

# Modelling the effects of ecology on wildlife disease surveillance

Laura Walton

Doctor of Philosophy  
University of York

Environment

September 2014

# Abstract

Surveillance is the first line of defence against disease, whether to monitor endemic cycles or to detect emergent epidemics. Knowledge of disease in wildlife is of considerable importance for managing risks to humans, livestock and wildlife species. Recent public health concerns (e.g. Highly Pathogenic Avian Influenza, West Nile Virus, Ebola) have increased interest in wildlife disease surveillance. However, current practice is based on protocols developed for livestock systems that do not account for the potentially large fluctuations in host population density and disease prevalence seen in wildlife.

A generic stochastic modelling framework was developed where surveillance of wildlife disease systems are characterised in terms of key demographic, epidemiological and surveillance parameters. Discrete and continuous state-space representations respectively, are simulated using the Gillespie algorithm and numerical solution of stochastic differential equations. Mathematical analysis and these simulation tools are deployed to show that demographic fluctuations and stochasticity in transmission dynamics can reduce disease detection probabilities and lead to bias and reduced precision in the estimates of prevalence obtained from wildlife disease surveillance. This suggests that surveillance designs based on current practice may lead to underpowered studies and provide poor characterisations of the risks posed by disease in wildlife populations. By parameterising the framework for specific wildlife host species these generic conclusions are shown to be relevant to disease systems of current interest.

The generic framework was extended to incorporate spatial heterogeneity. The impact of design on the ability of spatially distributed surveillance networks to detect emergent disease at a regional scale was then assessed. Results show dynamic spatial reallocation of a fixed level of surveillance effort led to more rapid detection of disease than static designs.

This thesis has shown that spatio-temporal heterogeneities impact on the efficacy of surveillance and should therefore be considered when undertaking surveillance of wildlife disease systems.

# Contents

<b>Abstract</b> .....	<b>2</b>
<b>Contents</b> .....	<b>3</b>
<b>List of Tables</b> .....	<b>6</b>
<b>List of Figures</b> .....	<b>7</b>
<b>Acknowledgements</b> .....	<b>9</b>
<b>Author's Declaration</b> .....	<b>10</b>
<b>1. Introduction</b> .....	<b>11</b>
1.1 Wildlife diseases and Humans.....	11
1.2 Wildlife disease and conservation.....	14
1.3 Wildlife diseases and the livestock industry .....	15
1.4 Wildlife Disease Surveillance.....	17
1.4.1 The Probability of Detection.....	20
1.4.2 Estimating the True Prevalence: Bias and Standard Deviation .....	22
1.4.3 Improving wildlife disease surveillance .....	24
1.5 Modelling Wildlife Populations, Disease and Surveillance .....	25
1.5.1 Temporal Modelling of Wildlife Populations with Disease .....	26
1.5.2 Spatial Temporal Modelling of Wildlife Populations, Disease and Surveillance .....	35
1.6 The Thesis.....	36
1.6.1 Aims .....	36
1.6.2 Thesis structure .....	36
<b>2. The Ecology of Wildlife Disease Surveillance</b> .....	<b>38</b>
2.1 Abstract .....	39
2.2 Introduction.....	39
2.3 Methods .....	41
2.3.1 Stochastic Model Description.....	41
2.3.2 Statistics generated from the model.....	42
2.3.3 Model Implementation.....	42
2.4 Results .....	44
2.4.1 Estimating Prevalence .....	44
2.4.1.1 Effect of host demography and disease dynamics .....	45

2.4.1.2 Surveillance design .....	47
2.4.2 The Probability of Detection.....	49
2.4.2.1 Effect of host demography and transmission dynamics.....	51
2.4.2.2 Limits to disease detection in wildlife disease systems.....	52
2.5 Discussion.....	54
<b>3. An Approach to Assessing Wildlife Disease Surveillance in Real Systems: tools and applications .....</b>	<b>57</b>
3.1 Abstract .....	58
3.2 Introduction.....	58
3.3 Methods .....	61
3.3.1 Stochastic Model Description.....	61
3.3.2 Model Implementation.....	61
3.3.3 Model Parameterisation.....	62
3.3.4 Simulations .....	63
3.3.5 Statistics generated from the model.....	65
3.4 Results .....	65
3.4.1 Estimating disease prevalence in badger and rabbit populations .....	66
3.4.1.1 Effect of disease system .....	66
3.4.2 Detecting disease.....	74
3.4.3 Explicit sources of bias.....	76
3.4.4 Imperfect diagnostic tests .....	78
3.5 Discussion.....	82
<b>4. Spatio-Temporal modelling of habitat composition, ecology and wildlife disease surveillance: impacts of emergent disease.....</b>	<b>86</b>
4.1 Abstract .....	87
4.2 Introduction.....	87
4.3 Methods .....	89
4.3.1 Stochastic Model Description .....	89
4.3.2 Model Implementation.....	91
4.3.3 Statistics generated from the model.....	93
4.4 Results	
4.4.1 Static surveillance: spatial distribution of effort.....	94
4.4.2 Static surveillance: stratified designs .....	96
4.4.3 Dynamic designs .....	101
4.5 Discussion.....	105

<b>5. General Discussion: Towards a new approach to wildlife disease surveillance.....</b>	<b>109</b>
5.1 Background .....	109
5.2 Fluctuations undermine wildlife disease surveillance .....	110
5.3 Tools for assessing wildlife disease surveillance in real systems .....	112
5.4 Spatial heterogeneity and the design of wildlife disease surveillance .....	113
5.5 Future Work .....	114
<b>Appendices .....</b>	<b>118</b>
<b>Appendix 1 .....</b>	<b>119</b>
1.1 Relationship between discrete and continuous (SDE) state-space model implementations.....	119
1.2 Parameterisations used .....	124
<b>Appendix 2 .....</b>	<b>128</b>
Appendix 2.1 .....	128
2.1.1 Estimating Badger Parameters .....	128
2.1.1.1 Discrete to continuous time .....	128
2.1.1.2 Translating from multi- to single-stage models.....	129
2.1.2 Estimating Rabbit Parameters .....	133
<b>Appendix 2.2.....</b>	<b>135</b>
2.2.1 Parameterisations used .....	135
<b>Appendix 3 .....</b>	<b>144</b>
3.1 Parameterisations used.....	144
<b>References.....</b>	<b>150</b>

# List of Tables

1.1	SI model with rate and effects for each event type	29
1.2	Expectations and variance-covariance rates	33
1.3	Expectation and variance-covariance	34
2.1	Model structure	43
4.1	Event, Rate and Effect on the State Space of the model	92

# List of Figures

1.1	Effect of disease prevalence on the sample size required to detect disease	22
1.2	Using binomial theory to estimate the expected bias and std dev in the surveillance estimate of prevalence	24
1.3	An example of a logistic growth curve	28
1.4	Deterministic and Stochastic Time Trajectories	29
2.1	Effect of host demography and disease transmission	46
2.2	Effect of surveillance design	48
2.3	Effect of host-pathogen and surveillance dynamics on probability of detection	52
2.4	Fluctuations reduce power to detect disease	53
3.1	Trace plots from SDE and Gillespie implementation runs of the model	64
3.2	Effect of infected population stability and disease transmission on surveillance in Badgers	68
3.3	Effect of infected population stability and disease transmission on surveillance in Rabbits	70
3.4	Effect of disease transmission and infected population death rate on surveillance	71
3.5	Effect of surveillance design	72
3.6	Effect of surveillance design	73
3.7	Effect of host-pathogen and surveillance dynamics on the probability of detection	75
3.8	Effect of trappability and surveillance dynamics in Badgers	77
3.9	Effect of test sensitivity and disease dynamics in Badgers	79
3.10	Effect of disease dynamics and test sensitivity in Badgers	81
4.1	Effect of the distribution of surveillance effort between patches	95
4.2	The effect of habitat quality on stratified sampling schemes	97
4.3	The effect of the suitability of less favourable on surveillance	99

4.4	The effect of surveillance effort and secondary transmission	100
4.5	Effect of switching	102
4.6	Switching and deployed effort	103
4.7	The effect of between patch transmission for different switching levels	104

# Acknowledgements

Firstly, I would like to thank my supervisor's Glenn Marion and Mike Hutchings for taking a leap of faith on a 22 year old girl about to graduate with an undergraduate degree in statistics. It has been a privilege to work so closely with two people of such knowledge and expertise. Glenn's relentless comments on my many re-drafts have been invaluable in this process. He has been my main supervisory support, I've always been able to go to him with any problem, and have never felt that any question was too simple. Mike's knowledge of ecology and his ability to relate my theoretical results to real systems has been instrumental in developing this research. His enthusiasm for this work has never faltered, which has had a positive influence on my own motivation and enthusiasm! This whole process has been a learning curve in more ways than one, and I couldn't have done it without the support of both Glenn and Mike.

Secondly, I would like to thank Ross Davidson, for his time, patience and expertise in C++. His guidance and problem solving skills have been extremely helpful and have aided my programming development, which is a skill that I will carry forward. I would also like to thank my supervisor Piran White for always being happy to help, answering questions and providing feedback. His independent perspective has helped greatly in formulating arguments and developing key messages.

Thirdly I would like to thank some important people in my life. My parents have been the biggest rock, always believing in me, even at my lowest points. They have listened to my worries, my joy, put up with my bad moods and took an interest in my theories and results, all with love and support. They have helped me above and beyond and I couldn't ask for a better family. I would also like to thank my friend Rachael for having an ear to listen and a shoulder to cry on! Rachael has celebrated the good times with me and distracted me from the bad, and has always found words of encouragement when I needed them most.

I would also like to thank Kevin Kane for employing me, having an income has come in very useful in my fourth year! He has been very understanding and flexible with my working hours in these hectic last few months, and I am truly grateful. Finally, thank you to Team AML for their friendship and humour which have made my transition to London so much easier. They have told me time and again that I should hurry up and finish my PhD so I can join them in cocktails during happy hour. Now that time has come, and I say cheers to that!

# Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other University. All sources are acknowledged as references.

# Chapter 1

## Introduction

Wildlife diseases have the potential, not only to impact greatly on the populations of wildlife species themselves, but also on human and livestock populations. A demonstrative example of how detrimental zoonoses can be to human health and the economy is the recent swine flu pandemic in 2009, which has origins in both domestic pigs and wild boar (Shoham 2011). The disease outbreak started in Mexico and the USA and quickly spread worldwide followed by vast media coverage. The official worldwide death toll according to the World Health Organisation (WHO) as of 28<sup>th</sup> March 2010 was 17,483 (Girard *et al.* 2010). As well as the impact on human health, the economic impact of Mexico alone was estimated as > \$3.2 billion (Girard *et al.* 2010). Wildlife diseases also have the opportunity to affect biodiversity and conservation efforts (Daszak *et al.* 2000; Smith *et al.* 2009; Pagán *et al.* 2012). Mathematical modelling is a tool which can be (and is widely) used to simulate disease systems (Renshaw 1991; Keeling & Rohani 2007) by manipulating equations in order to gain insights into the behaviour of the modelled system. Mathematical modelling will be used in this thesis in such a way to understand the effect that host fluctuations, disease dynamics and spatial heterogeneities have on the efficacy of wildlife disease surveillance. The literature surrounding wildlife disease is reviewed in the next sections and the impact it has on human health, wildlife health and conservation and the economy is outlined to show the importance of wildlife disease research.

### 1.1 Wildlife diseases and Humans

There is an increasing understanding in the literature that wildlife diseases pose a threat to human health (Bengis *et al.* 2004; Belant & Deese 2010; Kuiken *et al.* 2011). The World Organisation for Animal Health (OIE) stated in an editorial that 'Surveillance of wildlife diseases must be considered equally as important as surveillance and control of diseases in domestic animals' (Vallat 2008). In a comprehensive literature review carried out by Talyor *et al.*, it was reported that out of all the infectious organisms known to be pathogenic to

humans, 61% were zoonotic. It was also found that out of the 175 diseases considered to be “emerging” (Lederberg *et al.* 1992), 75% were zoonotic (Taylor *et al.* 2001). This was higher than expected by the authors and is indicative of the importance of wildlife disease research in terms of predicting and controlling emerging outbreaks, and promoting human health and safety. In comparison with human and livestock systems, with wildlife disease there are many added complications in terms of population demography and habitat location, and even though momentum is building behind wildlife disease research it is still the most poorly understood (Jones *et al.* 2008).

There are many examples of zoonotic diseases in humans for which wildlife species act as an intermediary for disease transmission. Nipah is an RNA virus initially detected in pigs in Malaysia in 1999 (when at the same time it appeared that pig farmers were suffering from an outbreak of viral encephalitis), it is closely related to the Hendra Virus discovered in 1994 in Australia. Host infection by Nipah virus is associated with a marked respiratory and neurological syndrome which can be followed by the sudden death of pigs. In the later stages of the initial outbreak of this disease, Nipah was characterised as causing a high mortality rate in humans when it emerged that the same causative agent was to blame for both the pig and pig farm workers mysterious illnesses. The outbreak had devastating consequences for the Malaysian Peninsular’s pig farming industry with an overall loss of 1.08 million pigs and a reduction in pig farms from 1885 to 829 (Nor *et al.* 2000). Although the catalyst of the original outbreak of Nipah virus in pig farmers was domestic pigs (Chua *et al.* 1999), the natural host for the disease is fruit bats. The main drivers associated with the spread of this virus were identified as deforestation, drought and the expansion of pig farming in Malaysia. Fruit bats were forced to move out of their natural habitat as a result of depleting resources available to them, and nearby agricultural areas with productive fruit orchards were an appealing choice. These areas were also home to a large number of pigs which increased the chance of transmission from bat to pig and as a knock on effect, from pig to human in the initial outbreak (Bengis *et al.* 2004). The subsequent 2001 Indian and Bangladeshi outbreaks of Nipah virus were assumed to be caused by the ingestion of fruit and fruit related products that had been contaminated with fruit bat saliva and urine (WHO 2004).

Hantavirus Pulmonary Syndrome (HPS) is an infectious respiratory disease which is endemic to the Americas. The natural host of HPS is thought to be rodents, and is typically carried by the deer mouse (*peromyscus maniculatus*) in the United States. Transmission from wildlife

to humans is through the respiratory route as a result of air-borne dispersion of rodent excretions (Bengis *et al.* 2004), and a study carried out in 1998 of 177 cases of HPS in 29 states, estimated the case-fatality proportion to be 45% (Young *et al.* 1998). Outbreaks of Hantavirus in humans in the United States have frequently been linked to changes in environmental conditions which are highly favourable to the rodent populations (Epstein 1995). The HPS epidemic in the American Southwest of 1993 followed an abundance of rainfall increasing the food sources available for the deer mice. The high population densities driven out of their burrows by flooding increased the chance for the virus to flourish and transmit within rodent populations and eventually onto human populations (Epstein 1995) which can be likened to the perturbation effect (Tuytens *et al.* 2000; Carter *et al.* 2007). Human activities such as herding livestock and cleaning out rarely used rodent infested areas have been attributed to increasing the risk of human exposure to HPS through contact with infected urine and faeces (Armstrong *et al.* 1995).

Avian Influenza (perhaps the highest profile wildlife disease threat given volume of past epidemics and the risk of a pandemic outbreak in the future (Swayne 2009)) is an infectious viral disease found in birds (particularly wild water fowl) caused by strains of the flu virus not unlike the strains found in humans. In wild bird populations it often causes no clinical signs. Some forms of avian influenza have mutated in such a way that they are able to transfer from birds to human populations. An example of this was demonstrated in 1997, when an outbreak of H5N1 (previously thought to only infect birds) was first reported in Hong Kong causing 18 cases of infection with a 33% mortality rate. The clinical progression in humans of this Influenza strain can be categorised into three stages. The first stage of the illness would usually show mild upper respiratory tract infection and fever, or be asymptomatic. The second stage is marked by additional symptoms of severe pneumonia, haematological liver and renal abnormalities. Finally the third stage shows a highly developed illness of acute respiratory distress syndrome, multiple organ malfunction and ultimately death (Tam 2002). The incidence of human infections reported in all major outbreaks of Avian Influenza to date has happened in people who have high level of interaction with poultry. It is thought these infections represent direct bird-to-human transfer of the virus, as fortunately there appears to have been negligible human-to-human spread. Nevertheless, the adaptation and mutation of these poultry viruses could lead to a new sub-type of Influenza which is capable of sustaining itself within a human population alone. This is a key concern to public health services around the world as the threat of a

highly contagious new pandemic could be potentially looming ever closer (Bengis *et al.* 2004).

## 1.2 Wildlife disease and conservation

Biodiversity can also be affected by the emergence and spread of wildlife diseases. If the pathogen in question induces a high enough mortality rate in the host, there is a risk of losing entire populations of species and endangering many others; disease induced reductions in population size may significantly increase the chance of local extinction due to demographic fluctuations. Recent outbreaks have demonstrated this, for example in Australia the chytrid fungus *Batrachochytrium dendrobatidis* (*Bd*) is thought to be behind the fall in population size and perceivable extinction of at least 14 high-elevation species of rainforest frog (Retallick *et al.* 2004). This fungus is not unique to Australia, and its effects on the amphibian community can be felt worldwide. Using information from a “last year observed” database, Marca *et al.* deduced that the frog genus *Atelopus* has undergone 67 species extinctions since 1980 (Marca *et al.* 2005), which have been generally attributed to *Bd*. *Bd* has been categorised as an amphibiotic emerging infectious agent (Daszak *et al.* 1999) and is considered to be pandemic. In 2008 the World Organisation for Animal Health reported *Bd* as a notifiable pathogen (OIE 2008).

The impact of pathogens on wildlife conservation can be damaging when endangered or threatened species are affected by an outbreak of an infectious pathogen. The Tasmanian devil population has continually declined by up to 90% from 1996 when a debilitating and aggressive facial cancer tumour was first reported known as Tasmanian Devil Facial Tumour Disease (DFTD) (McCallum *et al.* 2007). This has caused the Tasmanian Devil, which was originally listed as “low-risk” in terms of endangerment in 1996, to be officially categorised an “endangered” species as of 2008 by the International Union for Conservation of Nature (IUCN) Redlist (Hawkins *et al.* 2008). In the conservation biology literature emphasis is placed on the possible adverse consequences of wildlife disease due to a decline in genetic diversity (Epps *et al.* 2005; Schmid *et al.* 2009). The lasting effect of a decline in genetic diversity within a population is the reduction in the ability to adapt to changes, for example loss of habitat or fragmentation. A more immediate disadvantage is the incapacity to resist pathogen infection. DFTD is a prime example of how loss of genetic diversity within populations amplifies the risks posed by disease (Jones *et al.* 2004). This case highlights the

importance of surveillance and early action against emerging diseases threats; as DFTD was not quickly identified it has already spread across a large range of the species which makes eradication difficult (McCallum 2008).

Canine distemper (CDV) is a viral disease affecting the respiratory tract, gastrointestinal system, skin, reproductive tract, eyes, and nervous system. It is classified as producing clinical signs such as nasal discharge, transient fever, diarrhoea, and weight loss. CDV has been recognised as a pathogen of domestic dogs, however numerous CDV infections of wild species have been documented (Leisewitz *et al.* 2001). In 1994, the Serengeti National Park lion population were subjected to a devastating outbreak of canine distemper. This led to the over-all population reducing in size by approximately one-third (i.e. 1000 animals) (Roelke-Parker *et al.* 1996), a significant blow to the conservation and protection effort of the Serengeti National Park. However, as a consequence of intense monitoring of the lion prides, detailed observations were collected on the incidence of CDV and their movement. From this data it was deduced that it was unlikely lion movement patterns could account for the spread of the disease, which eventually led to the exploration of other potential reservoirs of the pathogen (Haydon 2008). It has been suggested that the most probable reservoir for CDV in the Serengeti lion epidemic was the domestic dogs of the local villages. Between 1991 and 1993, the seroprevalence of CDV increased in their population which preceded the 1994 outbreak in lions (Roelke-Parker *et al.* 1996). This 1994 outbreak of canine distemper illustrates quite clearly the risks posed by wildlife disease but also to the great advantages of surveillance. It would have been impossible or at the least very difficult to determine how the outbreak evolved without such rich data collection.

