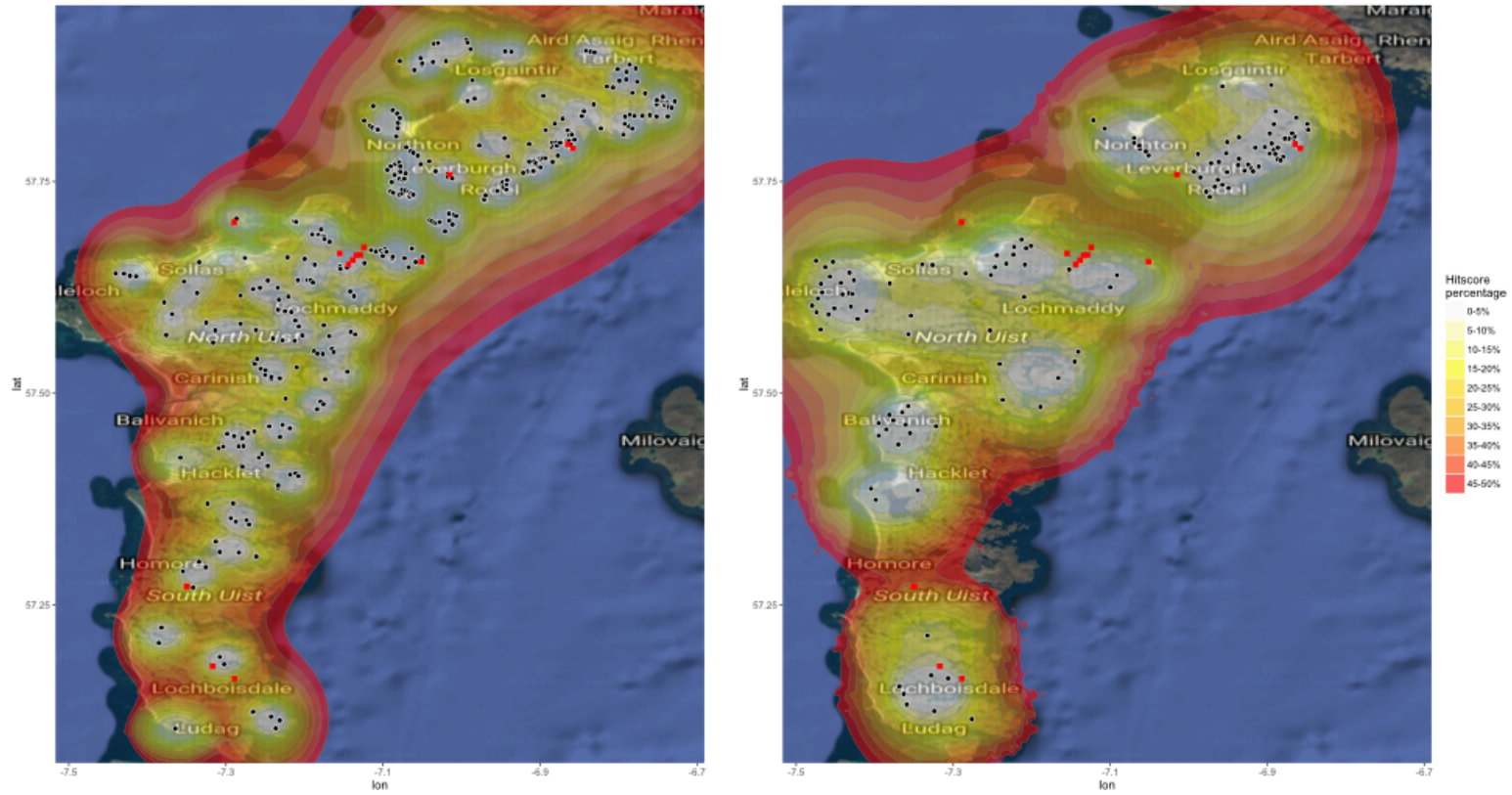


Figure 3.2. Geoprofiles showing the results of the GP analysis. Black dots show the locations of successful traps in (a) and the locations of mink sightings in (b). Contours show the top 50% of the geoprofile in both cases, in bands of 2%, with lighter colours corresponding to higher parts of the geoprofile. Sources (known den locations) are shown by the red squares

3.5 Discussion

In this chapter I show for the first time that citizen science data – in the form of sightings – and the DPM model could together be used to target interventions and estimate the area required for control in the early stages of an invasion, when control efforts are most likely to be effective. Although the hitscores for the trap dataset were significantly lower, the sightings data produced fundamentally the same result, as shown by the correlation between the two geoprofiles, even when the analysis was restricted to compare only the highest – and therefore most relevant – parts of the two surfaces. This was despite the fact that the trap dataset was both professionally acquired and much larger than the sightings dataset. In addition, our extension of the DPM model enhances the model's utility in data poor situations that might characterise the movement of an invasive species into a novel habitat.

GP's origins lie in criminology, where it was developed as a tool to prioritise suspect lists in cases of serial crime, since either time or resources often preclude detailed investigation of the large lists of suspects that typify such investigations (for example, 268,000 suspects and 5.4 million vehicle registrations in the case of the Yorkshire Ripper (Doney 1990)). Similar problems are also encountered in conservation, since time and resources are likely to be similarly limited (Stevenson et al. 2012). Added complications that apply in both criminology and conservation are (i) sample sizes may be low (models depending on large sample sizes are obviously useless when data points are murders), and (ii) that data may be incomplete (i.e. there may be unknown or unrecorded crimes, or populations of an invasive species). GP is generally robust to both of these problems, and can be successful on as few as five locations (Rossmo 2000). However, in conservation if not criminology, there may be the option of using

larger data sets even if the data are of lower quality – for example, data derived from citizen science rather than rigorous survey or trap data. Here, I show that in the case of the American mink in the Hebrides GP works very well at identifying known den sites using trapping data or when using sightings and – crucially – that the results of each are similar. The similarity of the results is particularly impressive since the sightings data set was smaller than the trap data, in addition to being (presumably) of lower quality.

In this chapter I also introduce an extension to the DPM model of GP originally described in Verity et al. (2014). This original paper placed GP within a Bayesian framework, while also allowing for multiple sources, even when the number of sources is unknown (previous Bayesian models had explicitly assumed a single source) (O’Leary 2009, 2010*a*, 2010*b*, 2012). However, the model did require the user to specify a fixed parameter, σ , describing the standard deviation (in that version of the model, in decimal degrees of latitude/longitude) of the bivariate normal distribution describing dispersal from the source. The version of the DPM model introduced here removes this limitation, and requires only an expected value of σ (now in km) and a prior shape parameter.

There are a number of other ways in which the basic model used here could be extended. Rather than a normal dispersal distribution, other types of distributions (for example, the Cauchy distribution or Student’s *t* distribution) could be used, and in particular distributions that are biased in particular directions, to account for the fact that dispersal may be easier in some directions than others, perhaps because of prevailing winds etc. In addition, more informed priors could be used – in the data set

described here, for example, setting a zero prior on the sea, eliminating the need for the (cruder) post hoc correction used here.

The results of the simulations show good correspondence between the actual σ value and the fitted value (Fig. 3.1b). The prior expectation of σ had a significant effect on the model (Fig. 3.1a), although eventually I would expect the prior to be overwhelmed by the data as the number of ‘crimes’ increased (in simulations there were as few as five crimes); in any case, this effect was minor compared to the effect of the actual σ value, with an F statistic of 114.5 compared to 2799.1. The model had a slight tendency to overestimate σ as the number of sources increased, almost certainly due to overlapping distributions of crimes from closely adjacent sources leading the model to select a single, larger cluster (as Fig. 3.1c shows, simulations with single sources typically have lower fitted values of σ than simulations with multiple sources). Again, this effect was small compared to the effect of the true value, with an F statistic of 15.8.

3.6 Conclusion

Together, the use of sightings data and the removal of the requirement to specify σ in the DPM model suggest that this approach could potentially be of use in other biological invasions, particularly in the early stages when interventions are likely to be most successful but when data quality may be low and the population poorly characterised. With the cost of invasive species running from millions to billions of dollars per occurrence (Vitousek et al. 1996; Wilcove et al. 1998; Gurevitch & Padilla 2004; Didham et al. 2005; Blackburn et al. 2011; Pyšek et al. 2012), it is clear that tools helping to inform management strategy in real time are highly advantageous, and

I suggest that this implementation of the DPM model provides one such tool. In particular, targeting control efforts in this way will reduce the area required for control, making interventions more efficient and more cost effective.

3.7 Appendix

Note: This appendix was written by Dr. Robert Verity, it is included here for completeness.

Here I build on the basic DPM model that is described in detail in the appendix of Verity et al. (2014) to allow for variable dispersal distance. Following the same notation, and assuming that crime sites \mathbf{x}_i (for $i \in 1, \dots, n$) are drawn from a bivariate normal distribution around their corresponding source locations $\boldsymbol{\mu}_{c_i}$, we obtain the following likelihood:

$$f(\mathbf{x} | \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^n \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_{c_i})^T (\mathbf{x}_i - \boldsymbol{\mu}_{c_i})\right\} \quad (1)$$

(where \mathbf{x}_i and $\boldsymbol{\mu}_{c_i}$ are column vectors, each with two rows). In order to infer σ^2 I must first choose a suitable prior for this parameter. I follow the convention of using an inverse-gamma prior on σ^2 :

$$f(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\delta+1} \exp\left\{\frac{-\beta}{\sigma^2}\right\}. \quad (2)$$

This prior has the significant advantage of being conjugate to the likelihood; if the prior distribution on σ^2 is written *Inv – Gamma*(δ, β) then the conditional posterior is given by *Inv – Gamma*($\delta + n, \beta + \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_{c_i})^T (\mathbf{x}_i - \boldsymbol{\mu}_{c_i})$). By drawing a new value of σ^2 from this conditional posterior distribution at each iteration of the MCMC – conditional on known source locations ($\boldsymbol{\mu}$) and cluster allocation (\mathbf{c}) at that point in the algorithm – we obtain a series of draws from the complete posterior distribution of σ^2 . The posterior distribution of source locations (i.e. the geoprofile) can then be considered to be integrated over the prior on σ^2 . I note that in the original Verity et al. (2014) application $\boldsymbol{\mu}$ was integrated out analytically, meaning only the

group allocation needed to be updated as part of the MCMC sampler. Here I can still make use of this gain in efficiency by integrating over $\boldsymbol{\mu}$ when updating the group allocation, however, $\boldsymbol{\mu}$ values must be realized immediately prior to updating σ^2 , as the posterior on σ^2 is conditional on these values.

Finally, when defining prior parameters we choose to work in the space of σ rather than σ^2 . The parameters set by the user are the expectation of σ , denoted ε , and the shape parameter δ . The distribution of σ can then be written

$$f(\sigma) \propto \left(\frac{1}{\sigma}\right)^{2\delta+1} \exp\left\{-\left(\frac{\varepsilon \Gamma(\delta)}{\sigma \Gamma(\delta - \frac{1}{2})}\right)^2\right\}.$$

The advantage of this setup is that ε has an intuitive interpretation as the expected dispersal distance. The prior variance of σ is given by

$$\text{Var}(\sigma) = \varepsilon^2 \left(\frac{\Gamma(\delta - 1)\Gamma(\delta)}{\Gamma(\delta - \frac{1}{2})^2} - 1 \right).$$

Chapter 4: A spatial approach to combatting wildlife crime

Faulkner et al. (2018)

4.1 Abstract

Poaching can have devastating impacts on animal and plant numbers, and in many countries has reached crisis levels, with illegal hunters employing increasingly sophisticated techniques. In this chapter, I show how geographic profiling – a mathematical technique originally developed in criminology and recently applied to animal foraging and epidemiology – can be adapted for use in investigations of wildlife crime, using data from an eight-year study in Savé Valley Conservancy, Zimbabwe that in total includes more than 10,000 incidents of illegal hunting and the deaths of 6,454 wild animals. Using a subset of these data for which the illegal hunters' identities are known, I show that the model can successfully identify the illegal hunters' home villages using the spatial locations of hunting incidences (for example, snares) as input, and show how this can be improved by manipulating the probability surface inside the Conservancy to reflect the fact that – although the illegal hunters mostly live outside the Conservancy, the majority of hunting occurs inside (in criminology, 'commuter crime'). The results of this analysis – combined with rigorous simulations – show for the first time how geographic profiling can be combined with GIS data and applied to situations with more complex spatial patterns – for example, where landscape heterogeneity means that some parts of the study area are unsuitable (e.g. aquatic areas for terrestrial animals, or vice versa), or where landscape permeability differs (for example, forest bats tending not to fly over open areas). More broadly, these results show how geographic profiling can be used to target anti-poaching interventions more

effectively and more efficiently, with important implications for the development of management strategies and conservation plans in a range of conservation scenarios.

4.2 Introduction

Illegal hunting represents one of the most severe threats to wildlife worldwide (Ripple et al. 2016). The severity of the threat is such that a growing number of species are suffering population declines and becoming threatened with extinction (Ripple et al. 2015, 2016). In Africa, wildlife hunting is conducted to obtain bushmeat for subsistence, as well as wildlife products such as ivory, rhino horn, pangolin scales and leopard skins for the international (and in some cases, local) trade (Biggs et al. 2013; Blanc et al. 2013; Lindsey et al. 2013, 2017). The resources available to tackle illegal hunting are severely limited, with the effect that protecting wildlife populations in the vast landscapes in which they occur is extremely challenging (Mansourian & Dudley 2008; Lindsey et al. 2016). There is an urgent need to develop technological solutions to provide law enforcement agencies with the edge over illegal hunters.

Although illegal hunting is prevalent even in times of relative peace, it can intensify during times of political instability (Cumming 2004). In Zimbabwe, illegal hunting began to rise with the onset of the land reform programme in which subsistence farmers were re-settled onto private farms and wildlife ranches (du Toit 2004). In 2001, settlers began to invade a large wildlife area in southeastern Zimbabwe, the Savé Valley Conservancy (SVC). Financial losses realized through illegal hunting in SVC were calculated to be at least USD 1 million per year (Lindsey et al. 2011), highlighting that the crisis is as much an economic problem as a conservation one.

In this chapter, I show how geographic profiling (GP) can be adapted for use in investigations of wildlife crime, using data from an eight-year study in Savé Valley Conservancy, Zimbabwe that includes more than 10,000 incidents of illegal hunting and the deaths of 6,454 wild animals.

Geographic profiling is a statistical technique that was originally developed in criminology to prioritise large lists of suspects in cases of serial crime such as murder, rape and arson (Rossmo 2000). More recently, the model has been successfully applied to epidemiological and biological data sets such as locating animal roost and nest sites using as input their foraging locations (Le Comber et al. 2006; Buscema et al. 2009; Martin et al. 2009; Raine et al. 2009; Le Comber et al. 2011; Le Comber & Stevenson et al. 2012; Verity et al. 2014; Faulkner et al. 2015, 2017). In criminology, the model uses the locations of linked crimes to calculate the probability of offender residence for each point within the study area. These probabilities are then ranked to produce a geoprofile, with suspects higher on the profile investigated first.

Despite the success of GP in a range of disparate fields within biology, the model's application has to date largely ignored a great deal of spatial complexity and differences in habitat, many of which are likely to be important (as a simple example, freshwater aquatic invertebrates will generally be restricted to ponds, lakes and streams). The particular case examined in this chapter provides another good example of this, as although the illegal hunters mostly live outside SVC, the animals they are hunting are found almost exclusively inside the Conservancy. In criminology, such a scenario results in what would be referred to as 'commuter crime'. In contrast to the normal assumptions of the model, in which the majority of offenders commit crimes

close to their anchor point (usually a home or workplace) (Brantingham & Brantingham 1981; Meany 2004), in commuter crime offenders must travel some distance to specific locations to commit their crimes because of the clustered nature of potential crime sites (for example, opportunities for high-value shoplifting are likely to occur in city centres, with few or no opportunities for the criminal near their home) (Canter & Larkin 1993).

In the case study presented here, I address the issue of commuter crime by a post-hoc manipulation of the geoprofile in which I adjust the model probabilities inside the Conservancy in a variety of ways to reflect the fact that the illegal hunters will in most cases live outside SVC. This chapter, thus has two main aims. First, I examine how an approach originally developed in crime science can be applied to wildlife crime. Second, I extend the GP method to show how post-hoc adjustment of the resulting geoprofile can improve model performance. Specifically, I ask: (1) Can geographic profiling be used to identify illegal hunters from hunting incidences alone? (2) Can geographic profiling be improved by incorporating geospatial data, in this case to deal with the issue of commuter crime?

4.3 Methods

Ethics

The data relating to the incidents of illegal hunting are a subset of data in an earlier study (Lindsay et al. 2011). As part of that study, anti-poaching scouts from the ranches comprising SVC were interviewed on a monthly basis and the locations of incidents of illegal hunting (eg poaching, snares) recorded. For a subset of these incidents, illegal hunters had been observed or caught as part of their routine patrols.

Where the hunter was known to the scouts, the location of their town or village (and not individual addresses) was recorded; it is these data that our study utilises. Thus, none of the data in this paper can be used to identify individuals (particularly since the data were collected 12 years ago). No additional data or analysis were shared with the police or with anti-poaching scouts.

General approach

This study examines how an approach originally developed in crime science can be applied to wildlife crime, and extends the GP method to show how post-hoc adjustment of the geoprofile can improve model performance. In the particular case I examine here, the majority of incidents of illegal hunting originate outside SVC, even though the incidents themselves mostly occur inside the conservancy. To address this, I first divide the geoprofile— a matrix describing, for each point in the study area, the probability that there is a source at that point —into areas inside SVC and outside SVC using a shapefile. I then adjust our estimate of the probability of source location inside the conservancy to reflect our belief that source locations within the conservancy are less likely than sources outside. I consider a range of manipulations in which I reduce the probability of source location for points inside the conservancy by factors from 0.1 to 0.000001; I also consider the extreme case where the probability of source location is set to zero inside the conservancy.

Study area

The Savé Valley Conservancy (20°24'48.10"S, 32° 8'19.61"E) is a large wildlife area (3,450 km²) in arid, southeastern Zimbabwe (Fig. 4.1). The Conservancy is comprised of 26 individual wildlife ranches held in ownership by private, government, and local

community entities. While there are no internal fences between ranches, 350 km of double, perimeter fencing has served as a boundary between wildlife within SVC and the surrounding high-density human settlements. The Conservancy is home to an abundance of wildlife species such as impala, zebra, wildebeest, buffalo, giraffe, elephant, wild dog, and both black and white rhino.

In 2001, trends of increasing wildlife populations within the Conservancy began to reverse with the implementation of Zimbabwe's land reform programme. Subsistence farmers began to settle within SVC and removed large tracts of perimeter fencing, enough to make over 400,000 wire snares (Lindsey et al. 2009) which are used to catch wildlife for bushmeat. In Zimbabwe, hunting using snares is prohibited by law (Trapping of Animals (Control) Act [Chapter 20:21]), as is the possession or sale of illegally obtained bushmeat (Parks and Wildlife Act [Chapter 20:14]).

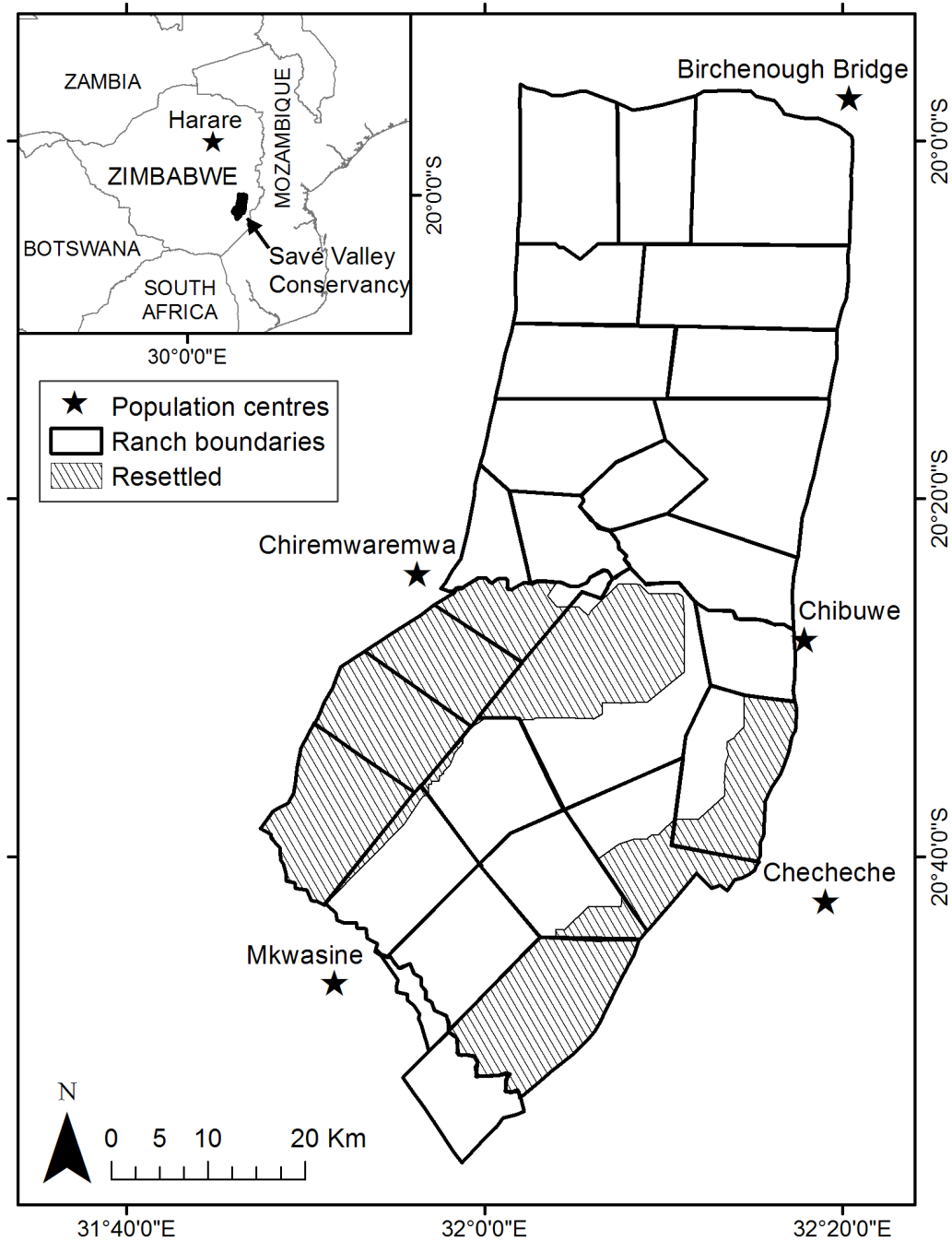


Figure 4.1. Map of Savé Valley Conservancy in southeastern Zimbabwe.

Data

Illegal hunting data were collected between August 2005 and July 2009. My collaborators received data monthly from anti-poaching managers on each ranch in SVC. I compiled: number of illegal hunting incidents, number of hunters and dogs, number of illegal hunters caught (or shot in the case of dogs), how they were caught,

number of snares recovered, number/species/gender/age of animals killed in each snare as well as the status of carcasses; i.e. recovered by illegal hunters, recovered by ranch, rotten or scavenged. Data on wildlife killed included records of observations of carcasses in snares, carcasses found in the possession of hunters, at their homes or in hunting camps, or from identifiable hair or body parts left in snares. The location of illegal hunting incidents was indicated by the anti-poaching managers on 1:50,000-scale maps overlaid with 1-km grid squares, with an average error of approximately 1 km (Lindsey et al. 2011). For this analysis I used a subset of these data for which the illegal hunter identities are known. This included 151 hunting incidents, and a total of 47 known illegal hunters. The most hunting incidents per individual was 32, with most individuals hunting just one time. The method of hunting varied: snares (66), dogs (60), fishing (13), snares and dogs (3), and other (9).

Geographic profiling: the DPM model

The DPM model is described fully in Verity et al. (2014) and extended in Faulkner et al. (2017). In brief, though, it can be explained as follows. Constructing a geoprofile can be broken down into two related problems – allocating ‘crimes’ to clusters, and finding the sources of the clusters. Solving these two problems together is difficult, but each is simple if the answer to the other is known. That is, if we know which crimes come from which source, finding the sources is straightforward, since they are most likely to be found at the spatial means of these clusters. Similarly, if we know where the sources are, allocating crimes to clusters is easy, since crimes are most likely to originate from the closest source. The solution is to alternate between these two problems in a process known as Gibbs sampling (Geman & Geman 1984). The Gibbs sampler begins by randomly assigning crimes to clusters, and then – conditional on

this clustering – estimates the locations of the sources. Then – conditional on these source locations – it reassigns crimes to clusters. These two steps are repeated many thousands of times using standard Bayesian Markov Chain Monte Carlo (MCMC) methods until the model converges on a posterior distribution of interest. Crucially, it is not necessary to decide on the number of clusters, since at each step there is a finite, positive probability that a crime comes from a previously unseen source.

Model implementation

The DPM model described here was implemented in R (R Core Team 2014) using version 2.0.0 of the package *Rgeoprofile* introduced by Verity et al. (2014) and extended in Faulkner et al (2017); this package is available at <https://github.com/bobverity/Rgeoprofile>. Models settings are explained in detail in Verity et al. (2014). Here, the settings used were *sigma_mean*=1, *sigma_squared_shape*=2, *samples*=10000, *chains*=10, *burnin*=1000. Broadly speaking *sigma* represents the standard deviation (in km) of the dispersal distribution around the source, and *sigma_mean* is the initial prior on this. The parameter *sigma_squared_shape* relates to the shape parameter of the inverse-gamma prior on *sigma*, with a value of 2 corresponding to a weakly informative distribution; see Faulkner et al. (2017) for details of the underlying mathematics. These settings correspond to a diffuse prior on *sigma* of 1km, implying that 39% of the poaching events occur within 1km from the source, 87% within 2km and 99% within 3km; however, the model will disregard this prior if the data warrant it. A value of 1km is a value typical of human patterns of movement (Rossmo 2000). The parameters *samples*, *chains* and *burnin* are all standard parameters relating to the MCMC.

Model evaluation

The model output is assessed in two ways. The model's performance in finding an individual source can be calculated using the hitscore. The hitscore is the proportion of the total area covering the crimes (in this case the hunting incidents) that has to be searched before that source is located. This is calculated by ranking each grid square within the total search area and dividing the ranked score of the grid square in which the source is located by the total number of grid squares to give a value between 0 and 1: the smaller the hitscore the more efficient the search strategy. For example, a suspect site with a hitscore of 0.1 would be located after searching one tenth of the total search area.

Overall model performance – across all sources – can be compared by calculating the Gini coefficient or Gini index. The Gini coefficient is essentially a measure of inequality (it is often used to look at wealth distribution) (Gini 1921). Here, I compare the proportion of illegal hunting incidents whose sources have been identified to the proportion of the total area searched. A strategy that finds sources exactly in proportion to the area searched would have a Gini coefficient of 0. In contrast, a perfect search strategy would have a Gini coefficient of 1. The higher the Gini coefficient, the more effective the search.

Simulations

To further test the accuracy of the model with and without the incorporated spatial data, I compared 1000 simulated data sets, each dealing with a simplified case with a study area spanning -1° to 1° longitude and -1° to 1° latitude, with a central ‘conservancy’ from -0.5° to 0.5° longitude and -0.5° to 0.5° latitude. I randomly generated 36 sources from a uniform distribution within the study area but outside the simulated conservancy, and 11 sources within the ‘conservancy’, again from a uniform distribution. The ratio of 36:11 was chosen because it reflected the spatial distribution of crimes in the SVC dataset. For each of these 47 sources, I generated a large number of crimes from a bivariate normal distribution with a standard deviation of 20km around the source, and sub-sampled from this distribution to select a maximum of 12 crimes per source such that all of the crimes occurred within the simulated conservancy (note that this constraint meant that for sources further from the conservancy, the realised number of crimes was in some cases less than 12; sources for which no crimes fell within the conservancy were excluded from the analysis). For each data set, eight analyses were carried out: the unmodified DPM model, and then using the same modifications used on the real data set (that is, multiplying by factors from 0.1 to 1×10^{-6} , and also by zero). To account for the paired nature of the design (each analysis was run on the same data set), the data were analysed using an analysis of variance on the differences obtained by subtracting the unmodified DPM hitscores from the hitscores for each of the other analyses; thus, negative values indicate cases in which the modified version of the model outperforms the unmodified DPM.

Spatial data

To account for the issue of commuter crime as mentioned previously, I incorporated spatial information into the model post-hoc. Shapefiles for SVC were superimposed on the geoprofile, and the probability of offender residence within SVC reduced by multiplying points within the Conservancy by 1×10^n , where n ranged from -1 to -6; in addition, I considered the case where the Conservancy was excluded entirely by multiplying by zero within SVC. Effectively, this forces the model to give greater weight to potential locations outside SVC to varying extent. This approach was compared to a simple ‘ring search’ strategy based on searching outwards from illegal hunting incidents in circles of increasing radii (see for example Smith et al. (2015)).

4.4 Results

Simulations

Across the 1000 replicates, the model identified the sources located outside the specified area (here, the area comprising the simulated ‘conservancy’) better when the model was adjusted (Fig. 4.2a). The hitscores improved as the adjustment on the surface increased, until it stopped having an effect at an adjustment of 0.001. (ANOVA: Adjusted surface $F_{7,226504} = 21953$, $p < 0.0001$; location (inside/outside) $F_{1,226504} = 3181562$, $p < 0.0001$, interaction $F_{7,226504} = 201110$, $p < 0.0001$).