### 1.3 Wildlife diseases and the livestock industry

When a disease is classified as notifiable, it is required by law that it is reported to government authorities if discovered. This can lead to restrictions placed on the movement of livestock from affected premises and subsequently impact on the “disease status” of a country possibly leading to a ban on trade until the country is considered “disease free”. Therefore, there is potential for tremendous economic impacts when wildlife disease can transmit to livestock. Bovine TB (bTB), caused by *Mycobacterium bovis*, is a focal point for wildlife control because of the adverse consequences of the disease on livestock production

(Donnelly & Hone 2010) and the significant affect it has on trading as a result of EU trading standards and procedures (Caffrey 1994; DEFRA 2011). bTB is categorised as a bacterium which causes chronic incapacitating disease in cattle, humans and various wild species, including the badger (*Meles meles*) (Bengis *et al.* 2004). During the comprehensive Randomised Badger Culling Trial (RBCT), designed to establish if culling badgers reduced TB in cattle, badgers were shown to be a source of infection to cattle (DEFRA 2013). The trial results suggested that culling of badgers over a fixed area of 150km<sup>2</sup> would lead to an average of 16% reduction in bTB incidence in cattle in the local area. However, there are also negative impacts of culling which can lead to an increases in disease e.g. the perturbation effect as mentioned previously (Prentice *et al.* 2014). This situation has put a strain on the economy as infected cattle have decreased production of milk and infected carcasses will be seized at abattoirs if detected (Amanfu 2006; Firdessa *et al.* 2012). Known infected cattle herds are put on a trade lockdown as required by the EU trading standards and are prohibited from moving until bTB free status is achieved. There are also added economic problems which come with the slaughtering of infected cattle. It is estimated that approximately 28,000 cattle were slaughtered in 2012 through bTB infection (DEFRA 2012) and the burden of compensation falls on the government or else must be absorbed by the individual infection sites which could result in farms going out of business if the financial impacts of the cull are too great.

Schmallenberg is an emerging livestock virus that has been detected in parts of Europe and the UK and it is transmitted via vectors such as mosquitoes and midges to the livestock hosts (DEFRA 2012) and wild ruminants (ECDC 2012). The virus is characterised by the host showing clinical signs including a decrease in milk production, watery diarrhoea, and occasional fever (Elbers *et al.* 2012), and there have also been reports of congenital malformations in ruminants (van den Brom *et al.* 2012). DEFRA has stated that the most likely cause of Schmallenberg in the UK is due to infected midges being blown across the channel, most probably from France (Conraths *et al.* 2013). Although this is not a notifiable disease as yet, it is being closely monitored and any farmers who encounter the disease or suspicious symptoms are advised to contact their local vet. Impact assessments by the EFSA (2012) and Harris (2014) suggest that if this virus were to spread further across the UK and Europe it is likely to have serious adverse economic impacts on the farming industry (EFSA 2012; Harris *et al.* 2014) as meat and milk production could be badly hampered.

Foot and mouth disease (FMD) is an infectious virus affecting ruminants and is characterised by clinical signs including loss of appetite, sudden death of young, lameness, blisters and reduced milk yield. Although recent outbreaks have been predominantly in livestock, wild ruminants can also play a role in the introduction and spread of FMD (Condy *et al.* 1969; Ward *et al.* 2007). The disease can have severe consequences for animal health and the economics of the livestock sector as exemplified by the 2007 outbreak in the UK. During August and September 2007, FMD caused large disruptions to the farming sector and cost hundreds of millions of pounds in control efforts and slaughtered animals (Cottam *et al.* 2008). For example in Scotland there was an export ban on live animals imposed until the close of the year, the effect of this was a reduction of market prices, although the real measure of this depends on the initial state of the market before the outbreak. This represented a considerable cost to farmers of livestock and in turn also represented losses to the overall agricultural supply chain (Scottish Government 2008). The burden of a restriction on movement requires farmers and other branches of the agricultural supply chain to diverge from the usual procedure and these effects can be exacerbated if farmers keep their stock for a longer period because of lower market prices, all of which demonstrate extra cost and strain on the livestock economy.

## 1.4 Wildlife Disease Surveillance

Surveillance is the first line of defence against wildlife disease and the threats it can pose. Wildlife disease surveillance aims to limit and end outbreaks of disease before they have the ability to cause major damage to public, livestock and wildlife health (Belant & Deese 2010) . When used in an efficient and comprehensive manner, surveillance can be instrumental in controlling and overcoming disease outbreak (FAO 2011). There is an increasing recognition of the necessity of wildlife disease surveillance (Jebara 2004; Kuiken *et al.* 2011). However, there are a range of issues associated with surveillance of wildlife disease, e.g. poor knowledge of basic ecology and distribution of host species, this makes it particularly challenging even when compared to surveillance in livestock and humans.

The wildlife disease surveillance strategies currently undertaken in Europe are few and far between (Artois *et al.* 2001). The protocols for these activities are still informal and as of yet there is no structure in place to facilitate coordinated surveillance/reporting of wildlife disease between countries (Kuiken *et al.* 2011). As previously mentioned, there is an

increasing acknowledgement that greater priority should be placed on wildlife disease surveillance, driven in part by the numerous examples of zoonotic outbreaks in the human population (SARS, Swine Flu, Avian Influenza etc). Leading international wildlife organisations and influential veterinary editorials have highlighted the importance of wildlife surveillance effort. The director general of the world health organisation for animal health (OIE) asserted that “Surveillance of wildlife diseases must be considered equally as important as surveillance and control of diseases in domestic animals” as well as concluding that the surveillance of wild animal disease is essential (Vallat 2008). In the first EWDA meeting for wildlife health surveillance on 15<sup>th</sup> October 2009 in Brussels, 25 representatives presented summaries of the wildlife health surveillance in their respective countries. Based on these summaries Kuiken *et al.* (2011) showed that there are significant differences in surveillance approach across Europe. The number of surveillance schemes in action per country, the intensity of those schemes, the number of animals examined, effort in terms of the number of people employed, and the sources of funding vary greatly. The authors categorised the differences in surveillance in each country by three different levels; no general surveillance, partial general surveillance and comprehensive surveillance, of which the UK fell into the later. As a result of this, there was a consensus among the participants that wildlife health surveillance in Europe would profit from a more formal network of people actively contributing to this research area (Kuiken *et al.* 2011).

When undertaking surveillance of a given population (i.e. a group of individuals of the same species interacting within some defined area), the task becomes a lot easier if the population is closely managed (i.e. livestock), but unfortunately this is not the case when considering wildlife populations. Compared with surveillance in livestock there are many added complications to wildlife disease surveillance, for example locating the population, estimating accurately the size of the population and understanding the demography and transmission dynamics of the population including interactions with other populations and the wider community (i.e. other species). These added complications can make it difficult to obtain the samples required of a successful surveillance system (Nusser *et al.* 2008). Many wildlife surveillance strategies depend on methodologies based on protocols developed for livestock systems. However, as discussed, wildlife populations are considerably more complex and to date there has not been a detailed exploration of whether methods used in livestock are suitable for application to wildlife populations.

Wildlife disease surveillance can be characterised under two broad categories, active surveillance and passive surveillance. Passive surveillance can be generally defined as the discovery and testing of naturally occurring deceased hosts (i.e. animals that have not died for the initial purpose of surveillance). There are instances of routine collection of hunter-killed samples and collection of road-kill animals, but a primary difficulty of passive surveillance is that the strategy generally relies on members of the public identifying and delivering a case for diagnostic testing (Rhyan & Spraker 2010). The potential of passive surveillance is hard to realise in practice since disease detection is quite frequently time sensitive (i.e. sensitivity of diagnostic tests may reduce sharply with time since death) and the incentive for the general public to report a case is relatively low. There is also a considerable chance of bias in the sample when relying on passive surveillance especially if the host-pathogen dynamic features significant disease induced mortality or if behaviour of infected individuals reduces or increases the chance that deceased individuals are encountered. Active surveillance in the context of this thesis is defined as the capture and subsequent testing of individuals driven by surveillance related objectives. A primary difficulty with this type of surveillance is that it can be more costly than other options and only limited funding is available (Lancoua *et al.* 2005). In livestock systems active and passive surveillance strategies are much simpler to implement, and as highlighted previously wildlife systems are much more complex. There are numerous complications when gathering samples in the field, including dynamic aspects of population turnover, habitat effects on density and distribution in space and time, behavioural aspects affecting sampling e.g. elusive nocturnal species trap shyness of animals etc. These complications typically result from dynamic processes which are subject to stochastic fluctuations making it more difficult to design and implement randomised sampling strategies.

In summary, both passive and active surveillance have the potential to be effective tools for wildlife disease research, but they can suffer shortcomings including under-reporting, and difficulties in designing effective surveillance strategies due to the complexity of host-pathogen systems (Stallknecht 2007). There is therefore a need to address these problems systematically; in particular there have been calls for improved pan-European mechanisms including defined standard protocols and data sharing (Genovesi & Shine 2002; Kuiken *et al.* 2011; “WILDTECH Report Summary” 2014). This is a long term goal which would aim for a coordinated approach to surveillance and monitoring, to offer increased protection from disease outbreaks and incursions. For this to work, unquantifiable biases need to be minimised (e.g. human behaviour in passive systems) and/or accounted for in subsequent

analysis. For the purposes of this thesis, we investigate only active surveillance as the characterisation of biases in passive surveillance requires a focus on specific surveillance scenarios and here we wish to explore generic aspects of wildlife disease surveillance. For example, results not shown indicate the ability of passive surveillance to detect disease depends strongly on the level of disease induced mortality and the rate at which the animals decay.

The key statistics we subsequently use to characterise the performance of wildlife disease surveillance systems are reviewed here. There are several statistics that can be estimated using surveillance information which give valuable information about the population itself, disease status and surveillance efficacy. However, in this thesis there are two primary statistics of interest used to investigate surveillance performance and the effect of population demography and disease transmission on surveillance efficacy, the estimate of prevalence and the probability of detection.

#### 1.4.1 The Probability of Detection

The simplest and most widely used approach to estimating the probability of disease detection is to assume constant prevalence  $p$  and an effective infinite population size (i.e. assume sampling is with replacement/the population size is not finite). These assumptions lead naturally to a binomial formula for the probability of detecting disease from  $n$  samples.

$$a = 1 - (1 - p)^n \quad (1)$$

where  $n$  = the sample size

$a$  = the probability of detecting disease

$p$  = the prevalence

The above formula can be used to carry out a power calculation as follows. Re-arranging equation (1) in terms of  $n$ , gives an estimate of the sample size required to obtain a probability of detection  $a$  given an assumed prevalence  $p$  can be derived.

$$n = \frac{\ln(1-a)}{\ln(1-p)} \quad (2)$$

This required sample size increases rapidly as prevalence tends to 0, as demonstrated in Figure 1.1. This can lead to “over sampling” especially for smaller populations, since the underlying assumption is that the population is infinite. To counteract this effect which would lead to the repeated sampling of individuals, a number of authors (Martin *et al.* 1987; Artois *et al.* 2009a; Fosgate 2009) have considered modifying this approach based upon the hyper-geometric distribution which accounts for finite population sizes. This approach leads to the following sample size calculation

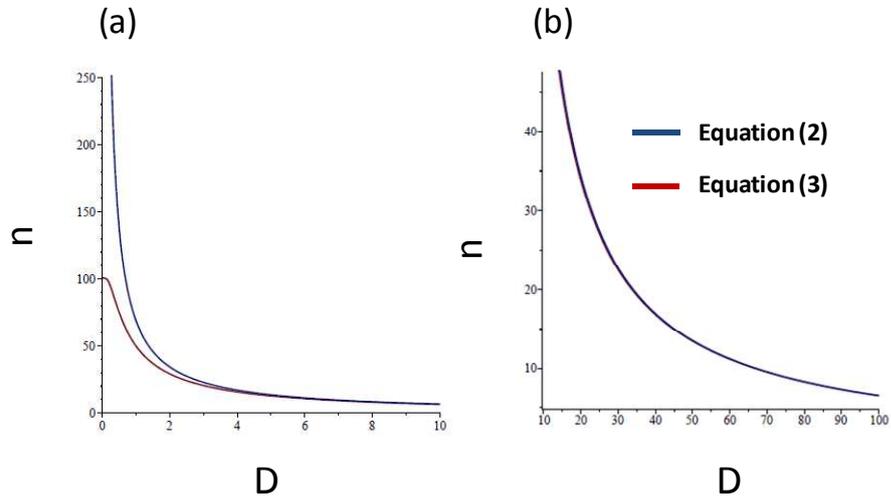
$$n = [1 - (1 - a)^{1/D}] \times [N - 0.5(D - 1)] \quad (3)$$

where  $N$  = the total population size

$D$  = the total number of diseased individuals within the population

$$p = \frac{D}{N} = \text{prevalence}$$

Figure 1.1 shows equation (3), demonstrating the effect of modifying equation (2) to account for finite population sizes. Without this modification, at low prevalence, the sample size required to detect disease presence can be greater than the population itself. As noted above this would entail sampling the same individuals more than once which is clearly inefficient and given perfect tests completely unnecessary, at least if the disease status of individuals is assumed not to change over time. This is of course consistent with the assumption of unvarying prevalence on which both equations (1)-(3) rely. The hyper-geometric correction implemented in equation (3) ensures the maximum sample size required is capped at the total population size.



**Figure 1.1: Effect of disease prevalence on the sample size required to detect disease.** Plots are shown for varying levels of diseased individuals when the probability of detection  $\alpha = 0.5$  for both equation (2) and (3). Plot 1.a shows the effect of number of diseased individuals on sample size requirement described by equations (2) and (3) for a fixed population size of 100 (i.e. maximum prevalence is 0.1). Plot 1.b shows the effect of number of diseased individuals on sample size requirement from equations (2) and (3) for a fixed population size of 1000 where again maximum prevalence is 0.1.

Equations (3) can be re-arranged to give an equation for the probability of detection for finite sized populations analogous to equation (1):

$$a = 1 - \left(1 - \frac{n}{N - 0.5(D-1)}\right)^D \quad (4)$$

Note that equation (1) is the form that will be used herein when referencing the binomial equation for the probability of detection.

#### 1.4.2 Estimating the True Prevalence: Bias and Standard Deviation

In addition to simply detecting the presence of disease, surveillance may also be called upon to accurately estimate the prevalence in a population. However, as we show in this

thesis achieving this can be quite difficult in the face of demographic fluctuations within the population. In contrast the standard approach is to ignore such fluctuations and assume constant prevalence  $p$  and population size. This leads to the conclusion that the number of infected individuals in a sample of  $n$  individuals, from a population with prevalence  $p$ , is drawn from a binomial distribution with  $Bin(n,p)$ . Which has mean  $np$  and variance  $np(1-p)$ . Therefore under these assumptions we find that the binomial estimate of prevalence is  $np/n=p$  and therefore

$$E[\text{surveillance estimate of prevalence}] = E[\text{true prevalence}] \quad (5)$$

i.e. the bias of the surveillance estimate (under the above assumptions) is equal to 0.

Given the variance in the binomial estimate of the number of infected cases in a sample of size  $n$ , the standard deviation (std dev) in the corresponding estimate of prevalence from surveillance is:

$$\frac{\sqrt{np(1-p)}}{n} = \sqrt{p(1-p)/n} \quad (6)$$

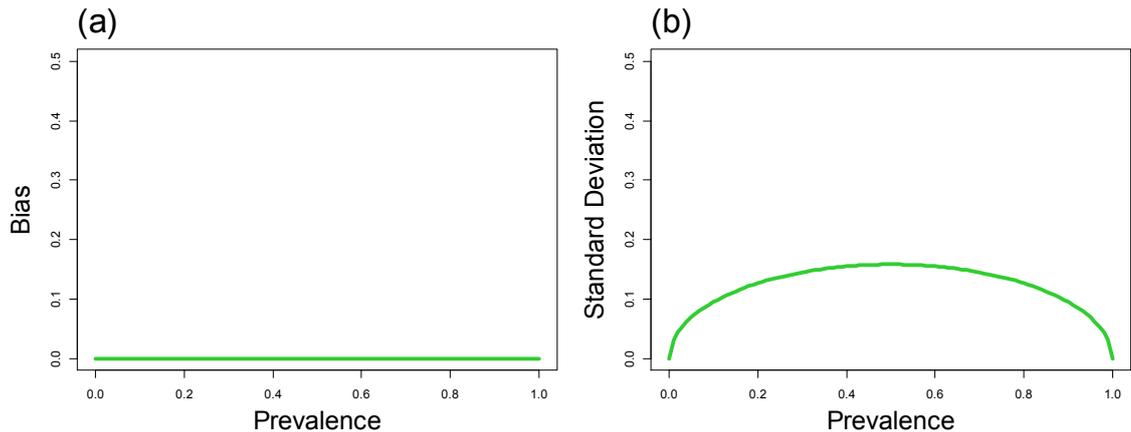
as before,  $p$  =prevalence and  $n$  = sample size.

As with the probability of detection in equation (4), there can be corrections made to equation (6) to account for the finite size of the population. Frequently in survey research, samples are taken without replacement and from a finite population of size  $N$ . In this instance, and especially when the sample size  $n$  is proportionally not small (i.e.  $n/N > 0.05$ ), a finite population correction factor (fpc) is used as a pre-factor on the right hand side of equation (6) to define both the standard error of the proportion. The finite population correction factor is expressed as:

$$fpc = \sqrt{N - n / N - 1} \quad (7)$$

Figure 2 show the bias and standard deviation using the binomial theory based on the assumptions of constant population size and prevalence. As stated above, the binomial

equation does not predict any bias in the surveillance estimate of prevalence as the predicted prevalence is equal to the true prevalence,  $p$ . However, the error in this estimate is non-zero and varies with true prevalence, as can be seen in Figure 2.b. Note that the correction factor in equation (7) is less than 1 for any  $N > 2$  and sample size  $n > 1$ , and therefore the error in the estimated prevalence shown in Figure 2.b will be reduced when accounting for the finite size of populations.



**Figure 1.2: Using binomial theory to estimate the expected bias and std dev in the surveillance estimate of prevalence.** Plots are shown for sample size  $n = 10$  and varying levels of prevalence in the population for equation (5) and (6). Plot 2.a shows the predicted bias in the prevalence estimate from surveillance. Plot 2.b shows the predicted std dev in the prevalence estimate from surveillance.

### 1.4.3 Improving wildlife disease surveillance

In recent years, in line with the heightened interest in wildlife disease discussed above, a number of authors have identified a need for both improved implementation and methodological developments to enhance the design and evaluation of wildlife disease surveillance (Stallknecht 2007; Hadorn & Stärk 2008; Artois *et al.* 2009a; Ryser-Degiorgis 2013). In order to effectively do this, it is important to consider the ecology of the population under surveillance as this will have an impact on the results obtained (Béneult *et al.* 2014). A strategy that worked for one type of natural population may not necessarily work as well in another. Understanding how the dynamics of the host-pathogen interaction affect the efficacy of surveillance is key, and potentially the most important step towards improvement of surveillance systems. However such effects have yet to be systematically considered in the literature. There have, however, been attempts to improve wildlife

disease surveillance design by incorporating weighting schemes based on habitat suitability of the observed population (Nusser *et al.* 2008; Walsh & Miller 2010). However, there are also other factors that should be taken into consideration if required, for example dispersal of populations, population fluctuations, disease stability, seasonality and environmental change. In this thesis we use mathematical modelling to explore how key ecological processes that govern wildlife populations, in particular demographic fluctuations, stochastic disease dynamics and spatial heterogeneity, impact on surveillance. Simulation modelling could make a positive contribution to this area, as testing out scenarios in the field is either very monetarily costly and/or time consuming or altogether unfeasible. By running simulations, what is expected from different surveillance strategies in a wide range of host-pathogen systems can be explored. Using such studies, it would be possible to better understand the results obtained from a given surveillance strategy.