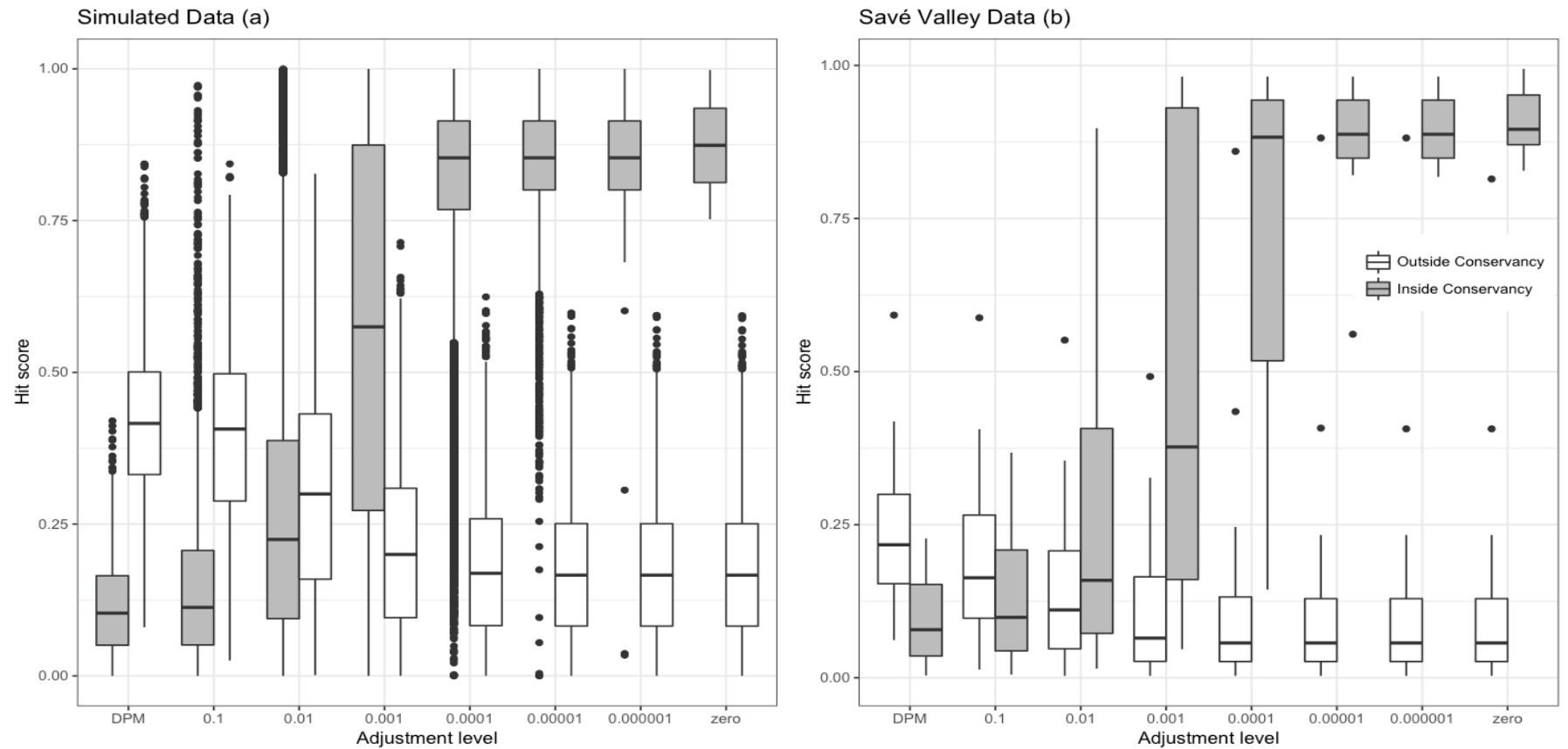


Figure 4.2. Boxplot of (a) simulated and (b) Savé Valley data. The plot shows the difference in hitscore for sources located inside and outside the conservancy (or simulated area) (grey and white boxes respectively).

Spatial data

The geoprofiles produced by the standard DPM model and the subsequent adjusted surfaces are shown in Figure 4.3. Figure 4.3a shows the basic DPM model results before I corrected for the commuter crime issue. Figures 4.3b and 4.3c show the geoprofiles when the probability values inside SVC were multiplied by 0.001 and 0. Hitscores improved as the adjustment on the surface increased and again the model identified the sources located outside the specified area better when the model was adjusted (ANOVA: adjusted surface $F_{7,360} = 7.993$, $p < 0.0001$; location (inside/outside) $F_{1,360} = 1241.61$, $p < 0.0001$, interaction $F_{7,360} = 77.328$, $p < 0.0001$) (Fig 4.2b). Proportions of illegal hunters located using the different methods of spatial targeting were also compared. All of the analyses using the adjusted geoprofiles located 50% of the illegal hunters by searching less than 20% of the area, with hitscores for sources outside SVC improving and hitscores for those inside SVC becoming worse.

The adjusted geoprofile (using a multiplication of 0.001 inside SVC) (Fig. 3b) also outperformed a simple 'ring search' strategy based on searching outwards from illegal hunting incidents in circles of increasing radii (Fig. 4.4). Although the GP hitscores were higher for the small number of sources inside the conservancy ($t = 6.00$, $df = 10$, $p = 0.0001$), they were lower for the larger number of sources outside the conservancy ($t = 18.5$, $df = 35$, $p < 0.0001$), searching on average 13% less of the total area than the ring search. Overall, the adjusted geoprofile identified the sources of more incidents of illegal hunting while searching a smaller area, with a Gini coefficient of 0.879 compared to 0.825 for the ring search, finding the sources for 50% of the incidents while searching 11% of the search area, as opposed to 18

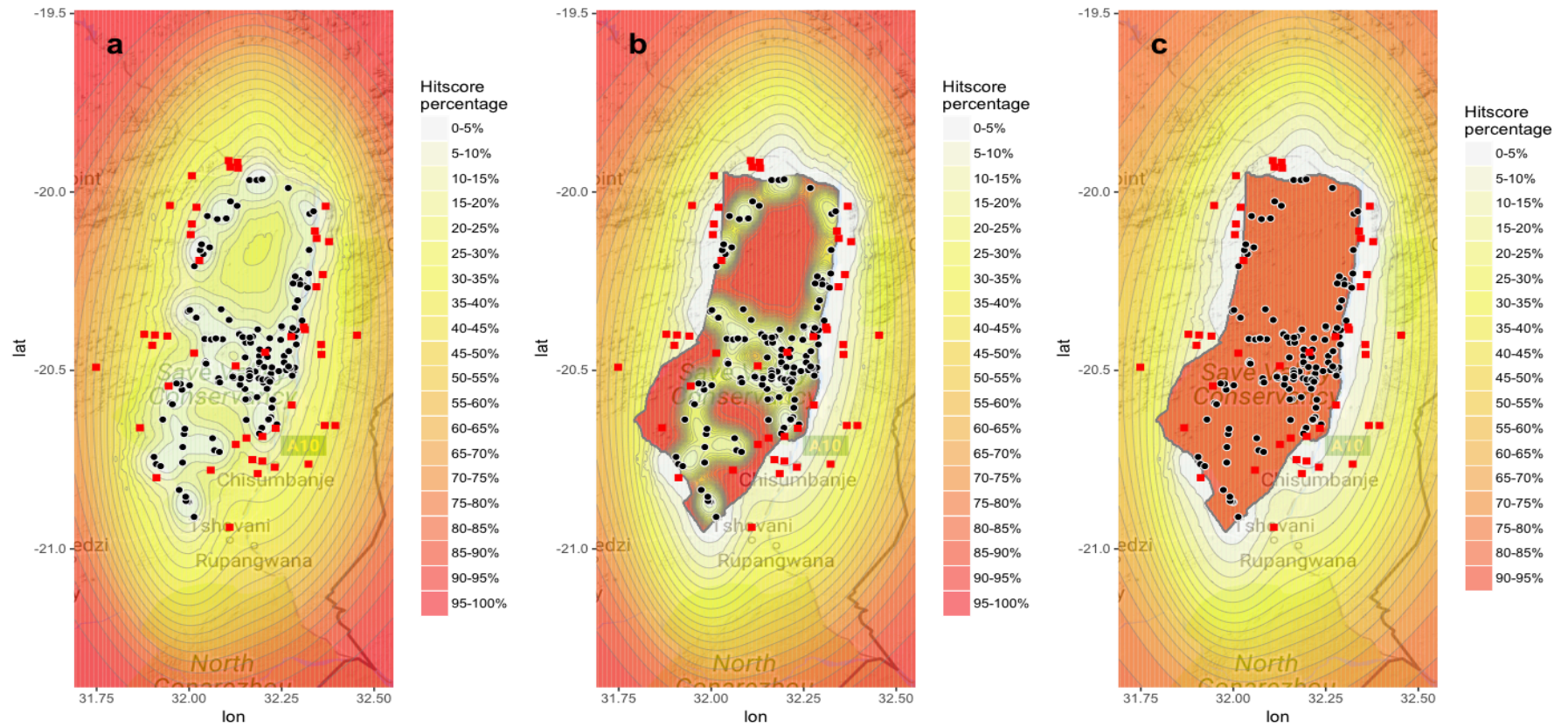


Figure 4.3. Geoprofiles showing the results of the geospatial analyses (a) standard – no adjustment, (b) 0.001 probability and (c) 0 probability. Locations of hunting incidents are shown as black circles and locations of illegal hunters by red squares. Contours show bands of 5%, with lighter colours corresponding to higher parts of the geoprofile.

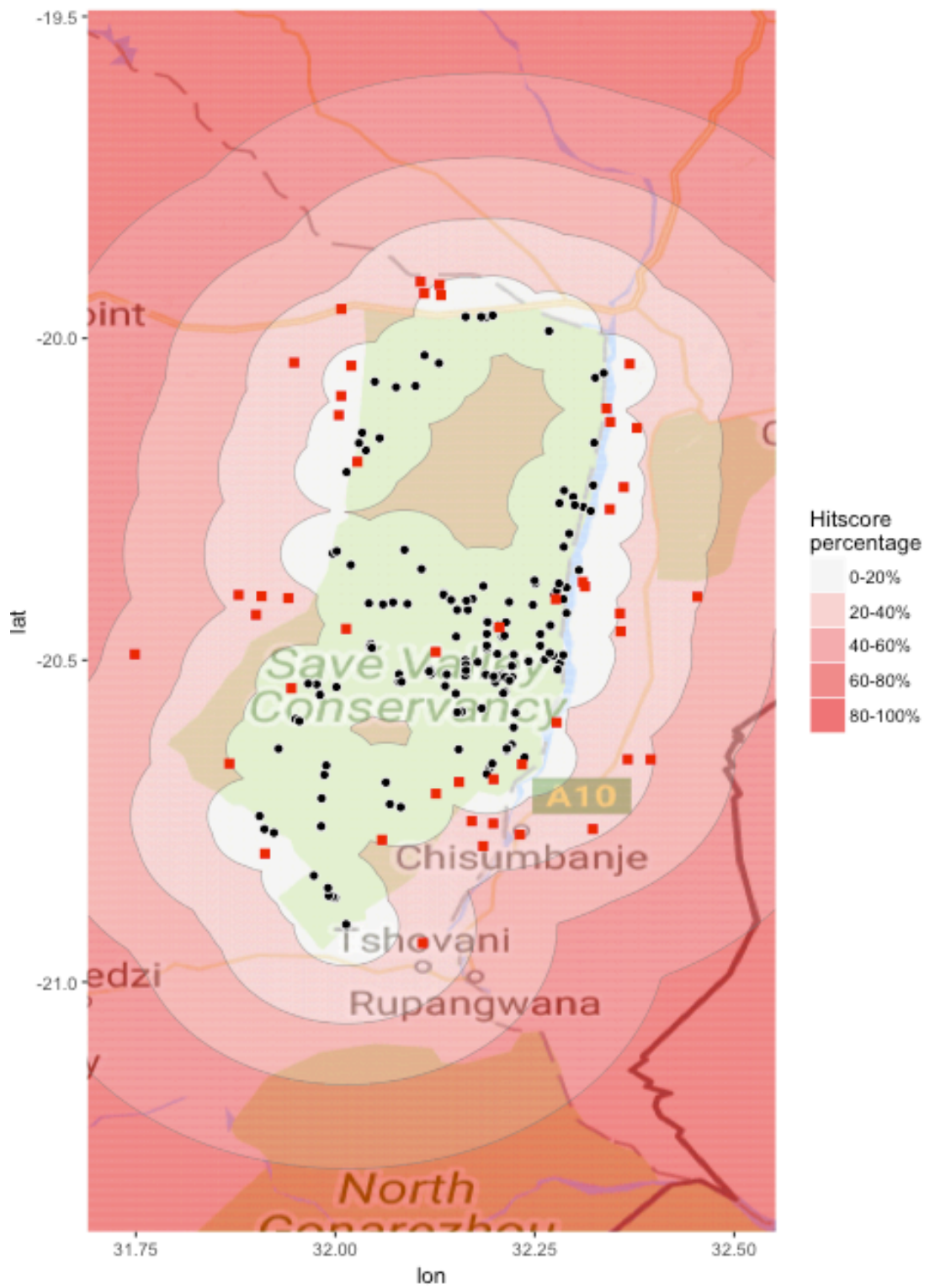


Figure 4.4. An alternative search strategy, based on searching outwards from incidents of illegal hunting in circles of expanding radii.

4.5 Discussion

Crimes that have been committed against the environment and animals – variously termed ‘green criminology’ (Lynch & Stretsky 2003), ‘conservation criminology’, and ‘environmental criminology’ (Gibbs et al. 2010) have had an increasing profile in recent years (Wellsmith 2011). The field of criminology has historically shown little interest in these issues, largely leaving environmental issues to other disciplines (Lynch & Stretsky 2003). This chapter shows that GP can be successfully used to identify areas where illegal hunters may live and could be used to target law enforcement interventions and community engagement efforts in these areas to prevent reoffending. In addition, I demonstrate for the first time how incorporating spatial information can improve the efficiency of the model, with the model outperforming an alternative ‘ring search’ strategy. Crucially, the DPM model identified the sources of 50% of illegal hunting incidents after searching just 11% of the study area, as opposed to 18%. Clearly, across the spatial scales that often characterise reserves and conservancies, such an improvement in efficiency may be of considerable benefit.

The origins of geographic profiling lie in criminology, and this study takes the modifications to the model that have been developed in biology back to this source. In criminal investigations, limitations of resources and time mean that a search prioritisation tool such as GP can be of great practical utility. The same can be said for conservation where resources and time are likely to be heavily limited (Stevenson et al. 2012; Faulkner et al. 2016).

There has been an increase in the scale of commercial hunting and the wildlife trade as the population expands and as techniques used by hunters improve (Fa & Brown

2009; Peres 2009; Di Minin et al. 2015; Naidoo et al. 2016). Traditionally conservation actions have been dependent on the hypothesis that different illegal wildlife actions occur in different places; commercial trade will occur closer to cities and coastal areas (Di Minin et al. 2015) and illegal hunting incidents will cluster in rural areas where the primary motivation for hunting is subsistence (Sanchez-Mercado et al. 2016). However, it has recently been shown that subsistence hunting and wildlife trade maybe spatially correlated (Sanchez-Mercado et al. 2016). In fact, spatial patterns of hunting will differ from case to case, just as the techniques used by the illegal hunters and the pressures driving hunting will vary between countries, time of year species and protected areas as illegal hunters adapt to – for example – difference in terrain and accessibility to protected areas and to the population changes that will occur amongst the animals (Risdiyanto et al. 2016). Geographic profiling provides one way of identifying locations that are the source of hunting – in most cases, areas where illegal hunters live – on a case by case basis. This could have important implications for the design and implementation of effective and efficient conservation actions since it could allow limited law enforcement resources to be focused on communities where it is needed most and help focus conservation efforts and generate economic benefits from wildlife to these local communities (Knapp 2012; Cooney et al. 2016). Such focusing of efforts is key. Law enforcement and protected area management is expensive and enormous budget deficits exist in African countries (Lindsey et al. 2016, 2017). Traditional anti-poaching patrols are reactive and attempt to find evidence of hunting after it has already happened, or after illegal hunters have already entered the area (Lotter & Clark 2014). Due to the large areas that are often involved and the difficulty associated with finding snares and traps, or of catching illegal hunters on the move, such interventions often fail to prevent hunting incidents and are of limited

efficacy. This method, especially if combined with information from intelligence operations has potential to allow for both preventative outreach efforts with the communities and households most involved in illegal hunting, and also much more targeted law enforcement efforts (Lotter & Clark 2014).

Beyond the interest of the particular case I describe here, this chapter illustrates how more complex spatial information can be incorporated within the DPM model framework. In many instances – notably in biology but also in criminology – treating the study area – the target backcloth in criminology – as homogenous will fail to take into account important information. For example, if we are searching for plants that only occur above 400m, or mosquitoes that only breed in water, it may well be the case that large parts of our study area can be excluded from the search, creating a more efficient search strategy. More complex manipulations of the model output – using continuous variables, rather than the categorical inside/outside here – are also possible – for example, if the probability of finding an anchor point is proportional to altitude, soil pH, distance from water, etc.

In some cases, of course, it will not be obvious precisely what manipulation of the final model output will be most appropriate and selecting a particular manipulation will require expert input. In this study, for example, it is clear that entirely excluding areas inside SVC from the search misses a number of sources (Fig 4.3c); multiplying by 0.001, on the other hand, effectively excluded large areas within SVC which are unlikely to be of interest, while still prioritising the areas of highest probability within the Conservancy (Fig 4.3b).

4.6 Conclusion

This chapter shows that geographic profiling can successfully identify areas where illegal hunters may live, using only the spatial locations of hunting incidents such as traps and snares. This has important implications for management strategies and conservation plans in terms of targeting particular areas with community based initiatives. I suggest that by being able to target control efforts in this way, will make hunting interventions more efficient and cost effective. More broadly, I demonstrate for the first time how incorporating additional spatial information can improve the overall efficiency of the DPM model.

Chapter 5: Human-wildlife conflict in Sumatra: using geographic profiling to identify areas of human-tiger conflict

Faulkner et al (in prep)

5.1 Abstract

The number of wild Sumatran tigers has been decreasing dramatically since the 20th century. There are a number of reasons for this decline, including habitat loss, habitat fragmentation and trade of illegal animal parts. Additional to these are human-wildlife conflict and persecution by humans in response to real or perceived livestock predation and/or attacks on people. As a result, there is an urgent need to be able to monitor and reduce the human-tiger conflict for the c.500 remaining wild Sumatran tigers (*Panthera tigris sondaica*) on the island of Sumatra. Here, I show how geographic profiling – a statistical technique originally developed in criminology and more recently applied to animal foraging and epidemiology – can be used to identify areas where the risk of human tiger conflict is high and indicate where retaliatory tiger killings are likely to occur. I use data collected from within 30km from the Kerinci Seblat National park over 13 years, including tiger sightings (106), attacks on livestock (83), attacks on people (12) and killing of tigers (27). First, I show that locations where tigers have attacked livestock or humans represent a good predictor for where tigers themselves will be hunted, finding 80% of these sites searching just 14% of the study area. I then show that GP can correctly predict where the livestock (or human) attacks may occur using as input areas where tigers have been sighted. I also show how incorporating a measure of farming intensity improves the model's search strategy and efficiency, with more human-tiger conflict events where farm density is higher. Being

able to predict areas of future human-tiger conflict will have important implications for the development of management strategies and conservation plans.

5.2 Introduction

Human-wildlife conflict has been a problem globally for many years now in both developing and developed countries and is a major conservation issue (Manfredo & Dayer 2004). Community-based natural resource management systems are expected to hold more promise for conservation of wildlife, with local people recognising wildlife as an asset (du Toit 2002). However, when wildlife presents a risk to human life or property these conservation efforts can be undermined (Gusset et al. 2009). Outside the boundaries of the protected areas the risk of livestock predation is a particular challenge. Community and individual support for predator conservation will depend largely on the value that local people place on the animal (Sillero-Zubiri & Laurenson 2001; Gusset et al. 2008). If we are unable to address the human-wildlife conflict at a community level we will ultimately be threatening this area of economic growth (Madden 2004; Gusset et al. 2008).

Throughout Asia there has been a dramatic decline in the numbers of tigers; recent population estimates put this figure at less than 4000 (Wilting et al. 2015). Tigers occupy only 7% of their original distribution range, with only a few forest patches that can maintain viable tiger populations (Nyhus & Tilson 2004; Seidensticker 2010; Sanderson et al. 2010; Wilting et al. 2015). Reasons for this decline include habitat loss, forest degradation and an increase in the illegal wildlife trade (O'Brien et al. 2003). In addition to this there is an increase in persecution by humans in the form of retaliation hunting as a response to livestock loss. When humans and wildlife occupy

the same areas and use the same resources tensions can be high and tolerances will decline as livestock, crops and personal security are threatened (Campbell-Smith et al. 2010; Hardwick et al. 2017). Human-wildlife conflict has contributed to the extinction of two tiger sub-species – the Bali tiger (*Panthera tigris balica*) and the Javan tiger (*Panthera tigris sondaica*) (Hoogerwerf 1970; Seidensticker 1987; O’Brien et al. 2003). In Sumatra there is an urgent responsibility to understand, monitor and reduce the human-tiger conflict, as there are only c.500 Sumatran tigers remaining in the wild (Nyhus & Tilson 2004).

Geographic profiling

Geographic profiling (GP) is a mathematical technique that was pioneered as a criminal investigative tool to deal with cases of serial crime such as arson, rape and murder. It was designed to solve the problem of information overload, and uses the spatial locations of serial crime to calculate the probability of offender residence at each point within the study area (Rossmo 2000). The probabilities are then ranked to produce a geoprofile – a probability surface in which the suspects are investigated in order of their height on the profile. More recently this method has been successfully applied to analogous problems in biology such as locating animal foraging sites and sleeping sites (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009; Le Comber & Stevenson 2012; Faulkner et al. 2015), identifying sources of disease outbreaks from addresses of disease carrying individuals (Buscema et al. 2009; Le Comber et al. 2011), identifying locations of poachers from the poaching locations (Faulkner et al. 2018) and invasive species biology, where it identifies the source populations of invasive species from the observed locations of individuals (Stevenson et al. 2012; Faulkner et al. 2017).

GP has been highly successful within criminology and in a range of biological fields, but until very recently the model has largely ignored complex spatial information, instead assuming spatial heterogeneity across the study area (Faulkner et al. 2018). Here, I address the issue of differences in habitat use and how this will be more important and relevant to tiger activity and subsequently to human-wildlife conflict. I believe human wildlife conflict will be greatest where tigers are leaving their territory within the forest and becoming more active in the human-dominated landscape. As a result I incorporated farm density into the model and ask if this improves the search efficiency in identifying known locations of human-tiger conflict.

Aims

In this study I ask whether GP can be used to predict areas human-tiger conflict. I ask (1) if the locations where tigers have attacked humans or livestock can predict where tigers are themselves hunted, and (2) if can I use GP to correctly predict where the livestock attacks from the locations where tigers have been seen. I revisit both of these questions after incorporating farm density as an additional variable. Overall, I aim to expand GP beyond the usual spatial analyses of this method and help develop management strategies in Sumatra. By highlighting areas of human-tiger conflict I may be able to suggest ways to better manage the Sumatran tiger's future conservation needs.

5.3 Methods

General approach

This study examines how a method used in criminology can be adapted and used to identify locations of human-tiger interactions. I extend the model to show how post-hoc adjustment of the geoprofiles to include other geospatial information can improve model performance. In this case, I address the complexity of the surface with respect to where human-tiger incidents may occur; I would expect more conflict to occur where there is a higher intensity of farming. I first divide the geoprofile – a matrix describing, for each point in the study area, the probability that there is a source at that point – into areas of different spatial complexities using spatial rasters, in this case a raster representing farming intensity. I then adjust our estimate of the probability of source location with respect to each individual raster grid cell. See Faulkner et al. (2018) for more details.

Study Area/data collection

Tiger encounter data were collected by the Kerinci-Seblat National Park (KSNP) authority and Fauna & Fauna international (Fig 5.1). KSNP comprises of 13,000km² of continuous mountainous forest, which is capable of supporting a healthy tiger population (Wikramanayake et al. 1998; Linkie et al. 2003). The National Park is recognized as a UNESCO World Heritage Site for its rich biodiversity and a global priority for the long-term survival of wild tigers (Dinerstein et al. 2007). KSNP is surrounded by lower elevation smallholder farms and spans the provinces of Jambi, West Sumatra, Bengkulu and South Sumatra. The altitude ranges from the lowland hill forests at 200m up to the montane forests with a highest point of 3,805m (Linkie et al. 2003). To protect the tigers, six Tiger Protection and Conservation units have been in

place around the park since 2000. Data from ranger patrol logbooks and datasheets were compiled and georeferenced (Linkie et al. 2015), these data included sightings (106), attacks on livestock (83), attacks on people (12) and poaching of tigers (27), each incident was recorded as an independent incident if it occurred at least one month apart.

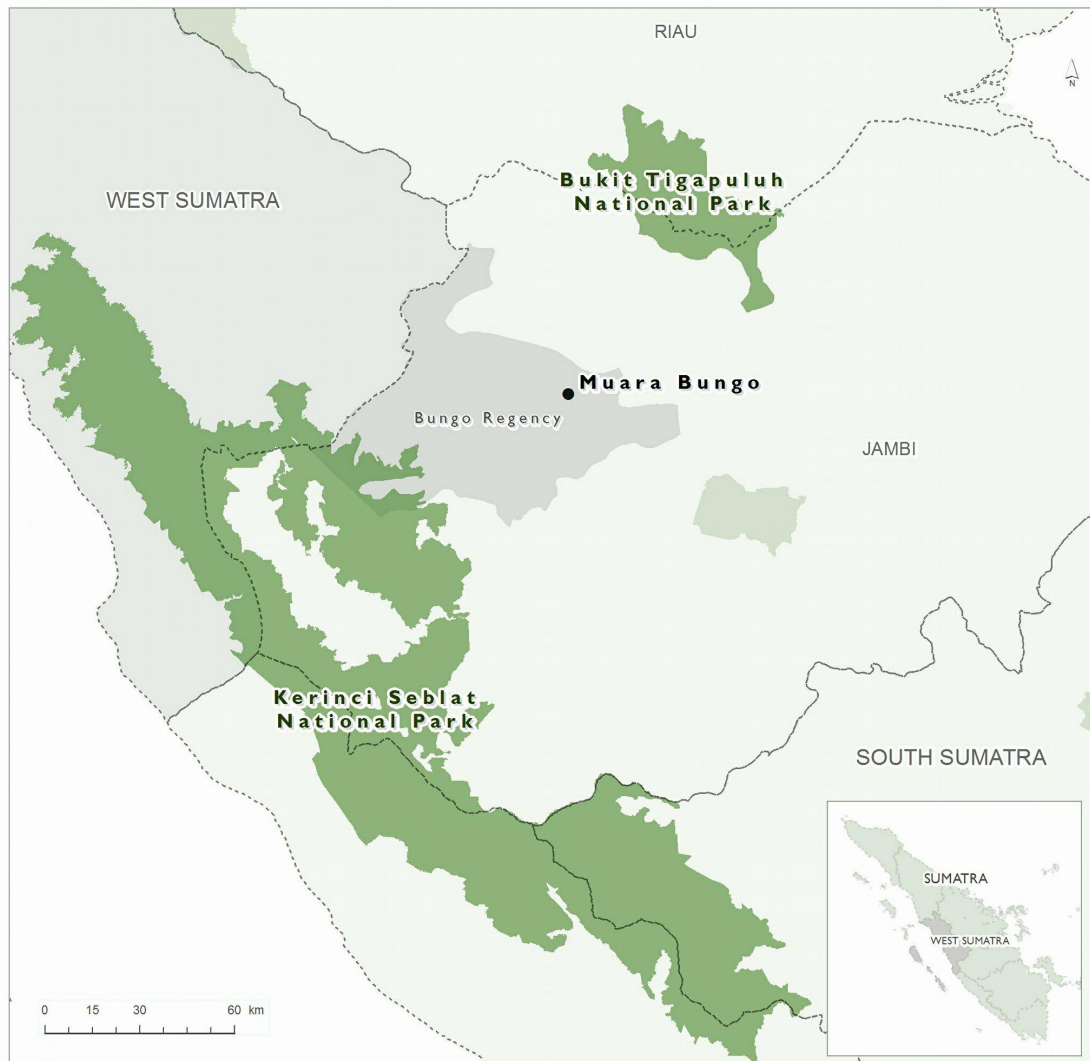


Figure 5.1. Kerinci-Seblat National Park, Sumatra.

Geographic profiling: the DPM model

I used the Dirichlet Process Mixture (DPM) model of geographic profiling, described fully in Verity et al. (2014) and extended in Faulkner et al. (2017, 2018), to analyse

the spatial locations of the tiger encounter data. Specifically, I used livestock/human attack and tiger sightings data as the input into the model and used the resulting probability surface to predict where future tiger encounters – such as tiger attacks on humans and livestock or tiger poaching events – were likely to occur. Constructing a geoprofile using the DPM model can conceptually be broken down into two parts; first, the tiger encounters are separated into clusters, with encounters that are close together more likely to be assigned to the same cluster. Second, source locations are estimated based on the initial clustering. The DPM model alternates between these two parts many thousands of times, using standard Bayesian Markov chain Monte Carlo (MCMC) based on a Gibbs sampler (Geman & Geman 1984), until the algorithm converges on a posterior distribution of interest.

Model implementation

The DPM model described here was implemented in R (R Core Team 2014) using version 2.1.0 of the package Rgeoprofile introduced by Verity et al. (2014) and extended in Faulkner et al. (2017); this package is available at <https://github.com/bobverity/Rgeoprofile>. A key parameter of the DPM model is sigma, the standard deviation of the bivariate distribution around the sources; a sigma value of 1 km implies that 39% of points (eg crimes, sightings etc) occur within 1 km of the ‘source’, 87% within 2 km and 99% within 3 km (see Faulkner et al (2017) for details of the underlying mathematics). Since the model fits sigma based on within- and between-cluster distances, the inclusion of multiple records at exactly the same location (as in the dataset here) will cause the model to fit artificially small values of sigma. To avoid this, I used the approach in which I first of all ran the model on the 105 unique locations only and fitted sigma from these data only. This produced a sigma

value of 27 km, which I then used to analyse the full dataset, using 50000 samples and 10 chains with a burnin of 10000. The parameters *samples*, *chains* and *burnin* are all standard parameters relating to the MCMC and are fully explained in Verity et al. (2014) and in the help documentation for Rgeoprofile 2.1.0

Model performance

The geographic profiling model output is assessed in two ways. For individual points, the model's performance can be assessed using a hitscore. This is the proportion of the total area covering the 'crimes' (for example sightings) that the model had to search before a source is located. The overall model performance – using all sources – is assessed using the Gini coefficient, which is obtained by plotting the proportion of sources found as a function of the proportion of the area searched; see Faulkner et al. (2018) for details. Gini coefficients range from 0 to 1, with an optimal search strategy producing a value of 1.