## 1.5 Modelling Wildlife Populations, Disease and Surveillance

Mathematical modelling is a tool, or more correctly a set of tools, used to represent different mechanisms in the natural world, and in particular enables prediction of system level impacts that results from interactions between multiple mechanisms (Lucio-Arias & Scharnhorst 2012). Such techniques have a long established role in mathematical biology. Translating descriptions of biological processes, such as host-pathogen interaction, into mathematical language is beneficial for many reasons, for example, the precise mathematical language aids in formulating ideas and recognising underlying assumptions as well as utilising mathematical techniques to manipulate equations in order to gain insights into the behaviour of the modelled system. Unfortunately non-linearities and population heterogeneities, which are of key importance in many biological systems, make such formal analysis of associated mathematical models difficult and typically intractable. However a key advantage of this approach to biological research is the use of computers in performing calculations and running simulations. This is a considerable time saver and also enables the exploration of many scenarios that would be unfeasible to study in the field. There are, however, compromises to be made with this approach. As with most natural interacting systems, the dynamics of a host-pathogen interaction are potentially extremely complicated and it is important to identify the key elements to include within the model since encompassing every aspect is typically unfeasible and such a comprehensive approach is unhelpful in terms of generating insights/understanding into how the studied system works.

“In principle models should be developed from the simple to the complex” (Murray 2002) and this ethos has been adopted throughout this thesis. With every added complexity introduced to the model description, the longer and more intensive are the methods required to handle the equations. Insights obtained from relatively simple model structures are more generic and can also aid in the understanding of more complex models.

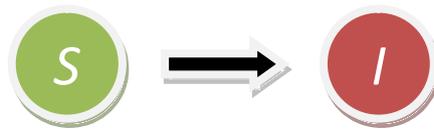
### 1.5.1 Temporal Modelling of Wildlife Populations with Disease

When simulating natural populations it is essential to account for key demographic processes in the model, e.g. births, deaths and immigration. An example is shown in Figure 1.3 which illustrates a population subject only to birth governed by a logistic growth rate. In this thesis, a primary interest is that of pathogen transmission, and we model both primary and secondary transmission. Primary transmission occurs when a susceptible from the modelled population becomes infected by routes including, infected water sources, contaminated food, and transmission from an individual outside the modelled population. Secondary transmission occurs when a susceptible individual from the modelled population becomes infected through contact with an infected individual within the modelled population. Vertical and pseudo vertical transmission represent the infection of offspring and young individuals by parents. However, unless such transmission is infallible vertical and pseudo vertical mechanisms are not capable of sustaining infection and in this thesis we do not consider such routes. Nonetheless such mechanisms could easily be incorporated as part of the models presented here. As well as these elements describing demographic processes and disease dynamics it will be important to model key aspects of the process surveillance in order to assess how it performs in differing circumstances.

Compartmental models have been used frequently to describe host-pathogen systems in wildlife disease research (Renshaw 1991; Murray 2002) as they aim to reduce the complexity of host-pathogen dynamics into a manageable number of disease status “compartments”. There are many different examples of such compartmental models; some of the more well-known include SI, SIS, SIR, SIRS and SEIR. In these compartmental models  $S$  represents a susceptible state,  $I$  an infected state,  $E$  a latent state (i.e. infected but not yet infective), and  $R$  represents a recovered state (recovered from disease and no longer infectious or able to be infected). There are two broad categories of mathematical modelling which can utilise such compartmental structure, deterministic and stochastic. In both cases the model state space typically represents the number of individuals in each category. Deterministic models ignore the random fluctuations that can be observed in real

systems, solving standard differential equations simultaneously to update the model at discrete time steps. The deterministic nature of these models means that, starting from a given input they will always return the same output. Continuous time stochastic models incorporate random variation by utilising probabilistic equations to determine a series of events which update the model at randomly generated time intervals and potentially (for multiple event types) in a manner that is also stochastic in nature. Such stochastic processes naturally lend themselves towards not only population level representations (e.g. where numbers of susceptible and infected individuals are tracked) but also to individual based models where the disease status over time of each member of the population is represented. There are pros and cons for both deterministic and stochastic approaches and the differences between them are now demonstrated with a simple birth-death process SI example.

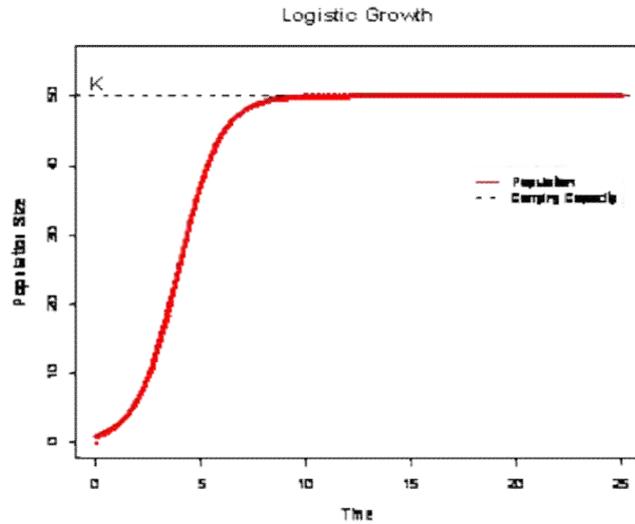
### A Simple Compartmental Model Worked Example



Consider the components of the state space  $S(t)$  and  $I(t)$  which represent the number of susceptible and infected individuals respectively at time  $t$ . Birth is represented by logistic growth defined by Verhulst (1845), and since this is a feature embedded throughout the models used in this thesis a more detailed explanation is required. Deterministic logistic growth is defined by the following equation:

$$\frac{dN}{dt} = rN \left( 1 - \frac{N}{K} \right)$$

where  $N$  is the total population ( $S + I$ ),  $r$  the intrinsic growth rate and  $K$  the carrying capacity. Figure 3 demonstrates a population experiencing deterministic logistic growth over time until reaching the carrying capacity,  $K$ . All individuals born via logistic growth are susceptible, which implies no vertical or pseudo vertical transmission.



**Figure 1.3: An example of a logistic growth curve** A Plot is shown to demonstrate how the population increases over time using the logistic growth equation. Despite an initial exponential increase growth is ultimately limited by the carrying capacity,  $K$ .

If there is also a per capita death rate,  $d$ , and a density dependent secondary transmission with contact rate  $\beta$ . The deterministic approach then describes an SI model with births and deaths as:

$$\frac{dS}{dt} = rN \left(1 - \frac{N}{K}\right) - \beta SI - dS \quad (8)$$

$$\frac{dI}{dt} = \beta SI - dI \quad (9)$$

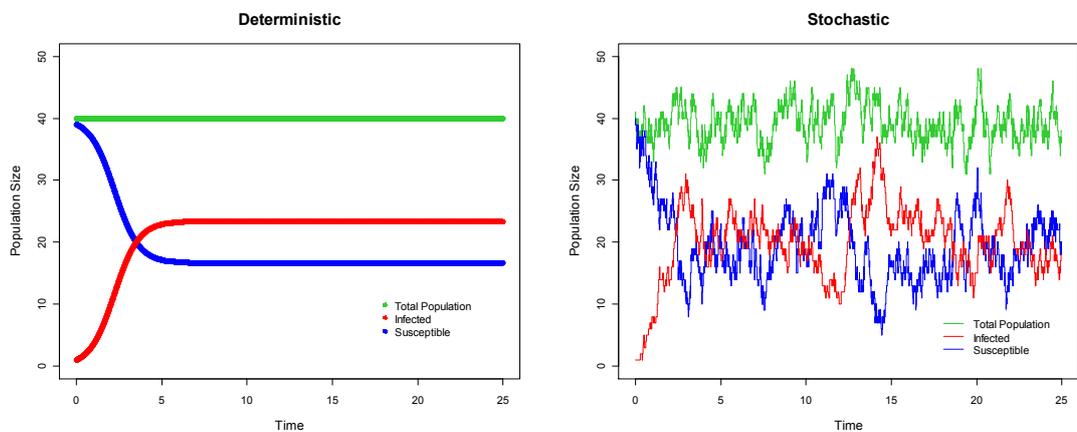
Given that  $N = S + I$

The biological mechanisms underlying this model are summarised in Table 1 below, which shows the rate at which each event occurs as well as the associated change in the population. In the next section we describe how the basic model description shown in Table 1.1 can be implemented as a stochastic model.

Event	Rate	Effect
Birth	$rN \left(1 - \frac{N}{K}\right)$	$S = S + 1$
Death of $S$	$dS$	$S = S - 1$
Death of $I$	$dI$	$I = I - 1$
Infection	$\beta SI$	$S = S - 1$ $I = I + 1$

**Table 1.1: SI model with rate and effects for each event type.** A table is shown which contains the rates of all events which can occur at each time step with the associated effect on the state space. This model can be implemented as a discrete state space stochastic process or with a continuous state space, as a deterministic model using ordinary differential equations or an analogous system of stochastic differential equations (see text for details).

Figure 1.4 demonstrates how  $S$ ,  $I$  and  $N$  population sizes develop over time using both the deterministic and discrete state space stochastic approach (described below).



**Figure 1.4: Deterministic and Stochastic Time Trajectories.** Shown here is the total population (green), the susceptible population (blue) and the infected population (red) over time for both a deterministic and discrete state space stochastic Gillespie implementation of an SI model with births and deaths. The parameter values used are  $d = 1$ ,  $\beta = 0.06$ ,  $r = 5$ ,  $K = 50$ , and starting with initial conditions of  $S = 39$ ,  $I = 1$ .

As Figure 1.4 shows, the outcome of a stochastic and deterministic simulation, although following roughly the same pattern, can be quite different. Deterministic modelling has its merits, but in the interests of understanding the impacts of variability on the dynamics of host-pathogen systems, a stochastic approach has been used in the models developed in this thesis.

The mechanisms summarised in Table 1.1 can be used to formulate an integer valued discrete state space continuous time Markov process. This then provides a natural stochastic description of the assumptions implicit in equations 8 and 9 (made explicit in Table 1.1). Under this Markov process, the probability that a given event occurs during a short time interval  $(t, t+\Delta t)$  is given by its rate multiplied by  $\Delta t$ . The stochastic model can, and in some of the models of this thesis will, be simulated using the Gillespie algorithm defined by Gillespie (1979). To run Gillespie's algorithm, starting from time  $t$  when the system is in state  $x(t)$ , the time to the next event,  $t_x$ , is randomly chosen from an exponential distribution with parameter,  $M_{total}(x(t))$ , which is the total rate (i.e. the sum of the rates of all possible events) evaluated at time  $t$ . The event, then occurs at time  $t + \tau$  and is chosen from the set of possible events with probabilities given by the rate of each possible event divided by  $M_{total}(x(t))$ . The derivation of the exponential waiting time distribution for a Markov process which is the basis of the Gillespie algorithm is as follows. Let –

$$G(t | \underline{x}(0)) = \text{probability that no event has occurred up to time } t,$$

when starting at  $\underline{x}(0)$  at time  $t = 0$ .

Then considering the change in  $G(t | \underline{x}(0))$  over a small time  $\delta t$  gives,

$$G(t + \delta t | \underline{x}(0)) = G(t | \underline{x}(0)) (1 - M_{total}(x(0))\delta t)$$

Where to first order in  $\delta t$ ,  $1 - M_{total}(x(0))\delta t$  = probability that no event has occurred in a small time interval  $(t, t + \delta t)$  i.e. 1- the probability that one of the set of possible events did occur.

Rearranging this equation and taking the limit as  $\delta t \rightarrow 0$  we find

$$\frac{\partial G(t | x(0))}{\partial t} \equiv \lim_{\delta t \rightarrow 0} \frac{G(t + \delta t | \underline{x}(0)) - G(t | \underline{x}(0))}{\delta t} = -G(t | \underline{x}(0)) M_{total}(x(0))$$

$$\Rightarrow \frac{1}{G(t | x)} \frac{\partial G(t | x(0))}{\partial t} = -M_{total}(x(0))$$

$$\Rightarrow \int_0^t \frac{1}{G} dG = - \int_0^t M_{total} dt$$

$$\Rightarrow \ln G(t | \underline{x}(0)) - \ln G(0 | \underline{x}(0)) = -M_{total}(x(0))t$$

$$\Rightarrow G(t | \underline{x}(0)) = e^{-M_{total}(x(0))t}$$

This last line follows on noting that, by definition, nothing has happened at time  $t=0$  and so  $G(0 | \underline{x}(0)) = 1$ . Note that this equation forms the basis of the Gillespie algorithm since it shows that the time to the next event has an exponential distribution.

For the model described in Table 1

$$M_{total} = rN \left(1 - \frac{N}{K}\right) + \beta SI + dS + dI$$

Thus starting at time  $t$  and the time is advanced to  $t + \tau$  with the inter event waiting time drawn from the exponential distribution. i.e.,  $\tau \sim \exp(M_{total}(x(t)))$  where  $x(t) = \{S(t), I(t)\}$  represents the state space at time  $t$ . After the time to the next event is calculated, the event which has occurred is chosen randomly by generating a number from the uniform distribution,  $y \sim U(0, M_{total})$ . The next event is:

Birth	$if\ y < rN\left(1 - \frac{N}{K}\right)$
Death of $S$	$if\ y < rN\left(1 - \frac{N}{K}\right) + dS$
Death of $I$	$if\ y < rN\left(1 - \frac{N}{K}\right) + dS + dI$
Infection	$if\ y < rN\left(1 - \frac{N}{K}\right) + dS + dI + \beta SI$

Note alternative orderings of event types (and associated rates) are allowed but the order used does not affect the statistical properties of the model. The state space is updated accordingly, the rates are recalculated and the above process is repeated until some maximum time,  $T_{max}$ , or an alternative stopping criteria, is reached. An alternative stochastic approach which ignores the discrete nature of populations (and hence is closer in spirit to the deterministic model) is that of stochastic differential equations (SDE's). The Gillespie implementation, e.g. of the SI model described above, is a continuous time discrete state-space Markov process in which the number of infected individuals ( $I$ ) and total individuals ( $N = S + I$ ) are represented as integer variables.

Table 1.2 shows the expectation and variance of the updates that would be obtained during a small time interval from the Gillespie algorithm implementation of the model based on the description of events shown in Table 1.1. (i.e. from the discrete state space SI model with births and deaths). Comparison with Table 1.2 enables both drift e.g.  $f_{N,B}(X(t))$  and diffusion e.g.  $g_{N,B}(X(t))$  functions to be identified

However the SDE approach makes use of continuous variables to represent the state space. Using the simple compartmental example above, we can represent the change in the system state variables during an infinitesimally small time interval  $dt$  as the following set of stochastic differential equations:

$$\begin{aligned}
dN(t) = & \left( f_{N,B}(X(t)) + f_{N,DS}(X(t)) + f_{N,DI}(X(t)) + f_{N,2ry}(X(t)) \right) dt \\
& + g_{N,B}(X(t))dB_B(t) + g_{N,DS}(X(t))dB_{DS}(t) \\
& + g_{N,DI}(X(t))dB_{DI}(t)(X(t)) + g_{N,2ry}(X(t))dB_{2ry}(t)
\end{aligned}$$

$$\begin{aligned}
dI(t) = & \left( f_{I,B}(X(t)) + f_{I,DS}(X(t)) + f_{I,DI}(X(t)) + f_{I,2ry}(X(t)) \right) dt \\
& + g_{I,B}(X(t))dB_B(t) + g_{I,DS}(X(t))dB_{DS}(t) \\
& + g_{I,DI}(X(t))dB_{DI}(t) + g_{I,2ry}(X(t))dB_{2ry}(t)
\end{aligned}$$

The reader will notice that here we have chosen to represent the dynamics in terms of the numbers of infectives and the total size of the population rather than using the number of susceptibles and another variable. Therefore the state space at time  $t$  is now represented by the vector  $x(t) = \{N(t), I(t)\}$ . The quantities  $B_B(t)$ ,  $B_{DS}(t)$ ,  $B_{DI}(t)$ ,  $B_{2ry}(t)$  are independent Brownian motions corresponding to each of the four event types. For small but finite  $dt$  the quantities  $dB_B(t)$ ,  $dB_{DS}(t)$ ,  $dB_{DI}(t)$ ,  $dB_{2ry}(t)$  can be interpreted as independent draws, from a zero mean Gaussian with variance  $dt$ , for each event type and each time point  $0, dt, 2dt, \dots, T \in (0, T)$ . Thus e.g.  $E[dB_B(t)] = 0$ ,  $E[ dB_B(t) dB_B(t) ] = dt$  and  $E[ dB_B(t) dB_{DS}(t) ] = 0$ . The so called drift term  $f_{N,B}(X(t))$  represents the expected change in population size  $N$  associated with the birth event conditional on the system being in state  $X(t)$ , whereas the diffusion term  $g_{N,B}(X(t))$  represents the variance in this update. There are analogous drift and diffusion quantities corresponding to each state variable for each event type. These are detailed in Table 1.3.

E-type	Event	$E[\delta N   X(t)]$	$E[\delta I   X(t)]$	$\text{Var}[\delta N   X(t)]$	$\text{Var}[I   X(t)]$	$\text{Cov}[\delta N, \delta I   X(t)]$
B	Birth	$rN \left(1 - \frac{N}{k}\right) \delta t$	0	$rN \left(1 - \frac{N}{k}\right) \delta t$	0	0
DS	Death of S	$-\mu S \delta t$	0	$\mu S \delta t$	0	0
DI	Death of I	$-\mu I \delta t$	$-\mu I \delta t$	$\mu I \delta t$	$\mu I \delta t$	$\mu I \delta t$
2ry	Secondary Trans-mission	0	$\beta I S \delta t$	0	$\beta_0 S \delta t$	0

**Table 1.2: Expectations and variance-covariance rates.** Expectations and variance-covariances in changes (during the time interval  $t$  to  $t+\delta t$ ) to the state space  $\{I(t), N(t)\}$  associated with each event type in the discrete state-space model described above for the SDE implementation. All such quantities are shown to first order in  $\delta t$ .

E - type	Event	$E[\delta N X(t)]$	$E[\delta I X(t)]$	$\text{Var}[\delta N X(t)]$	$\text{Var}[I X(t)]$
B	$f_{N,B}(X(t))dt$	$f_{I,B}(X(t))dt$	$g_{N,B}(X(t))^2 dt$	$g_{I,B}(X(t))^2 dt$	0
DS	$f_{N,DS}(X(t))dt$	$f_{I,DS}(X(t))dt$	$g_{N,DS}(X(t))^2 dt$	$g_{I,DS}(X(t))^2 dt$	0
DI	$f_{N,DI}(X(t))dt$	$f_{N,DI}(X(t))dt$	$g_{N,DI}(X(t))^2 dt$	$g_{I,DI}(X(t))^2 dt$	$g_{N,DI}(X(t))g_{I,DI}(X(t))dt$
2ry	$f_{N,2ry}(X(t))dt$	$f_{I,2ry}(X(t))dt$	$g_{N,2ry}(X(t))^2 dt$	$g_{I,2ry}(X(t))^2 dt$	0

**Table 1.3: Expectation and variance-covariance.** Expectation and variance-covariances in changes (during the time interval  $t$  to  $t+dt$ ) to the state space  $\{I(t), N(t)\}$  associated with each event type in the SDE model as described above. All such quantities are shown to first order in  $dt$ .

The SDE implementation is the diffusion limit of the Gillespie implementation if it is constructed in such a way that the first and second order moments of the stochastic updates in the differential equations correspond with those of the Gillespie implementation. This ensures that the results are consistent between the two implementations. It is easy to see that this consistency is achieved if the values of the drift and diffusion terms in the SDE model are chosen by comparing Tables 1.2 and 1.3. This then suggests, for example that  $f_{N,B}(X(t)) = rN(1 - N/K)$  and  $g_{N,B}(X(t)) = rN(1 - N/K)$  with assignments corresponding to other combinations of event types and state space variables made in an analogous manner (i.e. matching up the entries describing expectations and variances in updates shown in Tables 1.2 and 1.3). For events which change both state space variables the above formulation also ensures that associated updates also have the correct covariance, between changes in  $N$  and  $I$ .

There are pros and cons to both the SDE and Gillespie implementations, for example the Gillespie algorithm is computationally more intensive whereas using SDEs is faster and therefore facilitates more accurate estimation of model statistics (i.e. a greater number of realisations can be run). However, the discrete nature of the state-space under the Gillespie algorithm represents a more natural description of the population and the processes that affect it. In particular, it provides a more accurate representation of population dynamics for small populations.