Spatial information

To address the issue of landscape heterogeneity I incorporated farming intensity into the model. I would expect human-tiger conflict to be greater where farming intensity is higher. I matched the extent of the geoprofile (excluding areas in the sea, and those outside of the area for which data were available) with the farming intensity study area and multiplied the two surfaces together. Thus, areas with high farming intensity will be prioritised, while areas of low farming intensity will be less heavily weighted. Effectively, these manipulations force the model to give greater weights to different parts of the geoprofile based on farming intensity.

5.4 Results

Can the locations where tigers have attacked humans or livestock predict where tigers are themselves hunted?

Figure 2 shows the geoprofiles produced using the locations where tigers have attacked either humans or livestock as ‘crimes’ to predict where tigers have been hunted, first using the basic DPM model without the inclusion of any additional spatial information (Fig 5.2*ab*) and after including farm density (Fig 5.2*b*). The hitscores improved after I added the spatial surface (paired t-test; $t = 3.3526$, $df=25$, $p = 0.002551$). The Gini coefficient associated with the unadjusted surface was 0.738, with the model identifying 70% of the attacks locations by searching just 17% of the total area. This was improved by incorporating the farm density surface, increasing the Gini coefficient to 0.839, now the model identified 80% of the locations where tigers had been hunted by searching less than 14% of the search area (Fig 5.3).

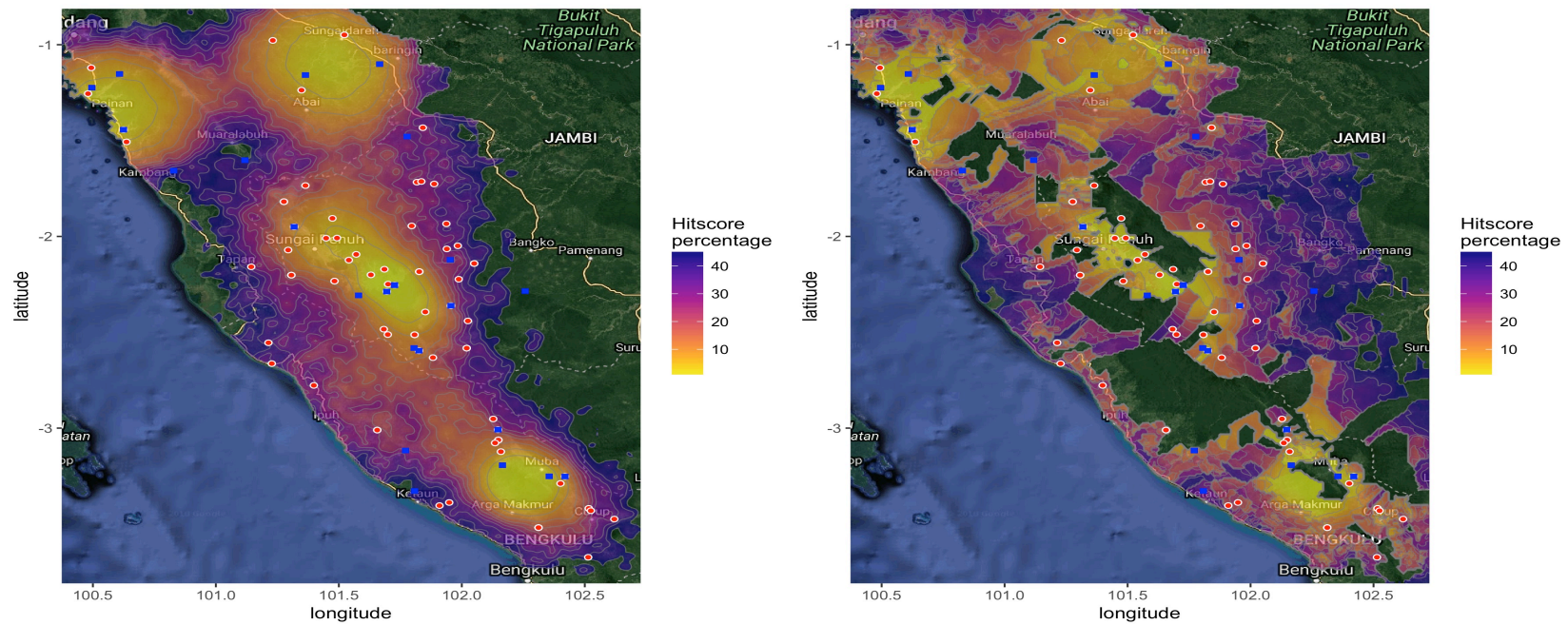


Figure 5.2. Geoprofiles showing the results of the geospatial analyses using the attack data to predict areas where tigers may get hunted (a) unadjusted surface and (b) surface adjusted with farm density. Locations of human-wildlife conflicts events are shown as red circles and locations where tigers have been hunted represented by blue squares. Contours show the top 40% of the geoprofile, with the highest priority areas shown in yellow.

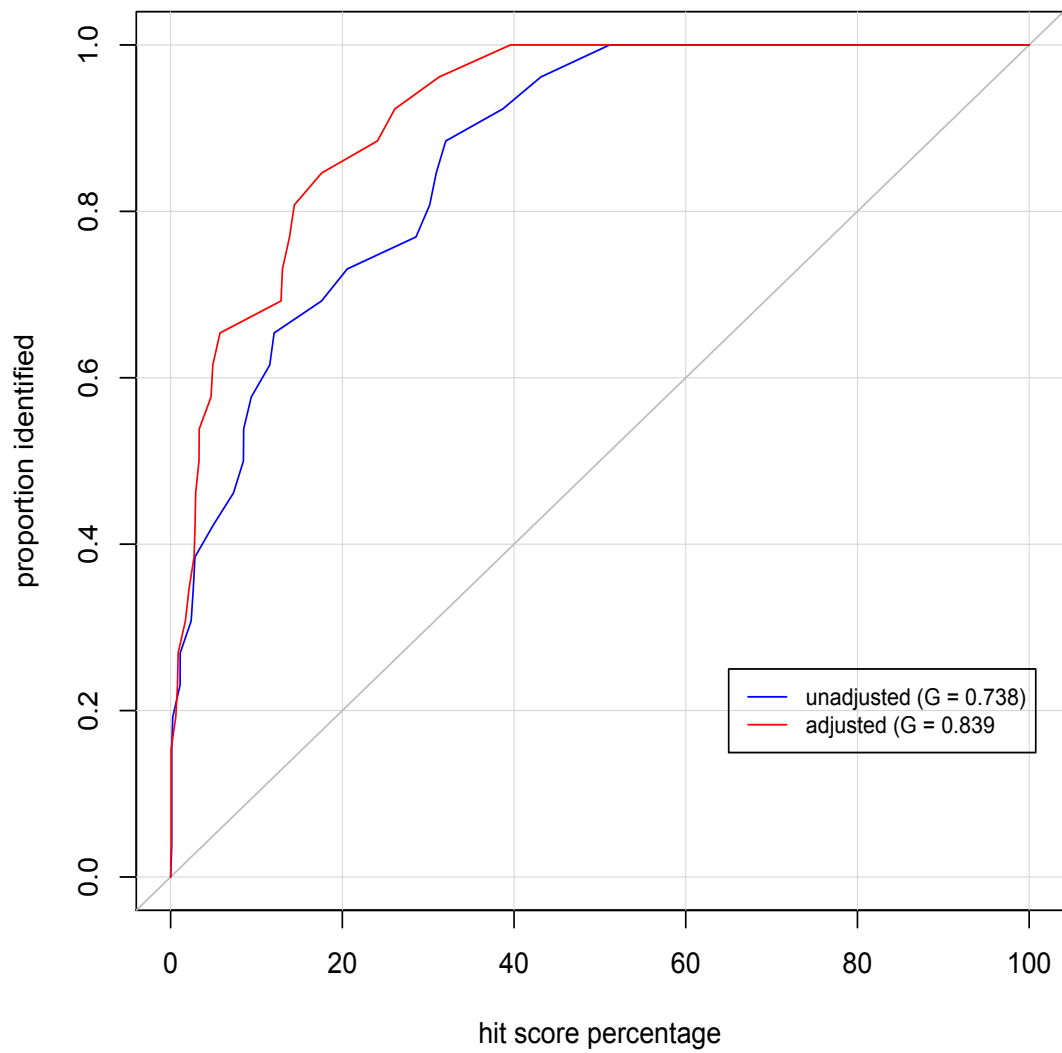


Figure 5.3. Lorenz plot for the unadjusted profile in blue (Gini coefficient 0.738), and the final adjusted geoprofile including a farming intensity surface in red (Gini coefficient 0.839).

Can we use GP to correctly predict the locations the livestock attacks from the locations where tigers have been seen?

Second, I ask if can use the DPM model to correctly predict where livestock attacks may occur from the locations where they have been sighted. The geoprofiles produced using the standard DPM model, and the subsequent adjusted surfaces are shown in Figure 5.4. Figure 5.4a shows the basic DPM model before incorporating the farming intensity data. Figure 5.4b shows the profile after I multiplied across the surface with the farming intensity spatial information. The hitscores improved once I added the geospatial information with the model identifying more areas of human-wildlife conflict events using a smaller search area (paired t-test; $t = 3.0683$, $df=92$, $p = 0.002828$). Overall, the adjusted surface identified more human-wildlife conflict events while searching a smaller area than the unadjusted surface with a Gini coefficient of 0.565, as compared to 0.458 for the original, unadjusted model, identifying 60% of the sources by searching just 23% of the total area, compared to searching 31% of the total area using the unadjusted surface (Fig 5.5).

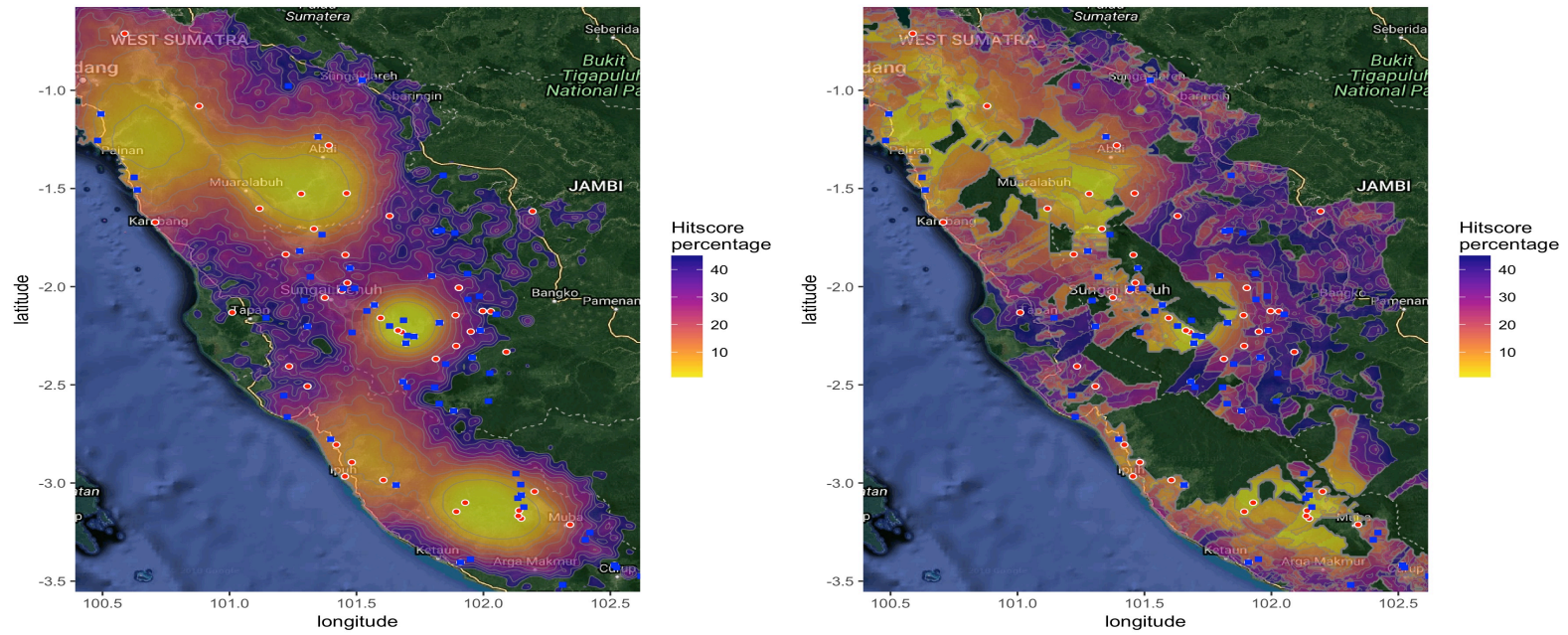


Figure 5.4. Geoprofiles showing the results of the geospatial analyses using the sightings data to predict areas where tigers may attack humans or livestock (a) unadjusted surface and (b) surface adjusted with farming intensity. Locations of sightings are shown as red circles and locations where tigers have attacked humans or livestock as blue squares. Contours show the top 40% of the geoprofile, with the highest priority areas shown in yellow.

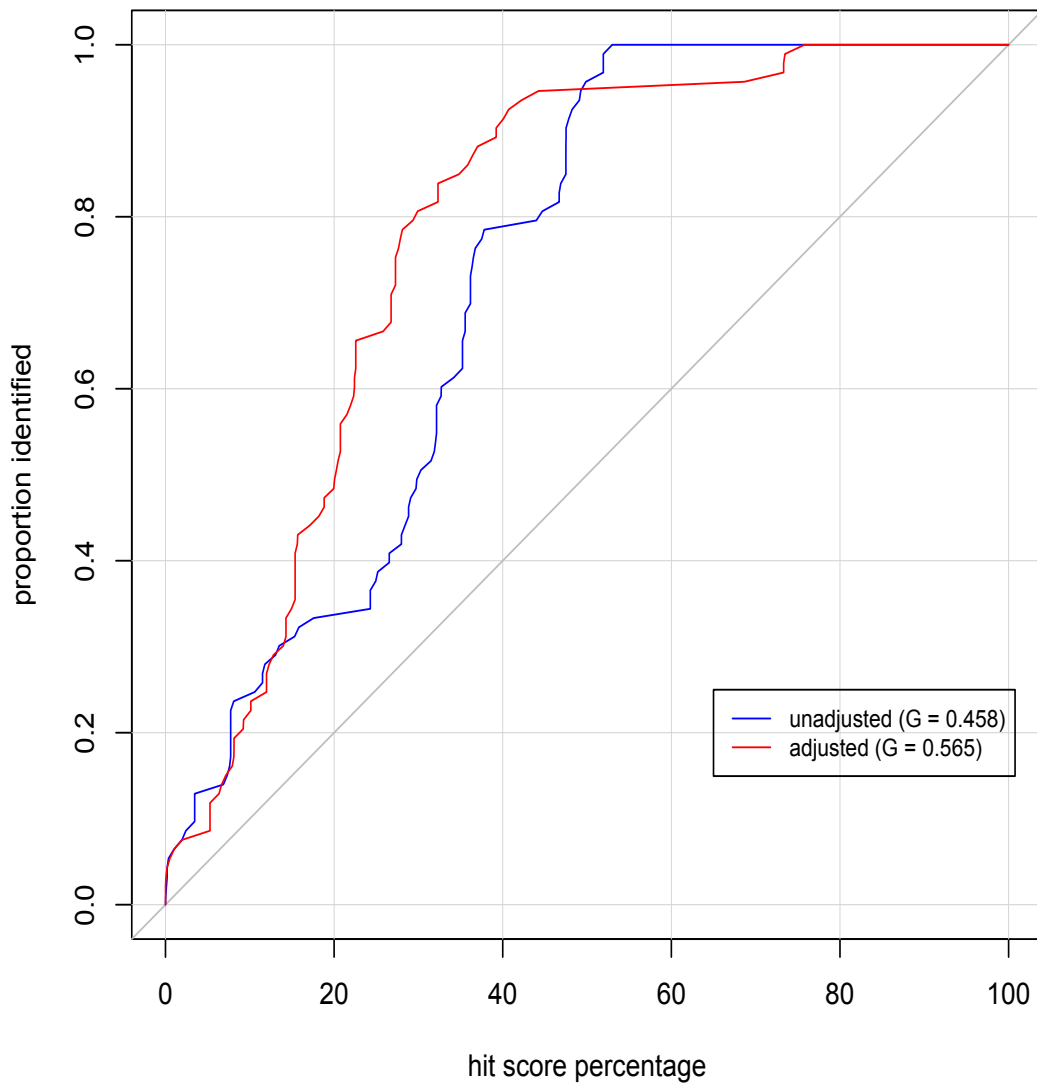


Figure 5.5. Lorenz plot for the unadjusted profile in blue (Gini coefficient 0.458), and the final adjusted geoprofile including a farming intensity surface in red (Gini coefficient 0.565).