### 1.5.2 Spatial Temporal Modelling of Wildlife Populations, Disease and Surveillance

As well as describing the temporal ecological interactions in wildlife systems, mathematical models can also include spatial temporal dynamics. Research has shown how extrinsic spatial heterogeneity (i.e. habitat and land use), has an impact on disease prevalence and persistence (Hagenaars *et al.* 2004). This finding is important in terms of disease surveillance and this is taken into consideration in practice by targeting known habitats of wildlife species (Nusser *et al.* 2008). However, as far as we are aware, there has been no research or practical application which has addressed how intrinsic spatial heterogeneity (i.e. as generated by demographic fluctuations and disease dynamics) will affect surveillance and the efficacy of surveillance strategies.

The implementation of spatial stochastic models used in this thesis (see Chapter 4) builds on the basis of a non-spatial model, incorporating dynamics and time increments as described in section 1.5.1. within a spatial meta-population. Here each node or “patch” in the defined space updates through time via birth, death and immigration events etc. Every patch is connected in the spatial area by a distance kernel describing the spread of disease between patches and it is this mechanism which controls the spread of disease from patch to patch. The distance kernel decays with distance and thus limits the extent to which each patch can transmit disease. The closer the patches are, the more likely they are to pass disease to one another. Because every process added to the model increases complexity, the simulations become ever more computationally expensive. There are many uses for both spatial and non-spatial modelling approaches (Tilman & Kareiva 1997), small scale spatial heterogeneity is less significant (i.e. a single population) and it is these instances when non-spatial methods may be more appropriate. However, when dealing with large scale meta-populations, as can be seen in this thesis, spatial heterogeneity is an important factor to include.

## 1.6 The Thesis

### 1.6.1 Aims

The overall aim for this thesis is to investigate how attributes of wildlife populations affect surveillance efficacy. We focus primarily on the statistical calculations of prevalence estimation and the probability of detecting disease. This is implemented in a generic exploration and then subsequently with more specific examples.

### 1.6.2 Thesis structure

#### **Chapter 2**

Chapter 2 uses a non-spatial stochastic simulation model to implement a systematic exploration of the effects of pathogen transmission and host population dynamics on the efficacy of disease surveillance systems. Our results suggest that for the vast majority of disease systems this leads to over confidence in terms of both the power to detect disease and the bias and precision of prevalence estimates obtained. Accounting for such ecological effects will permit improvements to surveillance systems and better protection against emerging disease threats.

#### **Chapter 3**

Chapter 3 utilises the results in Chapter 2 and applies these findings to two worked examples of disease systems in the wild: badgers and tuberculosis; and rabbits and paratuberculosis. We show that similar effects to those characterised in Chapter 2 can be seen in these disease systems and we explore other sources of complexity and bias which have the potential to affect surveillance efficacy. This demonstrates the potential of the non-spatial stochastic model to be used to quantify effects in real systems and also shows its potential as a tool to explore the potential impacts of known or putative sources of bias, illustrating the power of our approach to inform surveillance.

#### **Chapter 4**

Chapter 4 extends the non-spatial model used in Chapter 2 and 3 to explore spatial aspects of wildlife disease and wildlife disease surveillance and their subsequent effects on surveillance efficacy. This chapter focuses on disease incursion events, representing emerging or re-emerging disease threats, and in particular the amount and extent of spatial

spread of disease in the system at the point of first detection by the surveillance system. Different spatial surveillance designs are considered and compared to give a better understanding of the key mechanisms driving surveillance performance in spatial settings.

## **Chapter 5**

Chapter 5 is a general discussion which brings the results from the preceding chapters into the wider research context.

# Chapter 2

## The Ecology of Wildlife Disease Surveillance

This Chapter was originally written in the style of a paper with the intention of submitting to Ecology Letters. Submitted September 2014, First Author: Laura Walton.

## 2.1 Abstract

We present the first systematic exploration of the effects of stochasticity in pathogen transmission and host population dynamics on the efficacy of wildlife disease surveillance systems. The design of wildlife disease surveillance currently ignores fluctuations in these processes. Our results suggest that for many wildlife disease systems this leads to bias in surveillance estimates of prevalence and over confidence in assessments of both the precision of prevalence estimates obtained and the power to detect disease. Neglecting such effects thus leads to poorly designed surveillance and ultimately to incorrect assessments of the risks posed by disease in wildlife. Understanding such ecological effects will enable improvements to wildlife disease surveillance systems and better protection against endemic, emerging and re-emerging disease threats. Our results suggest a need for a wider exploration of the impacts of ecology on wildlife disease surveillance.

## 2.2 Introduction

Surveillance is the first line of defence against disease, whether to monitor endemic cycles of infection (Ryser-Degiorgis 2013) or detecting incursions of emerging or re-emerging diseases (Daszak *et al.* 2000; Kruse *et al.* 2004; Lipkin 2013). Identification and quantification of disease presence and prevalence is the starting point for developing disease control strategies as well as monitoring their efficacy (OIE 2013). Knowledge of disease in wildlife is of considerable importance for managing risks to humans (Daszak *et al.* 2000; Jones *et al.* 2008) and livestock (Frölich *et al.* 2002; Gortázar *et al.* 2007), as well as for the conservation of wildlife species themselves (Cunningham 1996; Daszak *et al.* 2000; Evenson 2008).

Recent public health concerns e.g. Highly Pathogenic Avian Influenza (Artois *et al.* 2009b), Alveolar Echinococcosis (Eckert & Deplazes 2004) and West Nile Virus (Valiakos *et al.* 2014), have heightened interest in wildlife disease surveillance (Vallat 2008) and led to a growing recognition that current approaches need to be improved (Mörner *et al.* 2002). For example, there is no agreed wildlife disease surveillance protocol shared between the countries in the European Union (Kuiken *et al.* 2011). Furthermore several authors have argued that improvements are needed to the structure, understanding and evaluation of wildlife disease surveillance (Bengis *et al.* 2004; Gortázar *et al.* 2007).

Much current practice for wildlife disease surveillance (Artois *et al.* 2009a) is based on ideas developed for surveillance in livestock including calculation of sample sizes needed for accurate prevalence estimation (Grimes & Schulz 1996; Fosgate 2005) and detection of disease within a population (Dohoo *et al.* 2005). A common feature of these methods is that they ignore fluctuations in host populations and disease prevalence. These assumptions lead naturally to sample size calculations (for both disease detection and prevalence estimation) and other analyses, based on the binomial distribution and associated corrections for finite sized populations such as the hyper-geometric distribution (Artois *et al.* 2009a; Awais *et al.* 2009). Fosgate (2009) reviews current approaches to sample size calculations in livestock systems and emphasises the importance of basing analyses on realistic assumptions about the system under surveillance.

However, although constant population size and prevalence may often be reasonable assumptions for the analysis of livestock systems, they are considerably less tenable in wildlife disease systems that are typically subject to much greater fluctuations in host population density and disease prevalence. For example, practicalities and changes in population density make it much harder to obtain a random sample of hosts of the desired sample size in wildlife disease surveillance programmes (Nusser *et al.* 2008) compared with livestock systems. Furthermore the importance of temporal (Renshaw 1991; Wilson & Hassell 1997), spatial (Huffaker 1958; Lloyd & May 1996; Tilman & Kareiva 1997) and other forms of heterogeneity (Read & Keeling 2003; Vicente *et al.* 2007; Davidson *et al.* 2008) in population ecology and in particular their role in the dynamics and persistence of infectious disease has long been recognised (Anderson 1991; Smith *et al.* 2005). However, such effects have yet to be systematically accounted for in the design of surveillance programmes for wildlife disease systems, or in the analysis of the data obtained from them. Although there have been some attempts to account for spatial heterogeneities in the design of wildlife disease surveillance by incorporating weighting schemes based on habitat suitability of the observed population (Nusser *et al.* 2008; Walsh & Miller 2010), we are not aware of any attempts to account for temporal fluctuations in prevalence or host population size. Here we address this gap by assessing the impact of stochastic fluctuations in host demography and disease dynamics on the performance of surveillance in a non-spatial context.

We demonstrate analytically that correlations in fluctuations of prevalence and population density bias prevalence estimates obtained from surveillance. Simulations, using logistic models of population growth and susceptible-infected disease dynamics, support this

finding and further show that variation in prevalence estimates can be considerably higher than would be apparent from standard calculations based on constant population size and prevalence. We also explore the impact of fluctuations in population density and prevalence on the ability of surveillance to detect the presence of disease. An approximate argument indicates that, in comparison with the detection rate obtained by assuming constant prevalence, the true probability of disease detection is reduced by fluctuations, and this is confirmed by subsequent simulation. The potential range of possible detection rates is assessed by simulating a spectrum of host-pathogen systems at two sampling levels to demonstrate the potential range of performance that could be expected when surveillance is deployed in the absence of knowledge of the underlying wildlife disease system.

## 2.3 Methods

### 2.3.1 Stochastic Model Description

The model represents a host population subject to demographic fluctuations (births, deaths and immigration) and the transmission of a single pathogen. At each point in time  $t$ , the state-space represents the total population size  $N(t)$ , with  $I(t)$  of these infected and  $S(t) = N(t) - I(t)$  susceptible. In addition the prevalence is then given by  $p(t) = I(t)/N(t)$ . The birth rate of individuals is logistic,  $rN(1 - N/k)$ , with intrinsic growth rate  $r$  and carrying capacity  $k$  reflecting the assumptions that population growth is resource limited. Individuals have a per capita death rate  $\mu$  and immigration occurs at a constant rate  $\nu$ .

A proportion  $\gamma$  of immigrants are infected, but otherwise all individuals enter the population (through birth or immigration) as susceptible since we assume vertical and pseudo-vertical transmission are negligible. Susceptible individuals become infected through primary transmission (contact with infectious environmental sources including individuals outside the modelled population) and secondary transmission (contact with already infected individuals from within the population). Primary transmission occurs at rate  $\beta_0 S(t)$  while secondary transmission occurs at rate  $\beta S(t)I(t)$ .

Disease surveillance is incorporated into the model in the form of active capture, testing and release at per capita rate  $\alpha$  for both susceptible and infected individuals. All surveillance testing is undertaken assuming perfect tests, which means that our measures

of the performance of surveillance reflect a best case scenario. A summary of this conceptual model is given in Table 2.1 which shows all demographic, epidemiological and surveillance events with their corresponding rate and effect on the state-space.

### 2.3.2 Statistics generated from the model

Since we allow immigration of susceptible and infected individuals neither the population nor the disease will become extinct and we therefore assume that long term averages are equivalent to ensemble expectations (typically approximated by averages over many realisations of the process). Each simulation is run for a period of time to allow the population to reach equilibrium before long run averages are calculated. For example, the expected mean  $E[M]$  and variance  $\text{Var}[M]$  of the population size are recorded along with the expected mean  $E[p]$  and variance  $\text{Var}[p]$  of disease prevalence. Similarly other statistics such as the covariance between the prevalence and population size are calculated as required.

During a so called *surveillance bout* individuals are captured at per capita rate  $\alpha$ , and both the total number and the number of infected individuals caught are recorded. Note this could be easily extended to account for imperfect disease diagnostics by recording the number testing positive but here we assume perfect tests. When the surveillance bout ends, either because a target number of individuals has been caught or because an upper time limit has been reached, the sample prevalence is recorded. In addition, if at least one infected individual was caught we note that disease was detected. Therefore over repeated surveillance bouts it is straightforward to estimate the probability of detection  $PD$  (the proportion of bouts where disease is detected) and the mean  $E[\hat{p}_{surv}]$  and variance  $\text{Var}[\hat{p}_{surv}]$  of the prevalence estimates obtained from active surveillance.

### 2.3.3 Model Implementation

The model is implemented as a continuous time continuous state space Markov process, based on a set of coupled Stochastic Differential Equations, SDEs (see e.g. Mao 2007) and simulated using the Euler-Maruyama algorithm (see e.g. Higham 2001). For details see section 1.1 in Appendix 1. The model is also implemented as a continuous-time discrete-state space Markov process (also described in section 1 in Appendix 1), which is simulated using Gillespie's exact algorithm (Gillespie 1976). The SDE implementation has been constructed so that it is the diffusion limit of the Gillespie implementation. To achieve this, the first and second order moments of the stochastic updates in the differential equations

are chosen to correspond with those of the Gillespie implementation, ensuring that the results are consistent between the two implementations. The Gillespie algorithm is computationally more intensive whereas using SDEs is faster and therefore facilitates more accurate estimation of model statistics (i.e. a greater number of realisations can be run) and more extensive exploration of parameter space. However, the discrete nature of the state-space under the Gillespie algorithm is a more direct implementation of the model described in Table 2.1 and provides a more accurate representation of population dynamics for small populations.

<b>Event</b>	<b>Rate</b>	<b>Effect</b>
<b>Birth</b>	$rN(1 - N/k)$	$S \rightarrow S + 1$
<b>Death of Susceptible</b>	$\mu S$	$S \rightarrow S - 1$
<b>Death of Infected</b>	$\mu I$	$I \rightarrow I - 1$
<b>Susceptible Immigration</b>	$(1 - \gamma) \nu$	$S \rightarrow S + 1$
<b>Infected Immigration</b>	$\gamma \nu$	$I \rightarrow I + 1$
<b>Primary Transmission</b>	$\beta_0 S$	$S \rightarrow S - 1$ $I \rightarrow I + 1$
<b>Secondary Transmission</b>	$\beta IS$	$S \rightarrow S - 1$ $I \rightarrow I + 1$
<b>Susceptible Capture and Release</b>	$\alpha S$	$S \rightarrow S$
<b>Infected Capture and Release</b>	$\alpha I$	$I \rightarrow I$

**Table 2.1: Model structure.** Event, Rate and Effect on the State Space of the model. Conceptually the effect of each event affects an individual and this is reflected in the discrete nature of the corresponding changes in the state space. However, given this underlying conception of the model there are a number of different implementations which can be considered including via the Gillespie algorithm and stochastic differential equations (see text for details).

## 2.4 Results

### 2.4.1 Estimating Prevalence

In order to develop an understanding of the properties of wildlife disease surveillance using the above model we now develop expressions describing prevalence estimates obtained by continuous surveillance i.e. continuously deployed effort resulting in per capita capture rate  $\alpha$ .

Consider the interval  $[0, T]$  during which the population history is  $\mathcal{H}[0, T] = \{(N(t), p(t)) : t \in [0, T]\}$  where  $N(t)$  and  $p(t)$  represent the population size and disease prevalence at time  $t \in [0, T]$  respectively (see above). Let  $n_T$  represent the total number, and  $i_T$  the number of infected individuals sampled during this time interval. Conditional on the history  $\mathcal{H}[0, T]$  the expectations of these quantities are

$$E[n_T | \mathcal{H}[0, T]] = \int_0^T \alpha N(t) dt \quad \text{and} \quad E[i_T | \mathcal{H}[0, T]] = \int_0^T \alpha N(t) p(t) dt.$$

Assuming perfect testing (as we do throughout this paper) the surveillance estimate of disease prevalence is simply the ratio  $\hat{p}_{surv}(T) = i_T/n_T$ . If the stochastic process representing the disease system is ergodic, and given the inclusion of immigration (see above) we can rule out extinction, the long time limit of this estimate can be equated with its ensemble average (expectation over all histories) i.e.,

$$\lim_{T \rightarrow \infty} \hat{p}_{surv}(T) = E[\hat{p}_{surv}] = \lim_{T \rightarrow \infty} \frac{\frac{1}{T} \int_0^T N(t) p(t) dt}{\frac{1}{T} \int_0^T N(t) dt} = \frac{E[N(t)p(t)]}{E[N(t)]}.$$

This can be re-expressed in the more suggestive form

$$E[\hat{p}_{surv}] = E[p(t)] + \frac{Cov[N(t), p(t)]}{E[N(t)]} \quad (1)$$

Thus when the covariance  $Cov[N(t), p(t)] = E[N(t)p(t)] - E[N(t)]E[p(t)]$  between the population size and the prevalence is non-zero the surveillance estimate of prevalence is a biased estimate of the true prevalence,  $E[p(t)]$ . Since  $Cov[N(t), p(t)]$  will be zero when either  $N(t)$  or  $p(t)$  are constant, this result leads to our first and most important conclusion,

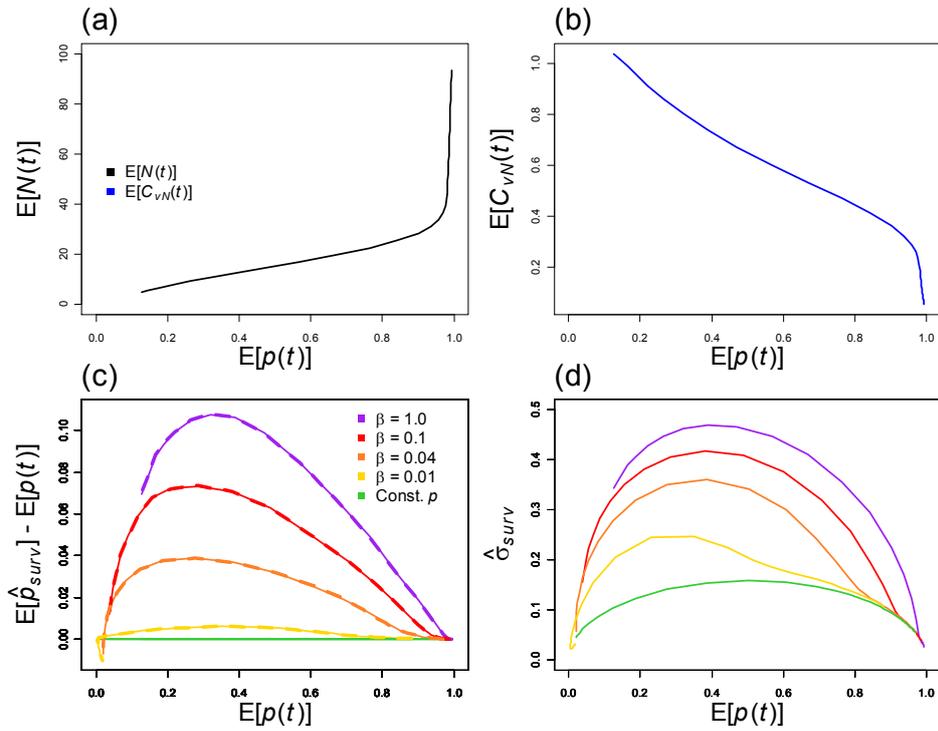
namely that demographic fluctuations and stochasticity in disease dynamics undermine the efficacy of surveillance.

#### *2.4.1.1 Effect of host demography and disease dynamics*

We now focus on surveillance estimates of prevalence based on finite sample sizes, and compare these to the continuous sampling theory prevalence estimate (see equation 1). Using the SDE implementation of the full model, Figure 2.1 illustrates how population fluctuations and disease dynamics in the host-pathogen system affect the efficacy of surveillance (in terms of the bias and variance of estimated prevalence). These results are generated by simulating the system for a range of values of the death rate  $\mu$ , with other parameters fixed. As the death rate increases the expected population size decreases and demographic fluctuations increase. For a given rate of disease transmission  $\beta$ , increasing the death rate reduces expected prevalence, and therefore simulating for different values of  $\mu$  generates the range of prevalence values shown. Details of the parameterisations used are given in Table S.1.1 (see section 1.2 in Appendix 1). The resulting relationship between demography and expected prevalence for particular disease characteristics (here a fixed transmission rate,  $\beta$ ) is illustrated in Figures 2.1.a and 2.1.b which show increasing population size and lower demographic fluctuations as expected prevalence increases. Simulations not shown here show that our results generalise, holding for transmission rates relative to a recovery rate (governing an additional transition from  $I$  to  $S$ ) and death rates relative to birth rate,  $r$ .

Figure 2.1.c shows the bias in the surveillance estimate of prevalence  $E[\hat{p}_{surv}] - E[p(t)]$  obtained from the same set of simulations. Results shown are based on  $10^6$  surveillance bouts with sample size  $m = 10$ , where for each bout sampling is conducted at rate  $\alpha$  until 10 individuals have been caught and tested. The bias predicted by continuous sampling theory (which does not account for sample size) is also shown, and in this case is extremely accurate i.e. it agrees with simulation results. Figure 2.1.c shows the bias in surveillance estimates of prevalence for four different transmission rates. It is important to note that the results shown are conditioned on the underlying prevalence  $E[p(t)]$ , and therefore for a given prevalence the populations associated with higher transmission rates are more variable than those with lower  $\beta$ . We therefore conclude that such variability increases the bias of surveillance estimates of disease prevalence. Finally, Figure 2.1.d shows the standard deviation in surveillance estimates of prevalence obtained from the same set of simulations. Comparison with the variability in prevalence estimates expected under the

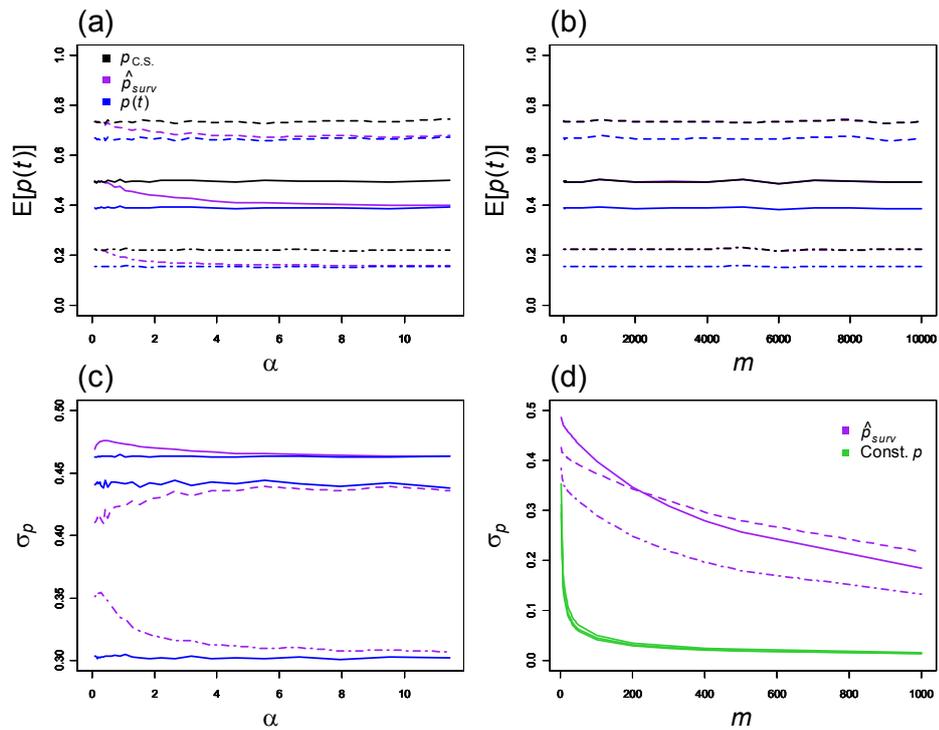
zero fluctuation assumption reveals that fluctuations in our simulated wildlife disease system considerably reduce the precision (increase the variance) of estimates obtained by surveillance. The variability of these estimates also increases with demographic fluctuations. Thus, the dynamics of the host-pathogen interaction are integral in determining the efficacy of surveillance in terms of prevalence estimation.



**Figure 2.1: Effect of host demography and disease transmission.** Data are shown for a range of values of the death rate  $\mu$  which controls the stability and size of the population, and thus determines disease prevalence for a given transmission rate,  $\beta$ . For  $\beta=1$  plot 2.1.a shows that expected population size increases with expected prevalence  $E[p(t)]$  (i.e. as  $\mu$  decreases) whilst plot 2.1.b shows that the coefficient of variation of the population size decreases. For the four values of  $\beta$  indicated and fixed sample size  $m=10$ , plot 2.1.c shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and plot 2.1.d the standard deviation in surveillance estimates of prevalence, versus the expected value of true disease prevalence in the system,  $E[p(t)]$ . Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see text) using the set of parameter values described in Table S.1.3 section 1.2 in Appendix 1.

#### 2.4.1.2 Surveillance design

An important determinant of surveillance efficacy is the design of the surveillance strategy itself. Figure 2.2 shows how the bias and variance of the estimate of prevalence changes as the intensity of surveillance (measured by the capture rate  $\alpha$ ) increases for fixed sample size (Figures 2.2.a and 2.2.c), and as the sample size,  $m$ , increases for a fixed capture rate (Figures 2.2.b and 2.2.d). For low capture rates and as  $\alpha \rightarrow 0$  the continuous sampling estimate given in equation (1) provides an accurate prediction for the level of prevalence estimated from surveillance based on fixed sample size, which as we saw above is a biased estimate of the true prevalence  $E[p(t)]$ . However, increasing the capture rate reduces bias, and as  $\alpha$  increases this bias tends to zero. In addition, for large capture rates, the precision of the surveillance estimate of prevalence matches the variability of the underlying wildlife disease system (see Figure 2.2.c). Thus for low capture rates the bias in surveillance estimates of prevalence is well described by continuous sampling theory (equation 1). However, for larger capture rates the properties of the surveillance estimate of prevalence increasingly reflect the expected true prevalence (i.e. bias reduces) and the variability in the prevalence of underlying disease system. In contrast, increasing sample size improves precision, but not bias (Figure 2.2.b). However, the precision is lower and improves less quickly with increasing sample size than that predicted by the standard approach that neglects fluctuations (see Figure 2.2.d).



**Figure 2.2: Effect of surveillance design.** In all plots results are shown for three wildlife disease systems with  $(\beta, \mu)$ : (1, 0.43) solid lines; (1, 0.4) dashed; and (0.1, 0.43) dot-dashed. Plots 2.2.a and 2.2.b show expected values of the surveillance estimate of prevalence (purple), the true prevalence (blue) and the continuous sampling theory prediction (black, see text for details). Plots 2.2.c and 2.2.d show the expected standard deviation (denoted,  $\sigma_p$ ) in both the true (blue) and the surveillance estimated (purple) prevalence. Plots 2.2.a and 2.2.c are plotted against a range of values of the capture rate  $\alpha$ , for  $m = 10$ , and 2.2.b and 2.2.d versus a range of sample sizes  $m$  for  $\alpha = 0.1$ . Plot 2.2.d also shows the constant prevalence estimate of the standard deviation based on the binomial (green). Parameter values used are described in Tables S.1.4 and S.1.5 (see section 1.2 in Appendix 1).

Figure 2.2 is based on simulated systems for which average host life spans are in the range of 2.3-2.5 time units and these results show bias in prevalence estimates for capture rates well above such levels. This suggests that demographic fluctuations will lead to bias in surveillance based estimates of prevalence unless surveillance intensity is high (i.e. corresponding to capture rates high enough to allow for individuals to be captured multiple times during their lifetimes). This does not imply that all individuals need to be tested, but just that the required samples should be gathered quickly relative to demographic fluctuations in order to reduce such bias. Nonetheless such capture rates are rare and only

occur in the most intensively sampled populations (see e.g. Delahay *et al.* 2000). Additional numerical results (not shown) indicate that as the sample size increases the capture rate required to obtain unbiased estimates increases, but even for large sample sizes when sampling is instantaneous sampling (i.e.  $\alpha \rightarrow \infty$ ) bias is zero and the standard deviation in the surveillance estimate of prevalence corresponds to that of the underlying wildlife disease system as shown above.

## 2.4.2 The Probability of Detection

In many cases the primary goal of wildlife disease surveillance is detection of disease rather than quantification of prevalence e.g. for emerging or re-emerging disease where detection is a precursor to further action, including heightened surveillance. If prevalence is assumed constant and equal to the long term, ensemble, average prevalence  $E[p]$  of the wildlife disease system, then the probability that disease is detected in a sample of size  $m$  is given by

$$PD^{Bin} = f(E[p], m) = 1 - (1 - E[p])^m \quad (2)$$

This formula, based on simple binomial arguments, and variants that also assume constant prevalence, are the standard basis for sample size calculations (see e.g. Fosgate 2009). However, we now demonstrate that  $PD^{Bin}$  is a misleading estimate of the probability of detection if prevalence fluctuates.

In real systems prevalence varies with time and therefore when conducting surveillance the prevalence values at the times each of the  $m$  samples are collected will vary. Nonetheless, for simplicity here we assume that the prevalence during a given surveillance bout (i.e. the collection of  $m$  consecutive samples) is constant and denoted  $p$ . The results shown in Figure 2.3.a, which compare the probability of detection measured directly in model simulations with approximations based on averaging over prevalence fluctuations both within and between and only between surveillance bouts, demonstrate that this is an accurate approximation. Then the expected probability of detection for sample size  $m$  is defined as

$$PD = E[f(p, m)] = E[1 - (1 - p)^m] \quad (3)$$

where the expectation is over the between bout prevalence distribution  $P(p)$ . For a single sample  $m = 1$ , equation (3) reduces to a linear form so that  $PD = PD^{Bin} = E[p]$ . However, if  $m > 1$ , then equation (3) is non-linear and therefore  $PD \neq PD^{Bin}$ . To illustrate this, we Taylor expand  $PD$  by assuming that the difference between the bout prevalence  $p$  and the long term average prevalence is small i.e.  $p = E[p] + \Delta_p$ . Then noting that  $E[\Delta_p] = 0$  and  $var[p] = E[\Delta_p^2]$  yields

$$PD \cong PD^{Bin} + \frac{1}{2}var[p] \left. \frac{\partial^2 f(p, m)}{\partial p^2} \right|_{p=E[p]} + H. O. T$$

This suggests (to leading order) that the true probability of detection will be lower than  $PD^{Bin}$ , since the second derivative  $\partial^2 f(p, m)/\partial p^2 = -m(m - 1)(1 - p)^{m-2}$  is negative for sample size  $m > 1$  and  $p = E[p]$ . In addition, the size of this deviation depends on the sample size and the variance in prevalence. Although the conclusions drawn are broadly correct, when compared with simulation results, the above Taylor expansion does not provide an accurate approximation to the probability of detection. However, analytic progress can be made with the following alternative approach. The approximation  $(1 - p)^m \approx e^{-pm}$  holds for  $m$  large (and is already accurate even for  $m = 10$ ) and enables us to write the probability of detection as

$$PD = 1 - E_p[(1 - p)^m] \cong 1 - E[e^{-pm}] = 1 - M_p(m) \quad (4)$$

where  $M_p(m) \equiv E[e^{-pm}]$  is the moment generating function associated with the between bout prevalence distribution  $P(p)$ . This suggests that if we could parameterise a suitable distribution to approximate  $P(p)$  then we could use the corresponding moment generating function to calculate the probability of detection.

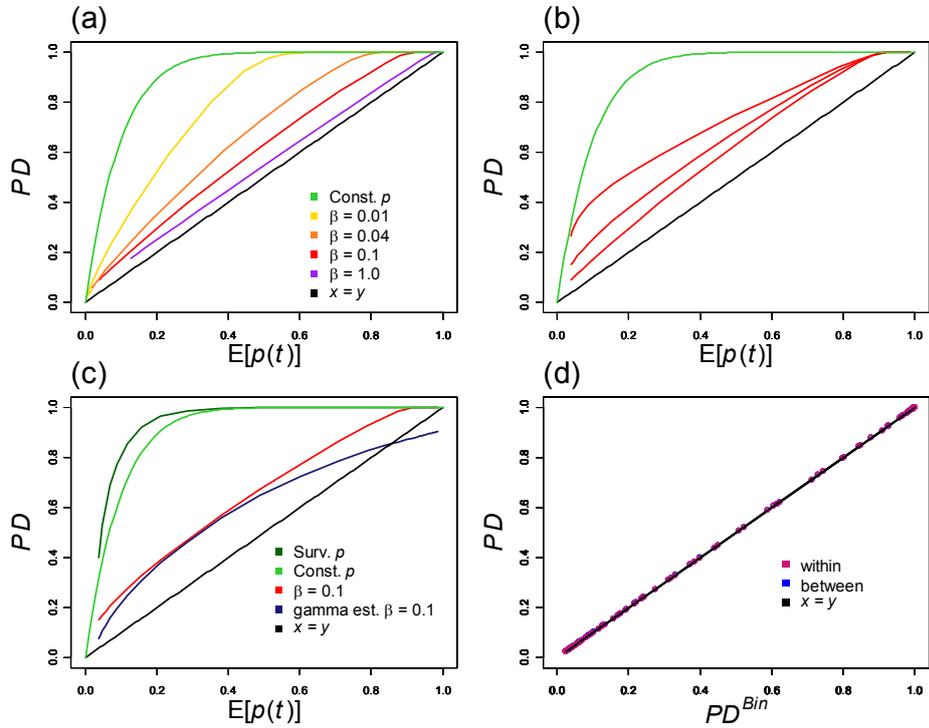
#### 2.4.2.1 Effect of host demography and transmission dynamics

The results shown in Figure 2.3 demonstrate the effect of host demography, transmission dynamics and surveillance design on the probability of detection. These results are obtained from the simulations described in Figure 2.1, except for those in Figure 2.3.d where these simulations are rerun for different values of the capture rate.

Figure 2.3.a suggests that a moment-generating function approximation (equation 4) based on the actual distribution of prevalence between surveillance bouts would be a very accurate approximation. Figure 2.3.b illustrates this approximation using an assumed gamma distribution parameterised with the mean and variance of  $P(p)$ . Although the gamma approximation is not completely successful it does provide a more accurate prediction of  $PD$  than  $PD^{Bin}$  and could improve sample size calculations in situations where simulation was not possible, but information about prevalence fluctuations was available. Moreover, the results of Figure 2.3.a show that such approximations could be improved by assuming a more accurate representation of the prevalence distribution  $P(p)$ . Crucially these calculations support the conclusion that the true probability of detection is less than obtained when ignoring fluctuations i.e. less than  $PD^{Bin}$ . Figure 2.3.b also shows the impact of biased prevalence estimation on disease detection for the case  $\beta = 0.1$ . Figure 2.1 demonstrated that in this case surveillance results in inflated estimates of prevalence  $E[\hat{p}_{surv}] > E[p(t)]$ . Ignoring the effect of fluctuations would therefore lead to an estimated detection probability even greater than  $PD^{Bin}$  which is based on the true average prevalence  $E[p]$ .

Figure 2.3.c shows the effect of interactions between disease dynamics and demography. As in the case of prevalence estimation, conditioned on a given expected prevalence, larger contact rates  $\beta$  are associated with greater fluctuations in the underlying wildlife disease system (i.e. greater transmission rates are needed to sustain a given prevalence). Here larger fluctuations translate into reduced probability of detection. In Figure 2.3.c for  $\beta = 1.0$  the probability of detection is only a little above the line  $PD = E[p]$  which corresponds to a single sample  $m = 1$ . Thus, in comparison with the zero fluctuation approximation  $PD^{Bin}$ , fluctuations reduce the effective sample size, for the  $\beta = 1.0$  case from  $m = 10$  to close to  $m = 1$ . Results not shown indicate that reduction in effective sample size increases with sample size (and see Figure 2.4). Figure 2.3.d shows the effect of capture rate on the probability of detection with more intense surveillance effort actually reducing the probability of detection. This is consistent with the above observations regarding  $\beta$ , since

less intense effort means that the required sample size takes longer to gather which reduces between bout fluctuations in prevalence.



**Figure 2.3: Effect of host-pathogen and surveillance dynamics on probability of detection.**

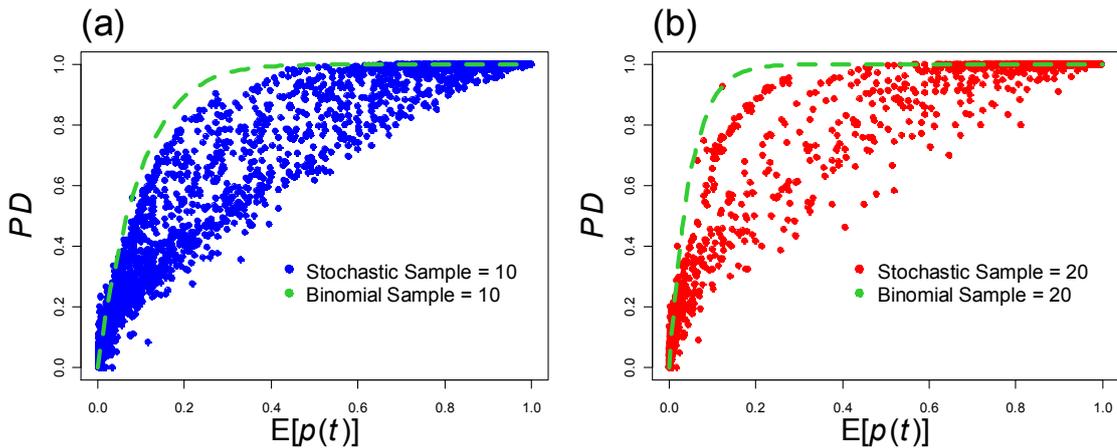
Results based on simulations used for Figure 2.1 (for details see Table S.1.6 in section 1.2 in Appendix 1). Plot 2.3.d estimated  $PD$  versus approximations based on modifications of equation (3) accounting for fluctuations in prevalence (i) within and between bouts and (ii) between bouts only. Plot 2.3.c shows  $PD^{Bin}$  based on both  $E[p]$  (green) and  $E[\hat{p}_{surv}]$  (black) and (for  $\beta = 0.1$ )  $PD$  and the approximation (equation 4) based on an assumed gamma distribution. Plot 2.3.a shows  $PD^{Bin}$  (green) and  $PD$  for various values of  $\beta$  (as shown yellow ( $\beta = 0.01$ ); orange ( $\beta = 0.04$ ); red ( $\beta = 0.1$ ); purple ( $\beta = 1.0$ )) versus actual prevalence  $E[p]$ . (b) shows  $PD^{Bin}$  (green) and  $PD$  for  $\beta = 0.1$  and the three capture rates  $\alpha = 0.01, 1.0, 10$ . In plots 2.3.a, 2.3.b and 2.3.c the black line indicates  $PD = E[p(t)]$ .

#### 2.4.2.2 Limits to disease detection in wildlife disease systems

Given that the nature of host demography and disease dynamics in wildlife disease systems will often be poorly understood, we carried out simulations similar to those explored above for a wide range of different host-pathogen combinations. These simulations focus on assessing the impacts of fluctuations on the ability of surveillance systems to detect disease

across a wide spectrum of wildlife disease systems. Since we focus on probability of detection the results are conditioned on the presence of disease and simulations are run using the Gillespie implementation of our model which explicitly handles the discrete nature of small populations. The details of these simulations, including the range of parameter combinations used, are described in detail in Table S.1.5 (see section 1.2 in Appendix 1).

Figure 2.4 shows how the probability of detection from surveillance, compares to the zero fluctuation approximation  $PD^{Bin}$ , at two different sampling levels, across this broad range of wildlife disease systems. Depending on the level of fluctuations in the system, the effective sample size can range from the actual number of samples taken all the way down to  $m \approx 1$ . These results suggest that ignoring the effect of fluctuations when designing surveillance could lead to studies that are underpowered in terms of disease detection. These results are consistent with those of Figure 2.3 based on the SDE implementation.



**Figure 2.4: Fluctuations reduce power to detect disease.** The two panels show the probability that disease is detected (conditional on non-zero prevalence) for target sample sizes 10 and 20. Each coloured dot represents the average of 100-1000 realisations of the model implemented using the Gillespie algorithm that met the sample target for a particular combination of parameters representing a distinct host-pathogen system (for details see Table S.1.7 in section 1.2 in Appendix 1). The green dashed line in both graphs represents  $PD^{Bin}$  the probability of detection assuming constant prevalence (see equation 2). It can be seen that  $PD^{Bin}$  generally over-estimates the power of the sample in that it predicts a larger probability of detection than is realised in the stochastic simulations.

## 2.5 Discussion

We believe this paper represents the first systematic exploration of the impact of pathogen transmission dynamics and demographic aspects of host ecology on wildlife disease surveillance efficacy. We have developed a generic framework in which surveillance in wildlife disease systems is characterised in terms of key demographic and epidemiological parameters alongside those representing the process of surveillance. Results were obtained using a combination of mathematical analysis and simulation, with different parameterisations used to represent a broad range of wildlife disease systems. We conclude that demographic and disease fluctuations reduce the power of surveillance to detect disease, and both bias and reduce the precision of estimates of prevalence obtained.