5.5 Discussion

In criminal investigation, geographic profiling is an important practical tool for use in situations when resources and time are limited and where improved targeting of interventions is critical. It is these same characteristics, which make it a potentially useful tool in conservation where time and resources are often similarly limited (Stevenson et al. 2012; Faulkner et al. 2017; Faulkner et al. 2018).

Given the limited resources that are available to conservation managers and agencies, there is an urgent need to work with governments and landowners to develop targeted strategies that are efficient and reliable. Here, I show that locations where human-wildlife conflict (livestock and human attacks by tigers) have occurred are the best predictor for where these endangered animals may themselves be killed. I also show that geographic profiling is successful at identifying areas where these incidents of human-wildlife are likely to occur. In addition, I demonstrate how incorporating additional spatial information can improve the performance of the model. Clearly, identifying areas where human-wildlife conflict is more likely to occur means that we can target particular areas where non-lethal and affordable mitigations could be implemented. This could allow intervention in particular areas and potentially prevent further retaliation killings. Bespoke land management plans could be put in place with focus groups, discussions, education and training.

The probability of human-tiger conflict, both human and livestock attacks and tiger hunting incidences, appears to be highest in diffuse edge areas around the Kernici National Park where there is greater overlap between tigers and humans, and this conflict is lowest

when there is little overlap. Considering both land use and tiger density together – as I do here – means that I was better able to identify areas of future conflict as a prelude to interventions aimed at reducing this conflict.

This study shows that attacks on livestock and humans by tigers are generally a good indicator for where tigers may get hunted. The unadjusted DPM model identified these areas efficiently, but model performance was improved by incorporating farm density. I showed that tiger attacks on livestock occur more frequently where there is a higher density of farmland – and hence people or livestock (Fig 5.2). Crucially, the DPM model identified 80% of the tiger poaching incidents by searching just 14% of the total area when I incorporated the farm density surface as opposed to 70% of poaching incidents searching 17% of the area, without the this additional spatial information (Fig 5.3).

Importantly, I also show that GP can successfully identify areas of human-tiger conflict such as attacks on livestock using just the locations of tiger sightings. Again, I show that I can improve the efficiency of the model by incorporating information on farm density, finding 60% of conflict sites after searching just 25% of the study area. Crucially, if I can identify where tigers may attack livestock I could potentially prevent these attacks from occurring.

With the global problem of human wildlife conflict increasing especially around protected areas, novel methods to diffuse these conflicts need to be developed (Manfredo & Dayer 2004; Wang et al. 2006). These problems will escalate when local people feel that the values and needs of the wildlife becomes a priority over their own requirements (Madden

2004). In order for tigers and humans to co-exist integrated management strategies must be in place, which take multi-land uses into account and can control the hunting and poisoning of tigers as a response to human-wildlife conflict. The attitudes of local communities concerning conservation policies are important and this must be taken into account when during the decision-making process (Wang et al. 2006; Hardwick et al. 2017). Figures 5.2 and 5.4 both show how incorporating complex spatial information into the model improves its performance. Although the model performs well prior to any manipulation, in many cases assuming a homogenous landscape will lose critical information. Faulkner et al. (2018) showed how a simple binary post-hoc manipulation of the surface (inside or outside a particular area) could be used to improve the search strategy when dealing with so-called ‘commuter crime’. Here, I extend this by incorporating a continuous surface representing farm density and use it to multiply across the whole geoprofile. More broadly within the field of GP, this approach could be extended to use other continuous surfaces such as altitude, rainfall, water depth and soil pH.

5.6 Conclusion

This study shows that whilst tiger hunting in response to a livestock attack is prevalent it is also predictable. I show that it is possible to use human-wildlife conflict data/ mapping and the optimised geographic profiling technique to predict areas where conflict such as poaching in response to livestock hunting by tigers is likely to occur. These predictions could enable bespoke spatial targeting of limited conservation resources. Solutions to combat these complex issues are not straightforward and increased persecuting of the offenders should not be considered. By predicting areas where these attacks are most

likely to occur we can focus efforts to increase community involvement in eco-tourism and engage in efforts to help increase anti-predation strategies. There has been a drive to develop initiatives that improve the landscape connectivity for tigers, by linking protected areas through corridors (Brodie et al. 2015), however these increased boundary areas will increase the risk of conflict with local people and this conflict risk is likely to lead to more tigers being killed in retaliation. Human-wildlife conflict will continue to increase when habitats are restricted and animals are forced into areas beyond their own decreasing territories (Gaveau et al. 2006; Jędrzejewski et al. 2017). This deforestation and habitat fragmentation is the overriding issue that needs to be addressed.

Chapter 6: Summary of research and main findings

6.1 Abstract

Geographic profiling is continually adapting and evolving as new challenges arise. This thesis has for the first time applied the method to conservation, addressing new challenges, such as spatial heterogeneity, and further developing the model. In this final chapter I summarise the research undertaken and explore the rationale for the thesis. I outline the key findings from each study within the thesis and discuss the strengths and limitations of this work. I also make recommendations for future applications – both in biology and criminology – and developments of this model for further use in conservation.

6.2 Main findings

GP was originally developed in criminology and is still commonly used in its original CGT form – which forms the basis of Rigel (ECRI, Canada) – by the Metropolitan Police, the FBI and others. The ability of the CGT model to reliably detect an anchor point associated with a criminal has been invaluable in many criminological cases, such as the Yorkshire Ripper and Washington sniper (Thurman & Jamieson 2015). While GP does not necessarily directly solve cases it helps to focus the investigation and helps manage large volumes of information. Just as resources will often limit criminal investigations, a fundamental challenge in conservation biology is deciding what areas are most worthy of conservation action (Ferrier et al. 2002). So, applying the model to conservation datasets was a logical step.

In Chapter 1, I discuss the history of geographic profiling, its more recent biological applications and the importance of considering a spatially heterogeneous environment.

In Chapter 2 I use the GP model to locate sleep trees for tarsiers in dense jungle in Indonesia, using as input the locations at which calls were recorded, demonstrating how the model can be applied to locating the nests, dens or roosts of other elusive animals and potentially improving estimates of population size, with important implications for management of both species and habitats. This is the first time that GP has been used for conservation purposes.

Chapter 3 builds on this work to show how spatial information can be used to improve a study of invasive mink in the Hebrides. I show that GP is a useful tool in situations where data are poor or not as reliable, such as citizen science, and that citizen science data can in fact be used to capture much of the information about sources of invasive species that is obtained from the harder to acquire and more expensive trap data. In this chapter I also introduce an extension to the DPM model (developed by R. Verity), which allows the model to infer source locations without the need to specify dispersal parameters. It is this chapter that highlights the importance of being able to introduce the issue of heterogeneous surfaces to within the DPM framework. Here I needed to compare two geoprofiles encompassing the Isle of Harris and the Isle of Uist in the Inner Hebrides, Scotland, but low scores assigned to the sea artificially inflated a correlation between the two. It became clear that having the ability to remove the sea from the analysis would increase the search efficiency of the model. This led the way to incorporating geospatial information into the model.

In Chapter 4, I turn to the issue of ‘commuter crime’ in a study of poaching in Savé Valley Conservancy (SVC) in Zimbabwe. This is the first time spatial information in the form of GIS data has been incorporated into the model. I address the problem that although poaching occurs inside SVC, the majority of illegal hunters live outside. I show how the model can be adjusted to reflect a simple binary classification of the landscape (inside or outside SVC). Although these data were collected 12 years ago, this study could potentially have an important impact on the way in which NGOs and other conservation organisations deal with issues such as illegal hunting, by making more efficient and cost effective decisions.

Finally, in Chapter 5, I combine more complex land use information (point estimates of farm density) with the GP model to improve predictions of human-wildlife conflict such as tigers attacking livestock or a tiger being hunted as a retaliation killing. I show how using a measure of farmland density improves the model’s search strategy and efficiency, with more human-tiger conflict events where farm density is higher. This study shows how GP, as it stands, can aid in managerial decisions and planning that deal with these conservation issues, but importantly how I can implement complex spatial information on a case by case basis to improve the efficiency of the model. This is the first time I use a continuous spatial surface – farming intensity – rather than a more binary approach.

6.3 Are these geospatial layers important for criminology?

In this thesis I have adapted a technique that originated in criminology, and applied it to conservation data before taking the model back to its beginnings and applying it to studies of wildlife crime. It is interesting now to consider how the research from this thesis could be applied to criminal investigations.

GIS and remote sensing tools are commonplace tools for law enforcement agencies. Criminological studies already use mapping software to visualise crimes and crime mapping is a key component of modern day crime analysis. Advancements in GIS technology have meant that that the need for GIS experts in law enforcement is increasing rapidly (Waters 1998). Using GIS, crime analysts can overlay geospatial layers such as census demographics, street networks and land use onto crime maps and look for any emerging patterns. For example the U.S. Forest service developed a system to locate illegal *Cannabis sativa L.* growth sites within the Chattahoochee National Forest of Georgia. They incorporated variables that were relevant such as soil conditions, sun exposure and privacy. As a result they had a very successful prediction rate (Rossmo 2000). Techniques employed by epidemiologists to predict disease outbreaks have been used to assess the likelihood of a predator being active in a particular area (Rossmo 2000). Although criminologists use GIS software, to date there is no record of incorporating this technology within the GP framework. There are times when it would be useful to be able to prioritise particular areas. I believe incorporating informative geospatial layers such as street networks, land use and demography into the current GP model used in criminology would be a beneficial adaptation. This could simply mean excluding areas of water

surrounding urban areas, or prioritising distances closer to canals if it was known that a criminal was acting just on tow paths.

6.4 Implications and recommendations for policy and practice

Despite the current variety and successes of modelling techniques in the field of biology, there is always room for more pragmatic algorithms. Geographic profiling was the product of an academic doctorate by Kim Rossmo in 1995 and is now commonly used by many law enforcement agencies. It is my hope that GP will now be taken up by ecological organisations and conservation bodies to become a useful tool in these disciplines. It is imperative that conservation management strategies and policies move forward and do not dwell on the past. With the increasing pressure on the natural systems we must critically review our approaches to conservation and look at novel and effective ways to manage our natural resources (Doak et al. 2014).

A novel application such as GP should be attractive to policymakers and conservation managers as it is clear, quantifiable and now has a proven track record in the field. The use of GP is transferable and it is applicable across many diverse jurisdictions and disciplines. This thesis alone has shown that it can be used to locate elusive animals, find sources of invasive species, and help identify areas of human wildlife conflict and wildlife crime. Previous studies have identified its application in epidemiology. All of these disciplines, including criminology, share a similar issue and that is they all cover large areas and large remits with limited resources. The natural world is complex and trying to align conservation practice with this ecological complexity can result in many challenges

such as over-homogenising habitat that will benefit some species whilst compromising others (Hiers et al. 2016).

An important issue I wanted to address in this thesis was the relationship between humanity and nature. Taking this into account when dealing with conservation management decisions could produce better long-term conservation and sustainability outcomes. Retaliation killings of wildlife are an often-understandable result of a livestock or human attack, as seen in Chapter 5. GP can help prevent these retaliation killings from occurring by predicting where livestock (or human) attacks may occur, rather than punishing poachers after the event. In Chapter 4 I address the issue of poaching and suggest that the areas the model predicts as having a large concentration of poachers should be targeted for education programs or financial incentives. In my opinion, the main benefit of GP is its ability to specify areas that could be targeted to promote sustainable behaviours. I feel it is important to validate the presence of indigenous people and empower individual action.

6.5 Future research

This thesis has for the first time used a technique used in criminology and applied it specifically to conservation data sets. I integrated different types of geospatial data into the model to improve its performance. However, despite its success in this field this research is far from complete. There are a number of key areas that would benefit from future development.

Integrating SDM models

In parts this thesis discusses the use of species distribution models in ecology as a standard. It is important to reiterate that these SDM models are different to GP and therefore not directly comparable. In fact, future applications of GP could combine the two. Being able to combine SDM models and GP models we potentially could produce a very powerful tool, which combines two robust methods. Traditionally, GP models do not encompass habitat preferences to identify where species are located and SDM models are based on habitat preferences but do not incorporate spatial information. Combining these two models could produce a map of where the species are coming from and a probability surface of where they are most likely to disperse (Le Comber et al. 2011).

User-friendly application

Visualising data on a map is a highly informative tool and an easily understood technique to present data. Future applications of GP could include the development of a user-friendly application that updates the geoprofile in real time so users can see the map update as they feed in their information. This becomes useful when studies are looking for an elusive or rare species such as the tarsiers in Chapter 2, or invasive species like the mink in Chapter 3. Members of the public could add data as they see an animal, these citizen science data sets could provide a potentially very useful addition to any studies. In addition, if anti-poaching scouts could have access to an application such as this it may be able to help with the war against poaching and illegal wildlife trade.

Genetics

There has been a surge in research studying the inter play between ecology and evolution (Lowe et al. 2017). Future developments of GP could include genetic information. This could be integrated into the DPM model in addition or separately to GIS information. Clusters could be weighted by their spatial location but also their genetic information.

6.5 Concluding statement

This thesis has shown that novel spatial methods such as GP can improve the understanding of animal spatial patterns, and as such can play an important role in the conservation of species. The adaptability of the model allows it to point to specific areas where elusive or invasive animals may be located, to areas where human-wildlife conflict occurs or go back to its criminological roots and identify areas where illegal hunters may be living. In addition, this thesis has explored the integration of complex spatial information into the model to improve its search strategy in cases such as these. The unadjusted geoprofiles produced by the model using the presence data alone provide fundamental information on the distribution and sources of cases. However, the current advancements in GIS technology allow for a much more sophisticated and tailored techniques to be implemented. These developments mean that GP has the potential to support much more effective and efficient investigations into conservation issues.