Current approaches to the design and analysis of surveillance in wildlife disease systems are largely based on methods developed for livestock (Grimes & Schulz 1996; Dohoo *et al.* 2005; Fosgate 2005) and typically ignore temporal fluctuations in prevalence and host population size. These assumptions lead to conclusions that, in the absence of extrinsic sources of bias, e.g. variation in habitat quality (Nusser *et al.* 2008; Walsh & Miller 2010) or biased capture (Tuytens *et al.* 1999), surveillance produces unbiased estimates of prevalence, and associated statistical power is determined completely by sample size and the underlying prevalence of disease. Our results demonstrate that such conclusions are not, in general, justified in wildlife disease systems where fluctuations bias prevalence estimates and reduce their precision compared with what would be expected in the absence of fluctuations. Similarly we find that the probability of disease detection increases with prevalence and sample size but at rates that progressively reduce as demographic and disease fluctuations increase. This suggests that wildlife disease surveillance programmes based on current theory are underpowered and produce biased results.

Here we have introduced a framework within which surveillance design is characterised by both sample size  $m$ , and the capture rate  $\alpha$  instead of simply sample size. Moreover, in this extended framework the performance of surveillance is assessed in light of the ecology of the wildlife disease system of interest (i.e. for particular population and disease parameters). A key insight is that sample capture (e.g. time taken to reach the sample target) is dependent on both the surveillance design and the ecology of the host (Nusser *et al.* 2008), represented here in terms of demography. Our results show that surveillance design (choice of  $m$  and  $\alpha$ ) can have a large impact on bias and precision of prevalence estimation and on the power to detect disease. Bias in prevalence estimates increases and

the precision of such estimates reduces with more unstable populations and greater fluctuations in disease. Such bias can be reduced by increasing capture rate, but for fixed sample size this also reduces the ability to detect disease. However, simulation results suggest that for all but the most intensive wildlife disease surveillance programs (Delahay *et al.* 2000) typical capture rates are not sufficient to eliminate bias. In contrast increasing sample size does not affect bias, but does improve statistical power in terms of both precision of prevalence estimates and disease detection. However, as sample size increases such improvements in power are not as fast as would be expected if fluctuations were ignored, as they are in current surveillance design and analysis.

The framework introduced here could be extended to account for details of particular wildlife disease surveillance programmes including the sensitivity and specificity of diagnostic tests, application to specific host populations and pathogens, multiple diagnostic test results (that are increasingly available), aspects of syndromic surveillance (Dórea *et al.* 2011) and more complex deployment strategies than those considered here. Characterisation of the outcome of surveillance e.g. in terms of bias, precision and probabilities of disease detection would aid the identification of efficient designs and interpretation of any results obtained. However, the development of more formal statistical analyses that account for demographic fluctuations and stochastic disease dynamics remains the subject of future research.

Surveillance is a critical prerequisite for defining and controlling wildlife disease risks and our results suggest that current approaches to the design and analysis of wildlife disease surveillance ignore fundamental aspects of ecology leading to inadequate assessment of risk. Moreover, these problems (unknowingly under powered studies and biased results) are likely to be very widespread given that the ecology of many wildlife species and the pathogens to which they are exposed, lead to significant temporal fluctuations in both population size and prevalence (Anderson & May 1979; Anderson 1991; Renshaw 1991; Wilson & Hassell 1997).

There is much current interest in quantifying risks from wildlife disease (Daszak *et al.* 2000; Jones *et al.* 2008) and this is stimulating debate on the need to improve wildlife disease surveillance (Bengis *et al.* 2004; Butler 2006; Gortázar *et al.* 2007; Béneult *et al.* 2014). This paper contributes to this debate, highlighting the need to consider the ecology of wildlife disease systems when designing or analysing surveillance programs (Béneult *et al.* 2014), emphasizing the importance of temporal heterogeneities induced by demographic stochasticity and disease dynamics. Further research is needed to assess the impacts of

alternative and complimentary heterogeneities including intrinsic and extrinsic forms of spatial heterogeneity and other population structures. There is a rich literature describing the effects of such heterogeneity in ecology and epidemiology (Huffaker 1958; Lloyd & May 1996; Tilman & Kareiva 1997; Keeling *et al.* 2000; Read & Keeling 2003; Keeling 2005; Vicente *et al.* 2007) and our results suggest that these are likely to important, but as yet unexplored, impacts on the efficacy of wildlife disease surveillance.

## Chapter 3

An Approach to Assessing Wildlife  
Disease Surveillance in Real Systems:  
tools and applications

### 3.1 Abstract

We show how the theoretical framework presented in Chapter 2 can be further developed and applied as a tool to assess the outcomes of wildlife disease surveillance in real systems. In so doing, the effects on surveillance of fluctuations demonstrated in Chapter 2 are quantified for two real host populations. Our results show, using examples of badger and rabbit populations, that the impact of demographic fluctuations and stochasticity of disease dynamics on surveillance efficacy is a justifiable concern for current control strategies of key wildlife populations and diseases. Using these tools we also offer novel analyses of recognised sources of bias in natural populations i.e. disease induced mortality, diagnostic test sensitivity and trappability. These are shown to either further contribute to, or mask the bias occurring from natural stochastic fluctuations within the population. Understanding these subtle effects and the impact they have on surveillance efficacy is key in the pursuit of better control strategies, which ultimately offer protection for wildlife, livestock and humans against the threat of disease.

### 3.2 Introduction

Chapter 2 identified that combined fluctuations in population and disease dynamics bias surveillance estimates of prevalence with respect to the true mean. It was also shown that the variability in the prevalence estimate is much larger (i.e. the precision of the estimate is much lower) than current methods (i.e. binomial theory) estimate and both effects increase with the size of fluctuations in the host-pathogen system. Furthermore variation in prevalence depresses the probability of detection compared to current theoretical estimates, based on binomial arguments, which ignore such effects. Overall, these results demonstrate the importance of assessing the host-pathogen system more carefully before designing surveillance strategies.

This Chapter aims to show how by building on the generic theoretical framework described in Chapter 2 we can employ stochastic models describing the dynamics of wildlife disease systems as tools which enable assessment of surveillance in real disease systems of current interest. We extend the stochastic model from Chapter 2 to model two endemic disease systems, and quantify surveillance in each in terms of the extent to which standard theoretical tools which ignore fluctuations are able to predict the performance of surveillance. Here we model badger and rabbit populations, simulating population dynamics

in conjunction with the spread of Bovine Tuberculosis (*Mycobacterium bovis*) (TB) and Paratuberculosis (*Mycobacterium avium* subsp. *paratuberculosis*) (paraTB) respectively. Both diseases are endemic in many countries, including the UK, and are associated with economic loss to livestock production systems and health risk to humans.

In addition rabbits and badgers can be thought of as representative of, respectively, *r* and *K* species under MacArthur and Wilson's (MacArthur & Wilson 1967) general theory. An *r* species has large population fluctuations seen as booms and busts of growth, resulting from high birth and death rates and typically lives in an unstable environment. A *K* species has population growth limited by resources and competition, lives in a stable environment, and has a relatively low birth and death rate; and hence lower population fluctuations. Thus, as host population fluctuations were shown in Chapter 2 to impact on surveillance, the results we obtain for rabbits and badgers will be relevant to other *r* and *K* species.

Badgers were first identified as a host for *M. bovis* in 1974 (Byrne *et al.* 2012b) and have since been included in TB control strategies in the UK. During the comprehensive Randomised Badger culling Trial (RBCT), which aimed to test the hypothesis that badger population reduction would reduce the TB prevalence in sympatric cattle, badgers were shown as a source of infection to cattle (DEFRA 2013). Under the RBCT proactive culling of badgers (population reduction irrespective of cattle TB status) was associated with an average of 16% reduction in TB incidence in cattle. However, reactive culling of badgers (population reduction in response to TB in cattle) was associated with an increase in TB in badgers (Carter *et al.* 2007) and cattle (Donnelly *et al.* 2003, 2006).

Rabbits were identified as a host of *M. avium* subsp. *paratuberculosis* in 1997 (Greig *et al.* 1997; Beard *et al.* 2001). Modelling studies suggest that the direct and indirect routes and rates of rabbit to rabbit transmission identified and quantified in the field (Judge *et al.* 2006) are sufficient for the disease to persist in rabbit populations (Judge *et al.* 2007). Rabbits excrete millions of bacteria in their faeces (Daniels *et al.* 2001) which are not avoided and are therefore ingested by grazing livestock (Judge *et al.* 2005). More recently paratuberculosis in rabbits has been associated with a failure of disease control operations in cattle (Shaughnessy *et al.* 2013), and may be associated with Crohn's disease in humans (Naser *et al.* 2014).

Wildlife disease surveillance is also subject to a number of well known inaccuracies and biases, including diagnostic limitations e.g. sensitivity, and biases in animal sampling e.g.

trapping biases. Diagnostic limitations are a result of the test sensitivity and specificity which determine how accurate test results are. The sensitivity and specificity of the test can bias results by reporting false information, giving the impression of more/less disease than is actually present in the population. Research funding in wildlife disease surveillance is limited and in any case developing more accurate tests may not always be feasible due to technical and scientific challenges. Sampling biases can also occur through the interaction of survey implementation and the behaviours of the sampled population. For example, animals burdened with disease can become more or less likely to be trapped (trap shy / trap happy) which can lead to unrepresentative results.

Accurate surveillance and monitoring of infection in these disease systems is essential for the quantification of the biological processes underpinning disease persistence and spread, accurate assessment of risk, and for the efficiency of disease control operations. In the current chapter we seek to understand the potential impacts of the above effects on two contrasting host systems. Therefore whilst in the previous chapter we explored a range of demographic parameters here we consider a range of possible pathogens, characterised in terms of both transmission rate  $\beta$  and disease induced mortality rate  $\mu_i$ , infecting populations of badgers and rabbits. Within this context TB in badgers and *paratuberculosis in rabbits* correspond to particular combinations of transmission rate  $\beta$  and disease induced mortality rate  $\mu_i$ . As in Chapter 2, we consider one form of surveillance; active capture and release (the capture and testing of individuals which are then released back into the population). Simulations are run to assess to what extent the results found in Chapter 2 are relevant to specific wildlife populations and diseases. We also investigate how the performance (e.g. sensitivity) of the diagnostic test and wildlife sampling biases (e.g. trappability) influence the efficacy of the surveillance strategy in terms of the probability of detection and the bias and variance (SD) in estimating the prevalence.

## 3.3 Methods

### 3.3.1 Stochastic Model Description

The model represents a host population subject to demographic fluctuations (births, deaths and immigration) and the transmission of a single pathogen species/strain. At each point in time  $t$ , the state-space represents the total population size  $N(t)$ , with  $I(t)$  of these infected and  $S(t) = N(t) - I(t)$  susceptible. In addition the prevalence is then given by  $p(t) = I(t)/N(t)$ . The birth rate of individuals is logistic,  $rN(1 - N/k)$ , with intrinsic growth rate  $r$  and a carrying capacity  $k$  reflecting the assumptions that population growth is resource limited. Individuals have a per capita death rate  $\mu$  and if there is disease induced mortality in the system, infected individuals have an additional disease induced death rate  $\mu_i$  i.e. infected individuals have a total death rate of  $\mu + \mu_i$ . Immigration occurs at a constant rate  $\nu$  with a proportion  $\gamma$  of immigrants are infected. As immigration is set as a constant rate, this can push the population size above the carrying capacity  $K$ , at which point the birth rate is set to 0.

Disease surveillance is incorporated into the model in the form of active capture, testing and release at per capita rate  $\alpha$  for both susceptible and infected individuals. The time for which individuals are removed from the population is assumed negligible. Surveillance testing is undertaken assuming perfect tests, unless otherwise specified by the sensitivity/specificity level. A summary of this conceptual model is given in Chapter 2 in Table 2.1 which shows all demographic, epidemiological and surveillance events with their corresponding rate and effect on the state-space. However, in this chapter, note that death of infectives occurs at rate  $\mu + \mu_i$ . For a more detailed description of the model and how the surveillance is modelled, see Chapter 2.

### 3.3.2 Model Implementation

The model is implemented firstly as a continuous time continuous state space Markov process, which is simulated using Stochastic Differential Equations (SDE) as described in Appendix 1. Secondly the model is implemented as a continuous-time discrete-state space Markov process, which is simulated using Gillespie's exact algorithm (Gillespie 1976). Descriptions of the implementation can be seen in Chapter 2. As noted above all events with their corresponding rate and effect on population size for the Gillespie implementation can be found in Table 2.1 in Chapter 2.

### 3.3.3 Model Parameterisation

Parameters used to simulate the badger and rabbit social groups were estimated using information gathered from the literature including Shirley *et al.* (2003) and Judge *et al.* (2007) respectively. Appendix 2.1 provides details but an outline of the procedure followed is described here. A key issue in establishing parameter values used here is the differences between the structure of our models and those used in the cited literature. As outlined above here we adopt a continuous time framework whereas Shirley *et al.* (2003) used a discrete time model. This difference necessitates a translation between transition probabilities associated with the characteristic time step used in the discrete time model and the event rates used here. Judge *et al.* (2007) uses continuous time also, however the rates are calculated for time measured in months and therefore needed some adjustment. Furthermore both (Shirley *et al.* 2003) and (Judge *et al.* 2007) model several age classes and both males and females. In contrast here we neglect such population structure and therefore must consolidate transition probabilities across many different age by sex classes in order to calculate a single rate per event considered i.e. birth and death rates, disease induced mortality and disease transmission rates. Details of these calculations are provided in Appendix 2.1.

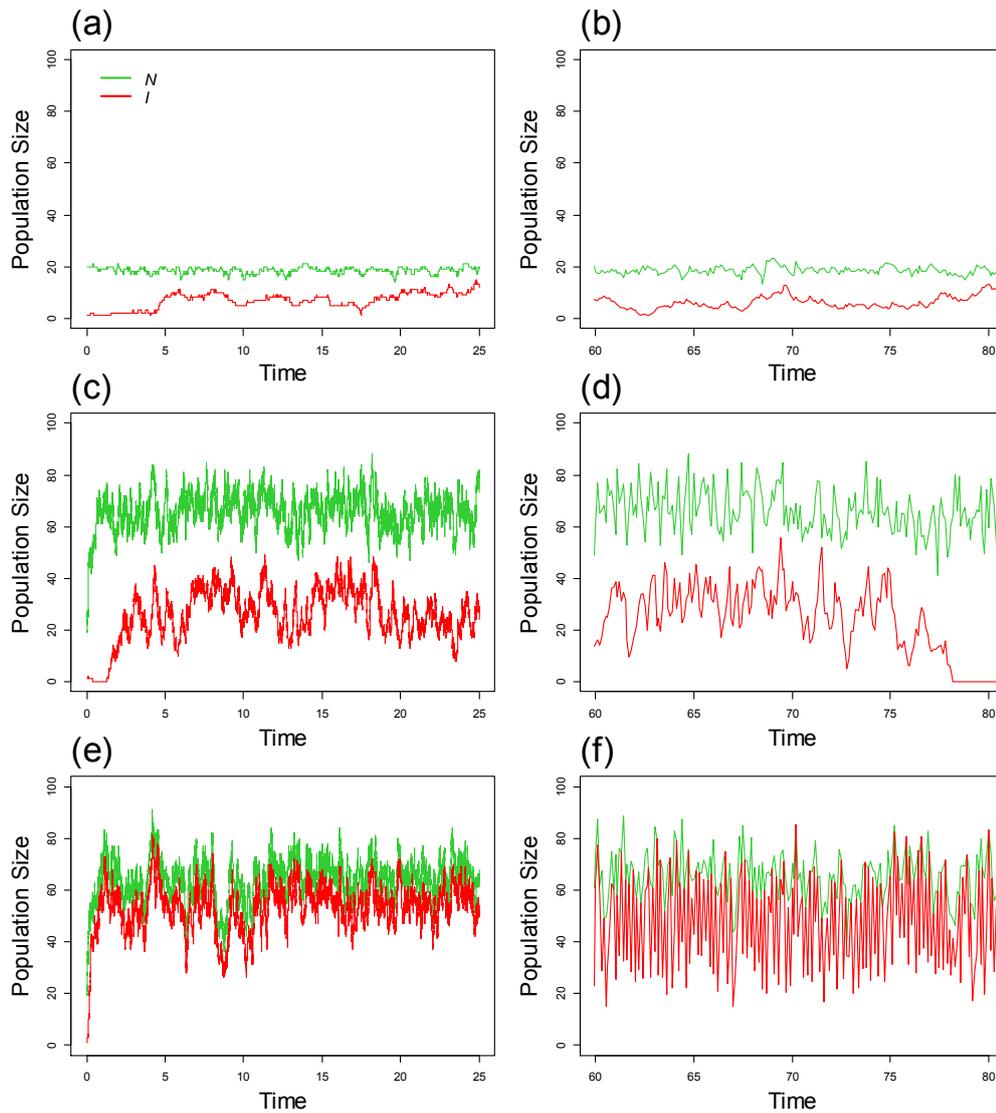
In addition to the above figures for badgers we also chose to model the average badger group size to be around 18 individuals based on the average group size observed at the Woodchester Park study site (Zijerveld 2012). Since in most parts of the UK badger group sizes are typically below 10 (Neal & Cheeseman 1996; Delahay *et al.* 2000) we set the carrying capacity (the maximum population above which the group size can't go) to be  $k=20$ . However, we note that our results and conclusions regarding the impact of fluctuations on surveillance in badgers are robust to alternative choices for the carrying capacity,  $k>20$ . We consider the mortality rates derived from (Shirley *et al.* 2003) to be accurate and therefore since births must match deaths at an equilibrium of approximately 18 this determines the growth rate parameter  $r$ . Considering figures for the number of offspring per breeding female per year given by Shirley *et al.* (2003) we find our estimates of the birth rate are consistent with observational studies (Woodroffe and MacDonald, 1995) which suggest approximately two breeding females. In the case of rabbits the birth rate was similarly determined from the derived mortality rate, the carrying capacity and the desired equilibrium population size. The parameterisations representing badgers and rabbits are shown in Tables S.2.1 and S.2.2 in Appendix 2.2.

### 3.3.4 Simulations

In contrast to the previous chapter here we explore a range of different diseases while keeping the demographic parameters fixed to represent both badger and rabbit populations. Within the model disease is characterised by disease induced mortality and transmission rate. Simulations were run using the SDE implementation of our model, firstly varying the disease induced (excess) death rate, but keeping other parameters fixed, leading to a range of prevalence. These simulations are repeated for several different transmission rates (some of which correspond to bTB transmission in badgers and paratuberculosis in rabbits) to demonstrate the effects of population fluctuations and disease dynamics on surveillance estimates of prevalence and probability of detection in badger and rabbit populations. Simulations were also run for a range of transmission rates (generating a range of prevalence) with other parameters fixed but repeated for several different disease induced mortality rates. Capture rates and sample size were also varied to explore surveillance design and differences between binomial sample size theory covered in Chapter 2 and simulation results in badger and rabbit populations.

Using the Gillespie implementation of the model, simulations were run firstly to show how trappability of infected individuals affects the probability of detection and prevalence estimates seen previously. Secondly, Gillespie algorithm simulations were run to show the effects of test sensitivity and transmission rate on the probability of detection and estimates of prevalence obtained by surveillance.

Figure 3.1 shows example trace plots from both the SDE and Gillespie implementations for rabbit and paraTB and badgers and TB. In the rabbit-paraTB system two transmission rates corresponding to the estimated range of transmission rates found by Judge *et al* 2006, are labelled low and high as they correspond to both ends of the likely prevalence range. In the case of badgers and TB we used one transmission rate which we consider representative of the literature (see above). Figure 3.1 demonstrates what can be seen in a single simulation of badger and rabbit populations with TB and paraTB respectively. The badger population is noticeably smaller and more stable whereas the rabbit population shows greater population fluctuations. There are close similarities between the Gillespie trace plots and the SDE trace plots which demonstrate their comparative similarities.



**Figure 3.1: Trace plots from SDE and Gillespie implementation runs of the model.** Data are shown for total population  $N$  (green) and infected population  $I$  (red) for three different real disease systems; Badgers and TB, Rabbits and low paraTB and Rabbits and high paraTB (corresponding to the range of paraTB transmission rates estimated in the literature used). The left hand plots 3.1.a, 3.1.c and 3.1.e show results based on the Gillespie implementation of the model, whereas the right hand plots 3.1.b, 3.1.d and 3.1.f show results from the SDE implementation of the model. The badger and TB results are shown in plots 3.1.a and 3.1.b. The rabbit-paraTB results are shown in plots 3.1.c and 3.1.d (low prevalence case) and 3.1.e and 3.1.f (high prevalence case). The plots shown were run using the set of parameter values described in Tables S.2.1 and S.2.2 in Appendix 2.2.

### 3.3.5 Statistics generated from the model

As previously discussed in Chapter 2, we assume long term averages are equivalent to ensemble expectations (approximated by averages over many realisations of the process). Each simulation is run for a period of time to allow the population to reach equilibrium before such statistics are calculated. For example, the expected mean  $E[M]$  and variance  $\text{Var}[M]$  of the population size are recorded along with the expected mean  $E[p]$  and variance  $\text{Var}[p]$  of disease prevalence. Similarly other statistics such as the covariance between the prevalence and population size are calculated as required. For a more detailed description of the calculation process see methods section of Chapter 2.

## 3.4 Results

In Chapter two we showed, for the case where there is no sample target and the period for which sampling is conducted is extended indefinitely, that the surveillance estimate of prevalence will always be biased unless the covariance of population size  $N$  and prevalence  $p$  is equal to 0 with the equation:

$$E[\hat{p}_{Surv}] = E[p(t)] + \frac{\text{Cov}[N(t), p(t)]}{E[N(t)]}$$

This result proved to be a reliable indicator of surveillance estimates of prevalence based on finite samples for the case where surveillance effort (capture rate) is low. Moreover simulations demonstrated that qualitatively these effects occur for a wide range of parameterisations. Building on these earlier results, simulations have been run to explore the same fundamental effects in models of demography and disease dynamics parameterised to represent badger and rabbit populations. In Chapter 2 simulations for a range of death rates were used to generate the full range of prevalence levels. However, varying the total death rate may have such an impact on the disease free demography of the population that respective model outputs are no longer representative of badgers and rabbits. We therefore consider varying the disease induced death rate so as not to majorly impact on the stability of the population itself but rather on the stability of the disease.

### 3.4.1 Estimating disease prevalence in badger and rabbit populations

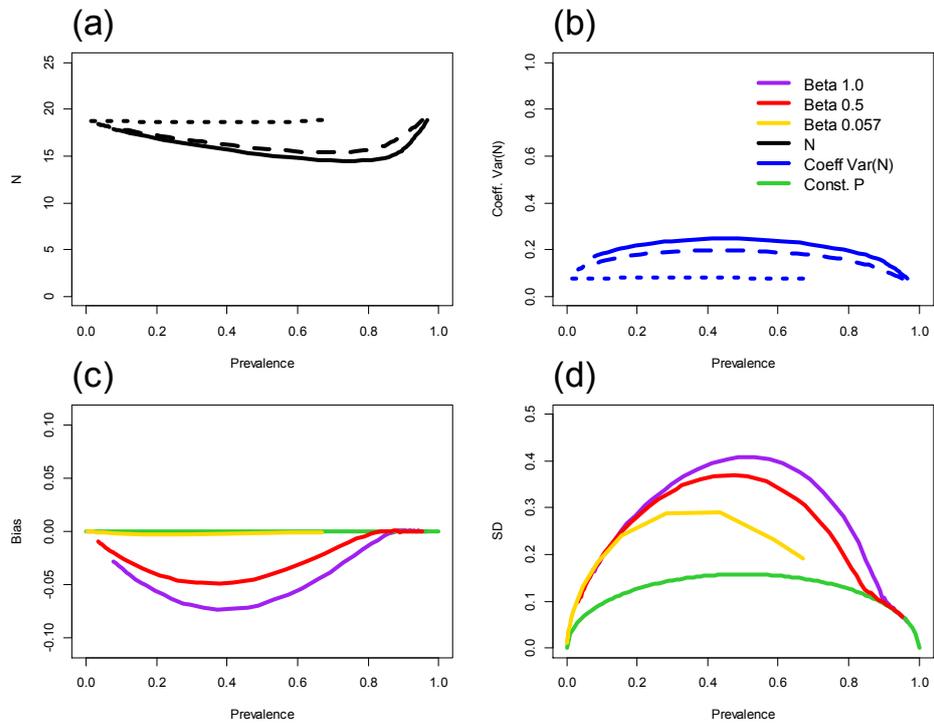
#### 3.4.1.1 Effect of disease system

Figure 3.2 explores the impact of different diseases in populations of badgers by varying disease induced mortality for several different values of  $\beta$ . These results illustrate the extent to which population fluctuations and stochasticity in disease dynamics in badger host-pathogen systems affect the bias and variance of the surveillance estimate of prevalence.

Figure 3.2.a shows the total population size  $N$  for a full range of disease prevalence, at three different levels of  $\beta$ . In Figure 3.2.a zero prevalence corresponds to high levels of disease induced mortality that rapidly remove diseased individuals from the population. Moving from left to right prevalence increases and due to disease induced mortality this reduces the population size. However because the increase in prevalence is driven by reducing levels of disease induced mortality this effect of disease on the population size diminishes as prevalence increases. Moving further to the right the reduction in disease induced mortality starts to compensate for the corresponding increase in prevalence and the population size is less affected by disease losses. Ultimately for a prevalence of 1 disease induced mortality is negligible and the population size is the same as the case with no disease.

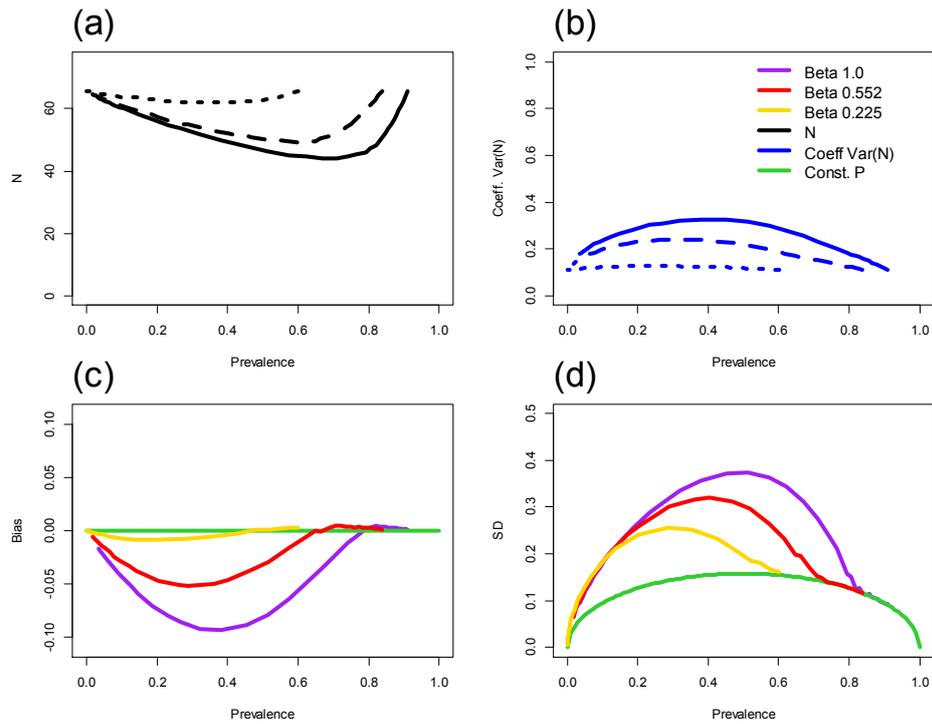
At any given level of prevalence, as the contact rate  $\beta$  increases the underlying dynamics of the system are more unstable; they require greater levels of contact to maintain the disease at the given prevalence. A given level of prevalence results from a balance between transmission and disease induced mortality so as the contact rate  $\beta$  increases so does the disease induced mortality required to generate the given prevalence. This increase in  $\mu_i$ , when  $\beta$  increases at a given prevalence, explains why population size reduces with increased contact rate. The variation in population size shown in Figure 3.2.b also reflects the pattern described above, where for both the disease-free state and the state of complete infection the variance in population size is at a minimum where it is determined purely by demographic, as opposed to epidemic, processes. However for other prevalence levels the variance is inflated by additional fluctuations produced by stochasticity in the disease dynamics.

Figure 3.2.c shows the bias in the surveillance estimate of prevalence  $E[\hat{p}_{surv}] - E[p(t)]$  for a sample size of 10 (i.e. sampling occurs at rate  $\alpha$  until  $m = 10$  individuals have been caught and tested) as obtained from simulating  $10^6$  surveillance bouts using the SDE implementation of the model described above. Figure 3.2.d shows the standard deviation in surveillance estimates of prevalence obtained from the same simulation results. Comparison with the variability expected under the zero fluctuation assumption reveals that fluctuations in our simulated badger population considerably reduce the precision of prevalence estimates obtained by surveillance. As we saw in Chapter 2 (and above) these effects increase with the contact rate, since at a given expected prevalence larger values of  $\beta$  are associated with more unstable dynamics. The  $\beta$  value of 0.057 (shown in yellow) represents TB in badgers. Although relatively little bias is observed, the SD in the prevalence estimate is seen to be just under double that of the constant prevalence prediction at the point where the prevalence is around 0.48 (for the yellow curve). This prevalence was selected as representative of badgers (for details see Appendix 2.2).



**Figure 3.2: Effect of infected population stability and disease transmission on surveillance in Badgers.** Data are shown for three different values of the transmission rate each for a range of values of the infected death rate  $\mu_i$  which controls the stability and size of the infected population and thus determines disease prevalence for a given transmission rate (see text). Each quantity considered is plotted against the resulting expected value of true disease prevalence in the system,  $E[p(t)]$ . Plot 3.2.a shows that expected population size for three contact rates  $\beta$  (1.0 – full line, 0.5 – dash line, 0.057 – smaller dash line) with increased prevalence (i.e. as  $\mu_i$  decreases). Plot 3.2.b shows the coefficient of variation of the population size for same values of  $\beta$ . Plot 3.2.c shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and plot 3.2.d the standard deviation in surveillance estimates for three different  $\beta$  (1.0-purple, 0.5-red, 0.057-yellow) (based on a fixed sample size  $m = 10$ ) of . The green lines show the bias (3.2.c) and standard deviation (3.2.d) obtained by assuming constant prevalence. Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the set of parameter values described in Table S.2.3 in Appendix 2.2.

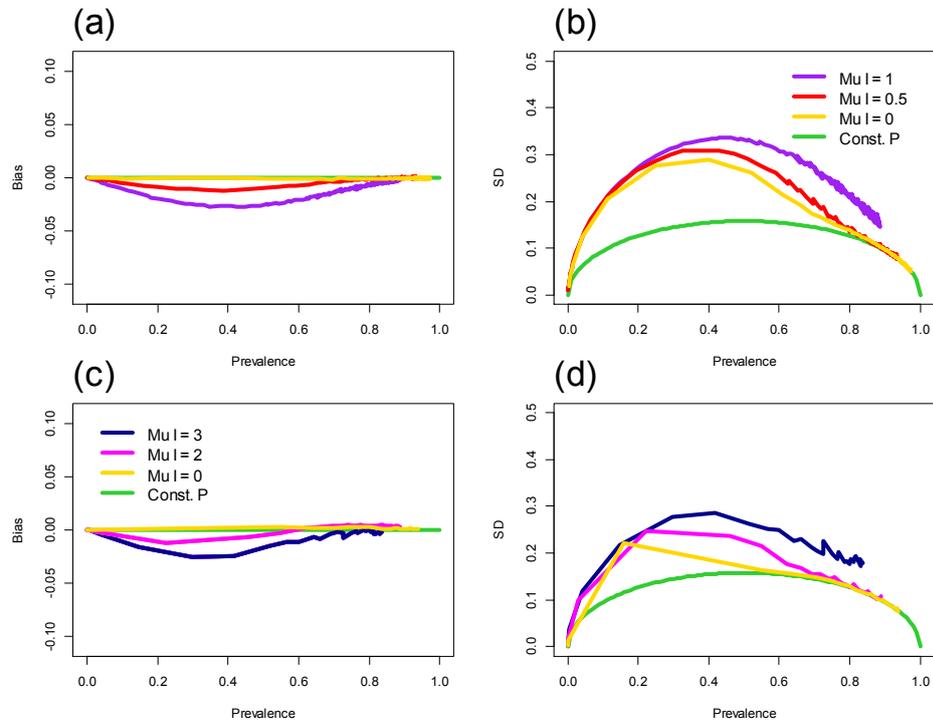
Figure 3.3 quantifies the same range of effects on prevalence estimates and population size and variance in our simulated rabbit population. Plot 3.3.a shows the total population  $N$  at three different levels of  $\beta$ . Figure 3.3.b shows the coefficient of variation for the same values of  $\beta$ . Plot 3.3.c shows the bias in the surveillance estimate of prevalence  $E[\hat{p}_{surv}] - E[p(t)]$  for a sample size of 10 (i.e. sampling occurs at rate  $\alpha$  until  $m = 10$  individuals have been caught and tested) as obtained from simulating  $10^6$  surveillance bouts using the SDE implementation of the model described above. Figure 3.3.d shows the standard deviation in surveillance estimates of prevalence obtained from the same simulation results. Similar effects to those found above from badgers are seen in the case of rabbits and over the whole prevalence spectrum. Moreover these results suggest that the effects described in a more theoretical context in Chapter 2 are of importance in these real wildlife-disease systems. Both the  $\beta$  values of 0.225 (yellow) and 0.552 (red) represent paraTB in rabbits. For cases of no disease induced mortality (e.g. paraTB), corresponding to large prevalences, we can see that in this instance, the constant prevalence estimate is very accurate. The effects seen in badgers and previously in Chapter 2 are more likely when the population becomes unstable (i.e. by introducing disease induced mortality). We see in both badger and rabbit examples, the effects diminish in the limit when there is no disease induced mortality (see Figure 3.2 and 3.3).



**Figure 3.3: Effect of infected population stability and disease transmission on surveillance in Rabbits.** Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the set of parameter values described in Table S.2.4 in Appendix 2.2 except for constant prevalence theory estimates shown in green. Data from model simulations are shown for three different values of the transmission rate  $\beta$  (1.0-purple/full line, 0.552-red/dash line, 0.225-yellow/smaller dash line) each for a range of values of the infected death rate  $\mu_I$  which controls the stability and size of the infected population and thus determines disease prevalence for a given transmission rate. All results are plotted against the resulting expected value of true disease prevalence in the system,  $E[p(t)]$ . Plot 3.3.a shows the expected size and Plot 3.3.b the coefficient of variation of the population. Plot 3.3.c shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and Plot 3.3.d the standard deviation in surveillance estimates (based on a fixed sample size  $m = 10$ ).

We can explore a different, but overlapping range of diseases by varying the transmission rate,  $\beta$ . We demonstrate similar effects in the bias and variance of the prevalence estimate in Figure 3.4 for both badgers (3.4.a and 3.4.b) and rabbits (3.4.c and 3.4.d) by varying the pathogen transmission rate, for three fixed disease induced mortality levels. The effects on the performance of surveillance, seen when varying  $\beta$  are similar to those seen when

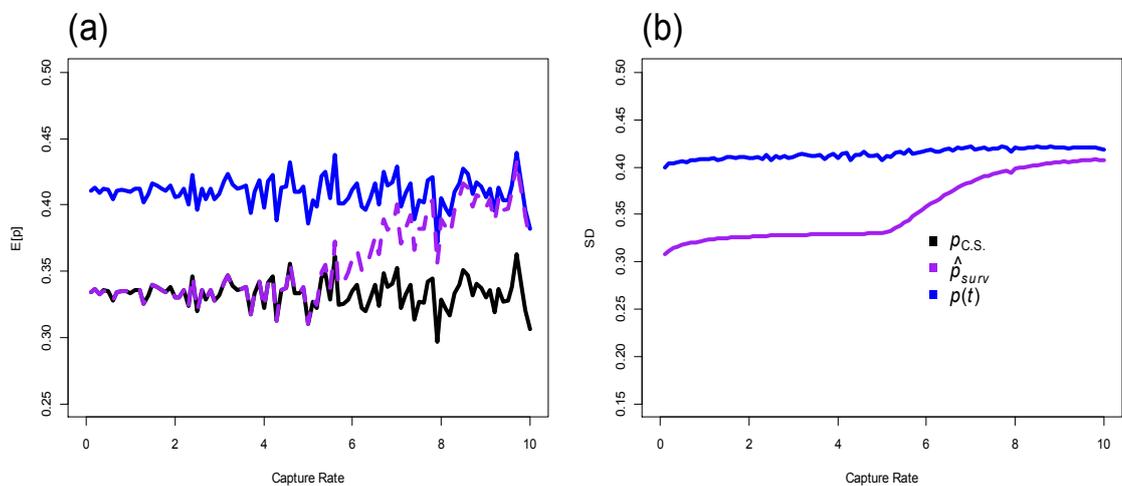
varying the disease induced mortality rate. ParaTB can be found on plot 3.4.c and plot 3.4.d when disease induced mortality is 0 (yellow line) for prevalence between  $\sim 0.4$  and  $\sim 0.8$ . Badgers and TB are located between the red and yellow lines for a prevalence of around 0.48.



**Figure 3.4: Effect of disease transmission and infected population death rate on surveillance.** Plot 3.4.a shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and plot 3.4.b the standard deviation in surveillance estimates (based on a fixed sample size  $m = 10$ ) of prevalence versus the expected value of true disease prevalence in the system,  $E[p(t)]$  for badger populations. Data are shown for three different values of disease induced mortality rate  $\mu_I$  (1.0-purple, 0.5-red, 0-yellow) each for a range of values of the infected death rate  $\beta$ . Plot 3.4.c shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and plot 3.4.d the standard deviation in surveillance estimates (based on a fixed sample size  $m = 10$ ) of prevalence versus the expected value of true disease prevalence in the system,  $E[p(t)]$  for rabbit populations. Data are shown for three different values of disease induced mortality rate  $\mu_I$  (3.0-navy, 2.0-pink, 0.0-yellow) each for a range of values of the infected death rate  $\beta$ . Constant prevalence theory estimates are also shown (green). Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the set of parameter values described in Table S.2.5 and S.2.6 in Appendix 2.2.

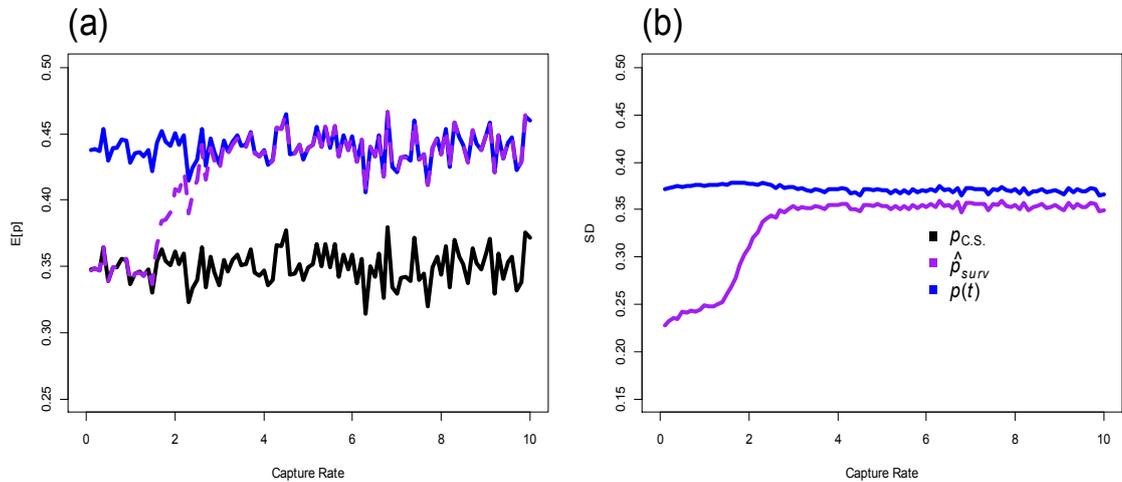
### Effect of surveillance design

In Chapter 2, it was shown that the surveillance strategy (sample size and capture rate) impact on the surveillance efficacy. As in Chapter 2 sample size was found not to have a big impact on the levels of bias in the examples that were explored here. The standard deviation of the prevalence estimates decreases by about 5% very rapidly and then continues to decrease very slowly as the sample size increases and this decrease is still much slower than predicted when fluctuations are ignored. Figure 3.5 shows the more interesting effect of capture rate on a badger population where the bias and variance of the estimate of prevalence changes as the intensity of surveillance (capture rate  $\alpha$ ) increases for fixed sample size (plot 3.5.a and 3.5.b).