The work I have presented in this thesis showcase the realised and potential application of GP to conservation and ecological data. It is my genuine hope that this work will lead

to further expansions of the model leading to more efficient and targeted interventions in ecology and conservation.

References

Altieri AH, van Wesenbeeck BK, Bertness MD, Silliman BR. 2010. Facilitation cascade drives positive relationship between native biodiversity and invasion success. *Ecology* **91**: 1269-1275.

Biggs D, Courchamp F, Martin R, Possingham HP. 2013. Legal trade of Africa's rhino horns. *Science* **339**: 1038-1039.

Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM. 2011. A proposed unified framework for biological invasions. *Trends in Ecology & Evolution* **26**: 333-339.

Blanc J, Burnham K, Dublin H, Milliken T, Mwangi P, Sangalakula L, Skinner D, Underwood F. 2013. Status of African elephant populations and levels of illegal killing and the illegal trade in ivory: a report to the African Elephant Summit. CITES Secretariat, Geneva. Available from <https://cites.org/eng/com/sc/61/E61-44-02-A1.pdf> (accessed July 2017).

Bocedi G, Zurell D, Reineking B, Travis JMJ. 2014. Mechanistic modelling of animal dispersal offers new insights into range expansion dynamics across fragmented landscapes. *Ecography* **37**: 1240-1253.

Bonesi L, Palazon S, 2007. The American mink in Europe: Status, impacts, and control. *Biological Conservation* **134**: 470-483.

Bowler DE, Benton. TG. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews* **80**: 205-225.

Brantingham PJ, Brantingham PL. 1981. Notes on the Geometry of Crime. Pages 27-54 in Brantingham PJ, Brantingham PL, editors. *Environmental Criminology*, Sage Publications, CA.

Brantingham PJ, Brantingham PL. 1993. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology* **13**: 3-28.

Brodie, J. F., Giordano, A. J., Dickson, B., Hebblewhite, M., Bernard, H., Mohd-Azlan, J., Anderson, J. and Ambu, L. 2015. Evaluating multispecies landscape connectivity in a threatened tropical mammal community. *Conservation Biology* **29**: 122-132.

Brundson C, Comber L. 2015. An introduction to R for spatial analysis and mapping. In Rojek R editor. SAGE publications, London.

Buscema M, Grossi E, Breda M. Jefferson T. 2009. Outbreaks source: A new mathematical approach to identify their possible location. *Physica A: Statistical Mechanics and its Applications* **388**: 4736-4762.

Campbell-Smith G, Simanjourang HVP, Leader-Williams N, Linkie M. 2010. Local attitudes and perceptions towards crop-raiding by orang-utans (*Pongo abelii*) and other nonhuman primates in Northern Sumatra, Indonesia. *American Journal of Primatology* **71**: 1-11.

Canter D, Larkin P. 1993. The environmental range of serial rapists. *Journal of Environmental Psychology* **13**: 63-69.

Capone DL, Nichols WW. 1976. Urban structure and criminal mobility. *American Behavioural Scientist* **20**: 199-213.

Carter R, Mendis KN, Roberts D. 2000. Spatial targeting of interventions against malaria. *Bulletin of the World Health Organisation* **78**: 1401-1411.

Chadès I, McDonald-Madden E, McCarthy MA, Wintle B, Linkie M, Possingham HP. 2008. When to stop managing or surveying cryptic threatened species. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 13936-40.

Clark P, Evans F. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* **35**: 445-453.

Cockle KL, Martin K, Drever MC. 2010. Supply of tree-holes limits nest density of cavity-nesting birds in primary and logged subtropical Atlantic forest. *Biological Conservation* **143**: 2851-2857.

Conrad CC, Hilchey KG. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment* **176**: 273-91.

Cooney et al. 2016. From Poachers to Protectors: Engaging Local Communities in Solutions to Illegal Wildlife Trade. *Conservation Letters* **10**: 367-374.

Cumming DHM. 2004. Performance of parks in a century of change. Pages 105-124 in Child B, editor. *Parks in transition: biodiversity, rural development and the bottom line*. Earthscan, London.

Dejean T, Valentini A, Miquel C, Taberlot P, Bellemain E, Miaud C. 2012. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology* **49**: 953-959.

Denoël M, Ficetola DF. 2015 Using kernels and ecological niche modeling to delineate conservation areas in an endangered patch-breeding phenotype. *Ecological Applications* **25**: 1922-1931.

Di Minin E, Laitila J, Montesino-Pouzols F, Leader-Williams N, Slotow R, Goodman PS, Conway AJ, Moilanen A. 2015. Identification of policies for a sustainable legal trade in rhinoceros horn based on population projection and socioeconomic models. *Conservation Biology* **29**: 545-555.

Didham RK, Tylianakis JM, Hutchison MA, Ewers RM, Gemmill NJ. 2005. Are invasive species the drivers of ecological change?. *Trends in Ecology & Evolution* **20**: 470-474.

Dinerstein E, Loucks C, Wikramanayake E, Ginsberg J, Sanderson E, Seidensticker J, Forrest J, Bryja G, Heydlauff A, Klenzendorf S, Leimgruber P, Mills J, O'Brien T, Shrestha M, Simons R, Songer M. 2007. The fate of wild tigers. *Bioscience* **57**: 508-514.

Doak DF, Bakker VJ, Goldstein BE, Hale B. 2014. What is the future of conservation? *Trends in Ecology and Evolution* **29**: 77-81.

Doney RH. 1990. The aftermath of the Yorkshire Ripper: the response of the United Kingdom police service. Pages 95-112 in Egger S, editors. *Serial murder: an elusive phenomenon*. Praeger Publishing, Westport, CT.

Du Toit RF. 2004. Review of wildlife issues associated with the land reform programme in Zimbabwe. World Wide Fund for Nature, Southern African regional programme Office.

Du Toit, JT. 2002. Wildlife harvesting guidelines for community- based wildlife management: a southern African perspective. *Biodiversity and Conservation* **11**: 1403-1416.

Elith J, Graham CH. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32**: 66-77.

Elith J, Philips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2010. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**: 43-57.

Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution and systematics* **40**: 677-697.

Epanchin-Niell RS, Hastings A. 2010. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecology Letters* **13**: 528-541.

Fa JE, Brown D. 2009. Impacts of hunting on mammals in African tropical moist forests: a review and synthesis. *Mammal Review* **39**: 231-264.

Faulkner SC, Stevenson MD, Verity R, Mustari AH, Semple S, Tosh DG, Le Comber SC. 2015. Using geographic profiling to locate elusive nocturnal animals: a case study with spectral tarsiers. *Journal of Zoology* **295**: 261-268.

Faulkner SC, Verity R, Roberts D, Roy SS, Robertson PA, Stevenson MD, Le Comber SC. 2017. Using geographic profiling to compare the value of sightings vs trap data in a biological invasion. *Diversity and Distributions* **23**: 104-112.

Faulkner SC, Stevens MCA, Romañach SS, Lindsey PA, Le Comber SC. 2018 A spatial approach to combatting wildlife crime. *Conservation Biology* **32**: 685-693.

Felson M. 2000. The routine activity approach as a general crime theory. Pages 205-216 In Simpson S, editors. *Of Crime and Criminality: The use of theory in everyday life*. Thousand Oaks, CA: Sage.

Ferrier S, Watson G, Pearce J, Drielsma M. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation* **11**: 2275-2307.

Foster RJ, Harmsen BJ. 2012. A critique of density estimation from camera trap data. *The Journal of Wildlife Management* **76**: 224-236.

Fourcade Y, Engler JO, Rodder D, Secondi J. 2014. Mapping Species Distributions with MaxEnt. Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias *PLOS one* **9**: e97122.

Fourcade Y, Besnard AG, Secondi J. 2017. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography* **27**: 245-256.

Gaveau, DLA, Wandono H, Setiabudi F. 2006. Three decades of deforestation in southwest Sumatra: Have protected areas halted forest loss and logging, and promoted re-growth? *Biological Conservation* **134**: 495-504.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004 *Bayesian Data Analysis*, 2nd edition, Chapman & Hall/CRC, Boca Raton, Florida.

Geman S, Geman D. 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine intelligence* **6**: 721-741.

Gibbs C, Gore ML, McGarrell EF, Rivers L III. 2010. Introducing Conservation Criminology; Towards interdisciplinary scholarship on environmental crimes and risks. *British Journal of Criminology* **50**: 124-44.

Gillespie GR, Lockie D, Scroggie MP, Iskandar DT. 2004. Habitat use by stream-breeding frogs in south-east Sulawesi, with some preliminary observations on community organization. *Journal of Tropical Ecology* **20**: 439-448.

Gini C. 1921. Measurement of Inequality of Incomes. *The Economic Journal*. Blackwell Publishing **31**: 124-126.

Grantham HS, Wilson KA, Moilanen A, Rebelo T, Possingham HP. 2009. Delaying conservation actions for improved knowledge: how long should we wait?. *Ecology Letters* **12**: 293-301.

Greig-Smith P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany* **13**: 293-316.

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM. 2013. Predicting species distributions for conservation decisions. *Ecology Letters* **16**: 1424-1435.

Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**: 993-1009.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological modelling* **135**: 147-186.

Gurevitch J, Padilla DK. 2004. Are invasive species a major cause of extinctions?. *Trends in Ecology & Evolution* **19**: 470-474.

Gusset M, Swarner MJ, Mponwanek L, Keletiele K, McNutt JW. 2009. Human–wildlife conflict in northern Botswana: livestock predation by Endangered African wild dog *Lycaon pictus* and other carnivores. *Oryx* **43**: 67-72.

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JRG. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**: 850-853.

Hardwick JL, Priston NEC, Martin TE, Tosh DG, Mustari AH, Abernethy KE. 2017. Community perceptions of the crop-feeding Buton Macaque (*Macaca ochreata brunnescens*): an ethnoprimateological study on Buton Island, Sulawesi. *International Journal of Primatology* **38**: 1-18.

Harvey CT, Qureshi SA, Maclsaac HJ. 2009. Detection of a colonizing, aquatic, non-indigenous species. *Diversity and Distribution* **15**: 429-437.

Hiers JK, Jackson ST, Hobbs RJ, Bernhardt ES, Valentine LE. 2016. The precision problem in conservation and restoration. *Trends in Ecology and Evolution* **31**: 820-830.

Hirzel AH, Hausser J, Chessel D, Perrin N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* **83**: 2027-2036.

Holden MH, Nyrop JP, Ellner SP. 2016. The economic benefit of time-varying surveillance effort for invasive species management. *Journal of Applied Ecology* **53**: 712-721.

Hoogerwerf A. 1970. *Ujung Kulon: The Land of the Last Javan Rhinoceros*. E.J. Brill, Leiden, The Netherlands.

Howard C, Stephens PA, Pearce-Higgins JW, Gregory RD, Willis SG. 2014. Improving species distribution models: the value of data on abundance. *Methods in Ecology and Evolution* **5**: 506-513.

Hulme PE. 2007. Biological invasions in Europe: drivers, pressures, states, impacts and responses. In Hester R & Harrison RM editors. *Biodiversity under threat*. Cambridge, UK: Cambridge University Press.

Hulme PE. 2008. Phenotypic plasticity and plant invasions: is it all Jack? *Functional Ecology* **1**: 3-7.

IUCN. 2013. IUCN Red List of Threatened Species. Version 2013.1. Online: www.iucnredlist.org. Accessed on 02.09.13.

Jędrzejewski W, Carreño R, Sánchez-Mercado A, Schmidt K, Abarca M, Robinson HS, Boede EO, Hoogsteijn R, Vilorio AL, Cerda H, Velásquez G, Zambrano-Martínez S. 2017. Human-jaguar conflicts and the relative importance of retaliatory killing and hunting for jaguar (*Panthera onca*) populations in Venezuela. *Biological Conservation* **209**: 524-532.

Jennings AP, Mathai J, Brodie J, Giordano AJ, Veron G. 2013 Predicted distributions and conservation status of two threatened small carnivores: banded civet and Hose's civet. *Mammalia* **77**: 261-271.

Jonsen, ID, Myers RA, Flemming JM, 2003. Meta-analysis of animal movement using state-space models. *Ecology* **88**: 3055-3063.

Keough H, Blahna D. 2006. Achieving integrative, collaborative ecosystem management. *Conservation Biology* **20**: 1373-1382.

Knapp EJ. 2012. Why poaching pays: a summary of risks and benefits illegal hunters face in Western Serengeti, Tanzania. *Tropical Conservation Science* **5**: 434-445.

Le Comber SC, Nicholls B, Rossmo DK, Racey PA. 2006. Geographic profiling and animal foraging. *Journal of Theoretical Biology* **240**: 223-240.

Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC. 2011. Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics* **10**: 35-42.

Le Comber SC, Stevenson MD. 2012. From Jack the Ripper to epidemiology and ecology. *Trends in Ecology and Evolution* **27**: 307-308.

Legg C, Nagy L. 2006. Why most conservation monitoring is but need not be, a waste of time. *Journal of Environmental Management* **78**: 194-199.

Lindenmayer DB, Blanchard W, McBurney L, Blair D, Banks S, Likens GE, Franklin JF, Laurence WF, Stein JAR, Gibbons P. 2012. Interacting Factors Driving a Major Loss of Large Trees with Cavities in a Forest Ecosystem. *PLoS ONE* **7**: e41864

Lindsey PA et al. 2013. The bushmeat trade in African savannahs: Impacts, drivers, and possible solutions. *Biological Conservation* **160**: 80-96.

Lindsey PA, Blame GA, Funston PJ, Henschel PH, Hunter LTB. 2016. Life after Cecil: channelling global outrage into funding for conservation in Africa. *Conservation Letters* **9**: 296-301.

Lindsey PA, duTroit R, Pole A, Romañach S. 2009. Savé Valley Conservancy: a large scale African experiment in cooperative wildlife management. Pages 163-184 in Child B,

Suich H, Spenceley A, editors. Evolution and innovation in wildlife conservation: parks and game ranches to transfrontier conservation areas, Earthscan, London.

Lindsey PA, Petracca LS, Funston PJ, Bauer H, Dickman A, Everatt K, Flyman M, Henschel P, Hinks AE, Kasiki S, Loveridge A. 2017. The performance of African protected areas for lions and their prey. *Biological Conservation* **209**: 137-149.

Lindsey PA, Romañach SS, Tambling CJ, Chartier K. 2011. Ecological and financial impacts of illegal bushmeat trade in Zimbabwe *Oryx* **45**: 96-111.

Linkie M, Martyr D, Holden J, Yanuar A, Hartana A, Sugardjito J, Leader-Williams N. 2003. Habitat destruction and poaching threaten the Sumatran tiger in Kerinci Seblat National Park, Sumatra. *Oryx*, **37**: 41-48.

Linkie M, Martyr DJ, Harihar A, Risdianto D, Nugraha RT, Maryati, Leader-Williams N, Wong W. 2015. Safeguarding Sumatran tigers: evaluating effectiveness of law enforcement patrols and local informant networks. *Journal of Applied Ecology*. **52**: 851-860.

Lotter W, Clark K. 2014. Community involvement and joint operations aid effective anti-poaching. *Parks* **20**: 19-28.

Lowe WH, Kovach RP, Allendorf FW. 2017. Population genetics and demography unite ecology and evolution. *Trends in ecology and evolution* **32**: 141-152.

Lynch MJ, Stretesky PB. 2003. The Meaning of Green; contrasting criminological perspectives. *Theoretical Criminology* **7**: 217-38.

MacKenzie DI, Nicholds JD, Sutton N, Kawanishi K, Bailey LL. 2005. Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* **86**: 1101-1113.

MacKinnon J, MacKinnon K. 1980. The behaviour of wild spectral tarsiers. *International Journal of Primatology* **1**: 361-379.

Madden F. 2004. Creating coexistence between humans and wildlife: global perspectives on local efforts to address human- wildlife conflict. *Human Dimensions of Wildlife* **9**: 247–257.

Manfredo MJ, Dayer AA. 2004. Concepts for exploring the social aspect of human-wildlife conflict in a global context. *Human dimensions of wildlife* **9**: 317-328.

Mansourian S, Dudley N. 2008. Public funds to protected areas. WWF, Gland, Switzerland.

Martin RA, Rossmo DK, Hammerschlag N. 2009. Hunting patterns and geographic profiling of white shark predation. *Journal of Zoology* **279**: 111-118.

Martin TE, Blackburn GA. 2010. Impacts of tropical forest disturbance upon avifauna on a small island with high endemism: implications for conservation. *Conservation and Society* **8**: 127.

Martin TG, Nally S, Burbidge AA, Arnall S, Garnett ST, Hayward MW, Lumsden LF, Menkhorst P, McDonald-Madden E, Possingham HP. 2012. Acting fast helps avoid extinction. *Conservation Letters* **5**: 274-280.

Martins TLF, Brooke Mde.L, Hilton GM, Farnsworth S, Gould J, Pain DJ. 2006. Costing eradications of alien mammals from islands. *Animal Conservation* **9**: 439-444 .

Mayfield H, Smith C, Gallagher M, Hockings M. 2017. Use of freely available datasets and machine learning methods in predicting deforestation. *Environmental Modelling & Software* **87**: 17-28.

Meaney R. 2004. Commuters and marauders: an examination of the spatial behaviour of serial criminals. *Journal of Investigative Psychology and Offender Profiling* **1**: 121-137.

Merker S, Thomas S, Volker, E, Perwitasari-Farajallah B, Streit B, Pfenninger M. 2014. Control region length dynamics potentially drives amino acid evolution in tarsier mitochondrial genomes. *Journal of Molecular Evolution* **79**: 40-51.

Merker S, Yustian I, Mühlenberg M. 2005. Responding to forest degradation: altered habitat use by Dian's tarsier *Tarsius diana*e in Sulawesi, Indonesia. *Oryx* **39**: 189-195.

Merker S. 2006. Habitat-Specific Ranging Patterns of Dian's Tarsiers (*Tarsius diana*) as Revealed by Radio tracking. *American Journal of Primatology* **68**: 111-125.

Miettinen J, Shi CH, Liew SC. 2011*b*. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology* **17**: 2261-2270.

Miettinen J, Shi CH, Tan WJ, Liew SC. 2011*a*. 2010 land cover map of insular Southeast Asia in 250-m spatial resolution. *Remote Sensing Letters* **3**: 11-20.

Myers JH, Simberloff D, Kuris AM, Carey JR. 2000. Eradication revisited: dealing with exotic species. *Trends in Ecology & Evolution* **15**: 316-320.

Naidoo R, Balmford A, Ferraro PJ, Polasky S, Ricketts TH, Rouget M. 2006 Integrating economic costs into conservation planning. *Trends in Ecology and Evolution* **21**: 681-687.

Naidoo R, Fisher B, Manica A, Balmford A. 2016. Estimating economic losses to tourism in Africa from the illegal killing of elephants. *Nature Communications* **7**: 13379.

Neilson E, Nijman V, Nekaris KAI. 2013. Conservation assessments of arboreal mammals in difficult terrain: occupancy modeling of pileated gibbons (*Hylobates pileatus*). *International Journal of Primatology* **34**: 823-835.

Neri-Arboleda I, Stott P, Arboleda NP. 2002. Home ranges, spatial movements and habitat associations of the Philippine tarsier (*Tarsius syrichta*) in Corella, Bohol. *Journal of Zoology* **257**: 387-402.

Nietsch A. 1999. Duet vocalizations among different populations of Sulawesi tarsiers. *International Journal of Primatology* **20**: 567-583.

Nogales M, Rodriguez-Luengo JL, Marrero P. 2006. Ecological effects and distribution of invasive non-native mammals on the Canary Islands. *Mammal Review* **36**: 49-65.

Nyhus PJ, Tilson R. 2004. Characterizing human-tiger conflict in Sumatra, Indonesia: implications for conservation. *Oryx* **38**: 68-74.

O'Brien TG, Kinnaird MF, Wibisono HT. 2003. Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape. *Animal Conservation* **6**: 131-139.

O'Leary M. 2009. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling* **6**: 253-265.

O'Leary M. 2010a. Implementing a Bayesian approach to criminal geographic profiling. First International Conference on Computing for Geospatial Research and Application, June 21-23 Washington, D.C.

O’Leary M. 2010*b*. Multimodel inference and geographic profiling. *Crime Mapping* **2**.

O’Leary M. 2012. New mathematical approach to geographic profiling. National Institute of Justice, Washington, D.C.

Papini A, Sordo L, Mosti S. 2010. Surface interactions of the epiphytic macroalga *Hincksia mitchelliae* (phaeophyceae) with the shoalgrass *Halodule wrightii*. 2010. *Journal of Phycology* **47**: 118-122.

Papini A, Mosti S, Santosuosso U. 2013. Tracking the origin of the invading *Caulerpa* (Caulerpales, Chlorophyta) with Geographic Profiling, a criminological technique for a killer alga. *Biological Invasions* **15**: 1613-1621.

Park R, Burgess E, Janowitz M. 1921. *Introduction to the Science of Sociology*. Chicago: University of Chicago Press

Parks and wildlife act 20 14. Available from <http://www.parlzim.gov.zw/acts-list/parks-and-wild-life-act-20-14> (accessed June 19, 2017).

Peres CA. 2009. Overexploitation. Pages107–130 in Sodhi NS, Ehrlich PR, editors. *Conservation Biology for All* Oxford, University Press, Oxford.

Peterson AT, Soberon J, Pearson RG, Anderson RP, Martinez-Meyer E, Nakamura M, Araujo MB. 2011. Ecological Niches and Geographic Distributions. Princeton University Press, Princeton, USA

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-175.

Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T, Tsomondo T. 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems & Environment* **84**: 1-20.

Pluess T, Cannon R, Jarošík V, Pergl J, Bacher S. 2012. When are eradication campaigns successful? A test of common assumptions. *Biological Invasions* **14**: 1365-1378.

Pozzi L, Hodgson JA, Burrell AS, Sterner KN, Raaum RL, Disotell TR. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution* **75**: 165-183.

Priston NE, Wyper RM, Lee PC. 2011. Buton macaques (*Macaca ochreata brunnescens*): crops, conflict, and behaviour on farms. *American Journal of Primatology* **73**: 1-8.

Pyšek P, Jarošík V, Hulme PE, Pergl J, Hejda M, Schaffner U, Vilà M. 2012. A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Global Change Biology* **18**: 1725-1737.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, <www.R-project.org>.

Raine NE, Rossmo DK, Le Comber SC. 2009. Geographic profiling applied to testing models of bumble-bee foraging. *Journal of the Royal Society Interface* **6**: 307-319.

Ripple WJ et al. 2015. Collapse of the worlds largest herbivores. *Science Advances* **4**: e1400103.

Ripple WJ et al. 2016. Saving the world's terrestrial megafauna. *BioScience* **66**: 807-812.

Risdianto D, Martyr DJ, Nugraha RT, Harihar A, Wibisono HT, Haidir IA, Macdonald DW, D’Cruze N, Linkie M. 2016. Examining the shifting patterns of poaching from a long-term law enforcement intervention in Sumatra. *Biological Conservation* **204**: 306-312.

Robertson PA, Adriaens T, Lambin X, Mill A, Roy S, Shuttleworth CM, Sutton-Croft M. 2015. The large-scale removal of mammalian invasive alien species in Northern Europe. *Pest Management Science* DOI: 10.1002/ps.4224.

Rossmo DK, Velarde L. 2008. Geographic profiling analysis: principles, methods, and applications. Pages 35-43 in Chainey S, Tompson L, editors. *Crime mapping case studies: practice and research*. John Wiley and Sons Ltd, Chichester, UK.

Rossmo DK. 2000. *Geographic profiling*. CRC Press, Florida.

Rossmo DK. 2012. Recent developments in geographic profiling. *Policing* **6**: 144-150.

Rossmo DK. 1993. A methodological model. *American Journal of Criminal Justice* **17**: 1-21.

Roy SS, Chauvenet ALM, Robertson PA. 2015. Removal of American mink (*Neovison vison*) from the Uists, Outer Hebrides, Scotland. *Biological Invasions* **17**: 2811-2820.

Roy SS. 2011. Strategies to improve landscape scale management of mink populations in the west coast of Scotland: lessons learned from the Uists 2001-2006. *Island Invasives: Eradication and Management*. IUCN, Gland, Switzerland.

Sánchez-Mercado A, Asmüssen M, Rodríguez-Clark KM, Rodríguez JP, Jedrzejewski W. 2016. Using spatial patterns in illegal wildlife uses to reveal connections between subsistence hunting and trade. *Conservation Biology* **30**: 1222-1232.

Sanderson EW, Forrest J, Loucks C, Ginsberg J, Dinerstein E, Seidensticker J, Leimgruber P, Songer M, Heydlauff A, O'Brien T, Bryja G, Klenzendorf S, Wikramanayake E. 2010. Pages 144-161 in *Tigers of the World*, R. Tilson, P. J. Nyhus, Eds. (Academic Press, London, ed. 2, 2010).

Seidensticker J. 1987. Bearing Witness: Observations on the extinction of *Panthera tigris balica* and *Panthera tigris sondaica*. Pages 1-8 In *Tigers of the World: The Biology, Biopolitics, Management, and Conservation of an Endangered Species* (eds R.L. Tilson & U.S. Seal. Noyes Publications, Park Ridge, USA

Seidensticker J. 2010. Saving wild tigers: A case study in biodiversity loss and challenges to be met for recovery beyond 2010. *Integrative Zoology* **5**: 285-299.

Shaw CR, McKay H. 1942. Juvenile delinquency and urban areas. Pages 435-446 in EDITOR. *The university of Chicago press*, Chicago.

Shekelle M, Groves C, Merker S, Supriatna J. 2008a. *Tarsius tumpara*: a new tarsier species from Siau Island, North Sulawesi. *Primate Conservation* **23**: 55-64.

Shekelle M, Merker S, Salim A. 2008b. *Tarsius lariang*. In: IUCN 2013. IUCN Red List of Threatened Species. Version 2013. <www.iucnredlist.org>

Sillero-Zubiri, C, Laurenson MK. 2001. Interactions between carnivores and local communities: conflict or co- existence? Pages 282-312 In Carnivore Conservation (eds J.L. Gittleman, S.M. Funk, D. Macdonald & R.K. Wayne). Cambridge University Press, Cambridge, UK.

Simberloff D. 2003. How Much Information on Population Biology Is Needed to Manage Introduced Species?. *Conservation Biology* **17**: 83-92.

Smith CM, Downs SH, Mitchell A, Hayward AC, Fry H, Le Comber SC. 2015. Spatial targeting for bovine tuberculosis control: can the locations of infected cattle be used to find infected badgers? *PLoS One* e0142710

Smith GD. 2002. Commentary: Behind the Broad Street pump: aetiology, epidemiology and prevention of cholera in mid-19th century Britain. *International journal of epidemiology* **32**: 920-932

Snäll T, Kindvall O, Nilsson J, Pärt T. 2011. Evaluating citizen-based presence data for bird monitoring. *Biological Conservation* **144**: 804-810.

Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC. 2012. Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography* **35**:704-715.

Stone PA, Snell HL, Snell HM. 1994. Behavioural diversity as biological diversity: Introduced cats and lava lizard wariness. *Conservation Biology* **8**: 569-573.

Struebig MJ, Turner A, Giles E, Lasmana F, Tollington S, Bernard H, Bell D. 2013. Quantifying the biodiversity value of repeatedly logged rainforests: Gradient and comparative approaches from Borneo. *Advances in Ecological Research* **48**: 183-224.

Sullivan BL, Wood CL, Iliff MJ, Bonney RE, Fink D, Kelling, S. 2009. eBrid: A citizen-based bird observation network in the biological sciences. *Biological Conservation* **142**: 2282-2292.

Synes NW, Osborne PE. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography* **20**: 904-914.

Thurman QC, Jamieson JD. 2015. *Police Problem Solving*. London: Routledge.

Trakhtenbrot A, Nathan R, Perry G, Richardson DM. 2005. The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions* **11**: 173-181.

Trapping of animal control act 20 21. Available from <http://www.parlzim.gov.zw/acts-list/trapping-of-animals-control-act-20-21> (accessed April 15, 2017).

Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP. 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications* **13**: 1790-1801.

Verity R, Stevenson MD, Rossmo DK, Nichols RA, Le Comber SC. 2014. Spatial targeting of infectious disease control: indentifying multiple, unknown sources. *Methods in Ecology and Evolution* **5**: 647-655.

Vitousek PM, D'Antonio CM, Loope LL, Westbrooks R. 1996. Biological invasions as global environmental change. *American Scientist* **84**: 468-478.

Václavík T, Meentemeyer RK. 2009. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* **220**: 3248-3258.

Wang SW, Lassoie J P, Curtis PD. 2006. Farmer attitudes towards conservation in Jigme Singye Wangchuck National Park, Bhutan. *Environmental Conservation* **33**: 148-156.

Waters JK. 1998 The geography of Crime. *Mercator's World* **3**: 46-51

Wellsmith M. 2011. Wildlife Crime: The Problems of Enforcement. *European Journal on Criminal Policy and Research* **17**:125-148.

Whitten T, Mustafa M, Henderson GS. 2002. The ecology of Sulawesi. Vol. 4. The Ecology of Indonesia. Periplus Press, Singapore.

Whitten T. 2000. The ecology of Sumatra Vol. 1. Tuttle Publishing.,

Wikramanayake ED, Dinerstein E, Robinson G, Karanth UK, Rabinowitz A, Olson D, Matthew T, Hedao P, Connor M, Hemley G. Bolze D. 1998. An ecology-based method of defining priorities for large mammal conservation: the tiger as a case study. *Conservation Biology* **12**: 865-878.

Wilcove DS, Rothstein D, Dubow J, Philips A, Losos E. 1998. Quantifying threats to imperiled species in the United States. *Bioscience* **48**: 607-615.

Wiltling A, Courtiol A, Christiansen P, Niedballa J, Scharf AK, Orlando L, Balkenhol N, Hofer H, Kramer-Schadt S, Fickel J, Kitchener AC. 2015. Planning tiger recovery: understanding intraspecific variation for effective conservation. *Science Advances* **1**: e1400175.

Zhang Y, Chen HYH, Taylor A. 2014. Multiple drivers of plant diversity in forest ecosystems. *Global Ecology and Biogeography* **8**: 885-893.