**Figure 3.5: Effect of surveillance design.** In all plots results are shown for surveillance in Badgers ( $\beta = 1$  &  $\mu_i = 1.75$ ). Plot 3.5.a shows expected values of the surveillance estimate of prevalence (purple), the true prevalence (blue) and the continuous sampling theory prediction (black) for a range of values of the capture rate for a fixed sample size  $m = 10$ . Plot 3.5.b shows the corresponding value of the expected standard deviation ( $\sigma_p$ ) in the surveillance estimate of prevalence (purple) and the true prevalence (blue) for a range of values of the capture rate for sample size  $m = 10$  for a range of values of the capture rate  $\alpha$ . Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the set of parameter values described in Table S.2.7 in Appendix 2.2.

Figure 3.6 shows the same effects are seen in a rabbit population. It can be seen below how the differences in the two populations (turnover and stability) impact on how the capture rate and sample size affect the mean and SD of the prevalence estimate.

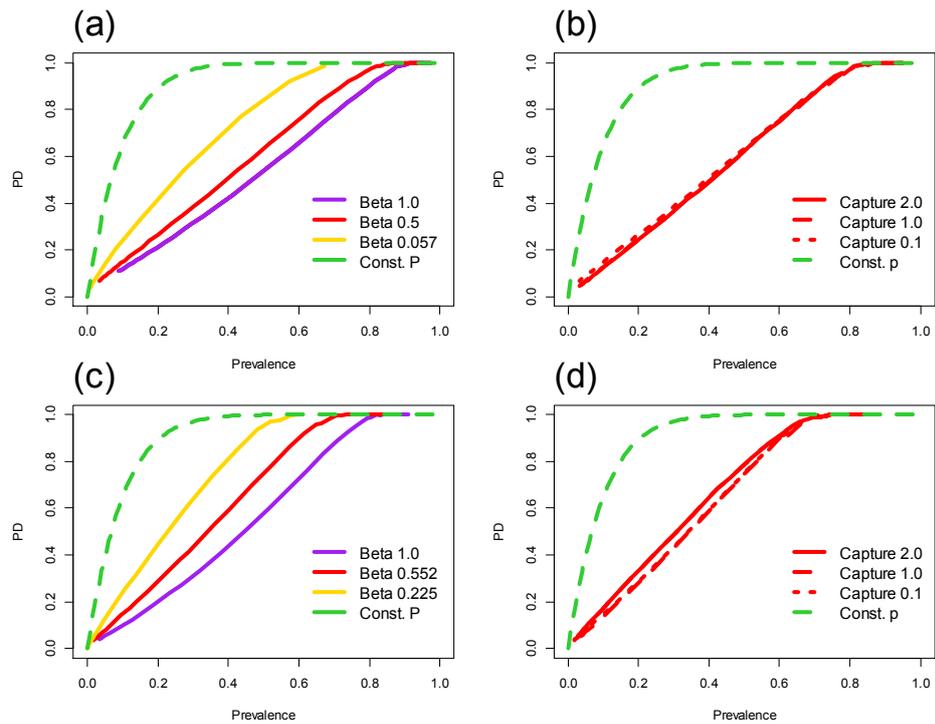


**Figure 3.6: Effect of surveillance design.** In all plots results are shown for surveillance in Rabbits ( $\beta = 1$  &  $\mu = 4.4$ ). Plot 3.6.a shows the expected values of the surveillance estimate of prevalence (purple), the true prevalence (blue) and the continuous sampling theory prediction (black) for a range of values of the capture rate for a fixed sample size  $m = 10$ . Plot 3.6.b shows the expected standard deviation ( $\sigma_p$ ) in the surveillance estimate of prevalence (purple) and the true prevalence (blue) for a range of values of the capture rate for sample size  $m = 10$  for a range of values of the capture rate  $\alpha$ . Results shown are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the parameter values described in Table S.2.8 Appendix 2.2.

Figures 3.5 and 3.6 show similar effects of the capture rate on the estimate of prevalence. Whereby as  $\alpha$  increases, the estimated prevalence moves from the continuous sampling theory to the true underlying prevalence in the system. The results show that less intense effort is needed in the rabbit population to achieve more accurate results (in comparison to the badger population). Given that  $\alpha$  is the *per capita* capture rate this difference is largely explained by the contrasting size of the two host populations, for which a lower capture rate in rabbits yields the required sample size approximately three times faster than in the badger population (which is approximately the ratio of population sizes for these hosts).

### 3.4.2 Detecting disease

In chapter 2 we elaborated on the binomial estimate of the probability of detection,  $PD^{Bin}$ , showing through simulation and analysis that it can over-estimate the probability of detection depending on the characteristics of the host-pathogen system (see Chapter 2 for details). Figure 3.7.a and 3.7.c shows the effect on detection probability of the interaction between two aspects of disease dynamics in badger and rabbit populations respectively, by varying disease induced mortality for three different levels of disease transmission. As in the case of prevalence estimation, conditioned on a given expected prevalence, larger transmission rates  $\beta$  are associated with greater fluctuations in the underlying wildlife disease system which results in reduced probability of detection. Figure 3.7.b and 3.7.d show the effect of capture rate on the probability of detection.



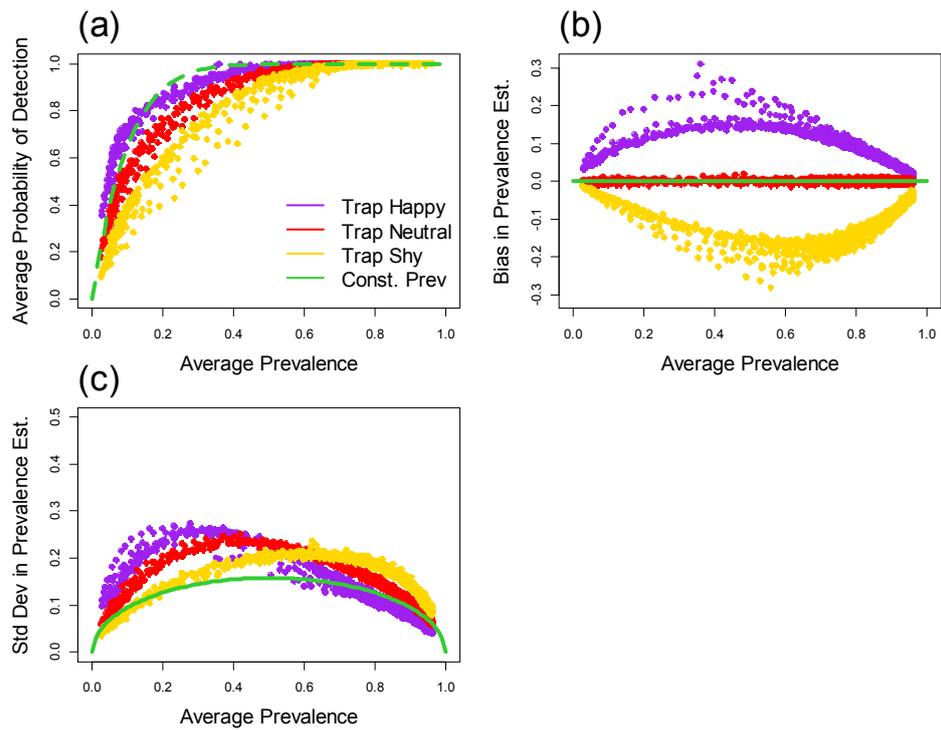
**Figure 3.7: Effect of host-pathogen and surveillance dynamics on the probability of detection.** Plot 3.7.a Probability of detection  $PD$  versus actual prevalence  $E[p]$  for badgers. Plots are generated by varying disease induced death rate as in previous figures with samples size  $m = 10$  and capture rate  $\alpha = 0.1$ . Green – shows the zero fluctuation binomial estimate for all prevalences  $[0,1]$  whilst simulations results accounting for fluctuations are:purple ( $\beta = 1.0$ ); red ( $\beta = 0.5$ ); yellow ( $\beta = 0.057$ ). Plot 3.7.b shows the case  $\beta = 0.5$  but for three capture rates  $\alpha = 0.1, 1.0, 2.0$ . Plot 3.7.c shows the probability of detection  $PD$  versus actual prevalence  $E[p]$  for Rabbits. Plots are generated by varying disease induced death rate as in previous figures with samples size  $m = 10$  and capture rate  $\alpha = 0.1$ . Green – shows the zero fluctuation binomial estimate for all prevalences  $[0,1]$  whilst simulations results accounting for fluctuations are:purple ( $\beta = 1.0$ ); red ( $\beta = 0.552$ ); yellow ( $\beta = 0.225$ ). Plot 3.d shows the case  $\beta = 0.552$  but for three capture rates  $\alpha = 0.1, 1.0, 2.0$ . All simulation results are based on  $10^6$  surveillance bouts using the SDE implementation of the model (see Chapter 2) run using the set of parameter values described in Table S.2.9, S.2.10, S.2.11 and S.2.12 in Appendix 2.2.

Figure 3.7 shows both the prevalence and transmission rate  $\beta$  affect the probability of detection. Disease induced mortality is varied giving rise to the range of prevalence plotted on the x axis. As prevalence increases, so does the probability of detection but at different

rates depending on the transmission rate. We can see that at lower transmission rates surveillance appears to have a better chance of detecting an infected individual for a given expected prevalence level  $E[p]$ . Comparing two transmission rates at the same prevalence the higher transmission rate corresponds to the more unstable infected population, leading to a lower probability of detection. The situation for badgers and TB can be seen on the yellow line in Plot 3.7.a at a prevalence  $E[p]$  of approximately 0.48 (corresponding to a disease induced mortality rate of approximately 0.165 i.e. the estimate for TB in badgers – see Appendix 2.2 for details). This shows that in the two systems studied disease and demographic fluctuations drive a significant reduction in the ability of surveillance to detect disease compared with the prediction from theory ignoring such fluctuations. We note from Figure 3.7 that capture rate barely impacts on the probability of detection, with only a very slight difference seen in the rabbit population which could be attributed to stochastic fluctuations in the simulations.

### 3.4.3 Explicit sources of bias

Above we showed that bias can arise from the interaction between disease dynamics, demographic fluctuations and surveillance even in the absence of any explicit bias in sampling (capture of individuals). However we now consider an explicit source of bias driven by animal behaviour in response to disease status, namely, trappability. For example, it has been debated in the literature as to whether disease affliction will lead badgers to be trap happy or trap shy, that is more or less likely to be captured depending on their disease status (Tuytens *et al.* 1999; Byrne *et al.* 2012a). Figure 3.8 shows the impact that both of these scenarios, as well as a neutral baseline, would have on the probability of detection and the mean and standard deviation of the prevalence estimate. We ran simulations varying both the surveillance capture rate and the transmission rate. This spanned scenarios of low and high transmissibility of disease and low and high effort. In order to ensure that, in the trap neutral example, bias is kept to a minimum, disease induced mortality is fixed at a relatively low level. From here on in the Gillespie implementation of the stochastic model has been used to produce the results shown. As we are exploring specific details of the dynamics of the badger population in terms of individual behaviours and surveillance testing scenarios, the Gillespie algorithm provides a more appropriate representation of stochasticity.



**Figure 3.8: Effect of trappability and surveillance dynamics in Badgers.** Plots are generated by varying the transmission rate  $\beta$  with samples size  $m = 10$  for three trappability levels of infected individuals modelled as different capture rates: yellow (Trap Shy – half less likely to catch an infected badger compared with an uninfected individual); red (Trap Neutral); purple (Trap Happy – twice as likely to catch an infected badger compared with an uninfected); For the same levels of trappability shown in Plot 3.8.a, Plot 3.8.b shows the bias  $E[\hat{p}_{surv}] - E[p(t)]$ , and plot 3.8.c the standard deviation in surveillance estimates (based on a fixed sample size  $m = 10$ ) of prevalence versus the expected value of true disease prevalence in the system,  $E[p(t)]$ . Constant prevalence theory estimates are also shown (green). Results shown are based on 1000 surveillance bouts using the Gillespie implementation of the model (see text) run using the set of parameter values described in Table S.2.13 in Appendix 2.2.

We can see in the graphs above that trap shyness leads to a reduction in the probability of detection as well as a negative bias in estimating the prevalence. Intuitively this relationship makes sense, if individuals are less likely to be trapped because of infection burden then it will be harder to detect the disease and this will also lead to a reduction in the prevalence estimate compared to the true underlying prevalence (hence negative bias). Conversely, trap happiness leads to an increased probability of detection and a positive bias in the

prevalence estimate. The parameterisations used here have been chosen to reduce the effects of fluctuations on the process of surveillance in order to highlight the impact of trappability. However in cases, such as those described earlier, where fluctuations have a greater impact, the effects of trappability will interact with the effects of demographic fluctuations and stochasticity in disease dynamics to determine probabilities of disease detection, and the bias and precision of prevalence estimates obtained via surveillance. Even in the results shown here the effect of population fluctuations on the standard deviation of the prevalence estimate can be seen for the neutral case where, as shown earlier, precision is reduced compared with standard theory. However, these effects are modified by animal behaviour. Trap shyness increases precision (reduces the standard deviation of the estimate) at low prevalence and reduces precision at high prevalence. In contrast, trap happiness on the part of infected individuals has the opposite effects i.e. precision is reduced at low and increased at high levels of prevalence.

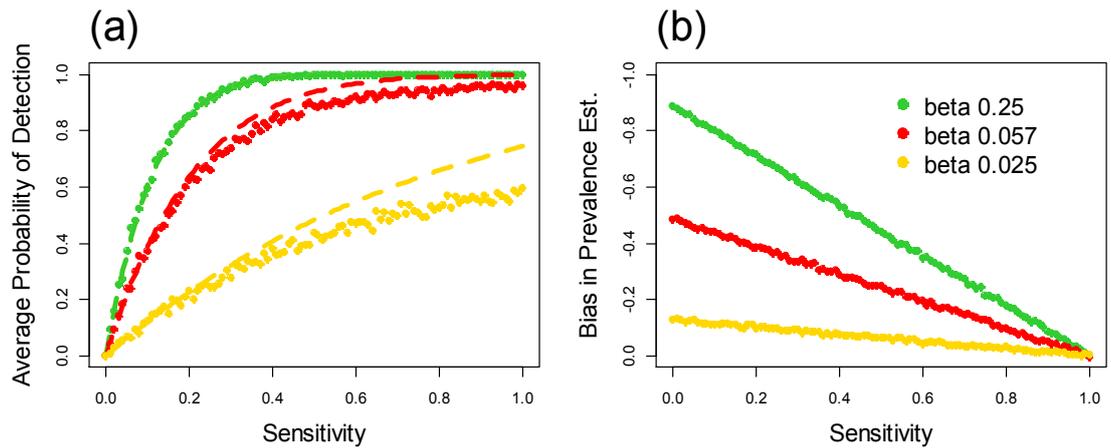
#### 3.4.4 Imperfect diagnostic tests

Our previous explorations of the effects of fluctuations and surveillance strategy on the efficacy of surveillance have assumed perfect diagnostic tests (i.e. sensitivity and specificity of the test = 1.0). It is worth recalling the definitions of both sensitivity and specificity of the test, namely

*Sensitivity = probability of detecting disease in an individual that is truly infected*

*Specificity = probability of a negative test in an individual that is truly uninfected*

Here we focus on the issue of false negatives, so consider sensitivities less than 1, whilst keep specificity = 1. In Figure 3.9 we demonstrate how the sensitivity of the test can impact on both the probability of detection and the bias in the prevalence estimate for a badger population. All parameters are held fixed at three different transmission rates (0.25, 0.057, 0.025) with disease induced mortality = 0.165 whilst the sensitivity of the test is varied and specificity is constant at 1.0.



**Figure 3.9: Effect of test sensitivity and disease dynamics in Badgers.** 3.9.a shows the probability of detection  $PD$  versus sensitivity of the diagnostic test for Badgers. Plots are generated by varying the sensitivity of the tests used in surveillance with sample size  $m = 10$  and capture rate  $\alpha = 0.5$  and  $\mu_I = 0.165$  for three transmission rates  $\beta$ : yellow (0.025); red (0.057); green (0.25) with associated prevalence of approximately 0.127, 0.483 and 0.887 respectively. In each case simulation results are shown by dots and theory based on constant prevalence by dotted lines. Plot 3.9.b shows the bias  $E[p(t)] - E[\hat{p}_{surv}]$  in surveillance estimates (based on a fixed sample size  $m = 10$ ) of prevalence versus the expected value of true disease prevalence in the system,  $E[p(t)]$ . Data are shown for three different levels of transmission rate  $\beta$  as in 3.8.a. Results shown are based on 1000 surveillance bouts using the Gillespie implementation of the model (see text) run using the set of parameter values described in Table S.2.14 Appendix 2.2.

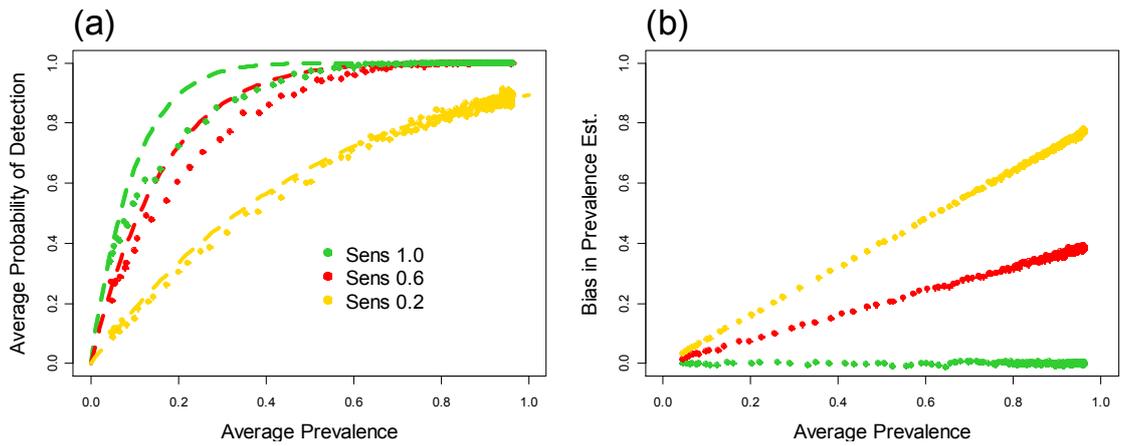
Figure 3.9 shows that, scenarios in which the lack of a sensitive test can mask the effect of fluctuations in demography and disease dynamics on the probability of detection. In the low transmission rate examples shown here as the sensitivity of the test increases the discrepancy between the true probability of detection and the prediction based on assuming constant prevalence increases. In these cases as test sensitivity is increased the supposed gains in ability to detect disease predicted by binomial theory assuming constant prevalence are not realised. We can see for  $\beta = 0.025$  that even though the bias in prevalence estimate is highest at lower sensitivity levels, this is when the probability of detection follows the constant prevalence curve closely. It is only when sensitivity levels are increased, and bias in the prevalence estimate reduced, that there is a difference between the detection rate predicted by constant prevalence theory and the true (simulated) detection rate. It is important to note that the probability of detection does improve with increased sensitivity, but not at the same rate as would have been expected if fluctuations

are ignored. For the chosen disease induced mortality rate, the three values of the transmission rate correspond to low medium and high prevalence. Fluctuations have a small effect on detection rates at high prevalence and an increasingly large effect as prevalence decreases. This is why the above noted effect of sensitivity masking errors in the constant prevalence theory estimates of the probability of detection are most evident for lower values of  $\beta$ .

The bias curves plotted in Figures 3.9 and 3.10 show  $E[p(t)] - E[\hat{p}_{surv}]$  and can largely be explained by reference to constant prevalence theory (note here we have sought examples where the role of fluctuations is minimised). Consider a case where the prevalence is constant at a value of  $E[p(t)]$  and the test sensitivity is given by  $Se$ . Then the expected value of the surveillance estimate of prevalence is  $Se * E[p(t)]$  and

$$Bias = E[p(t)] - Se * E[p(t)]$$

The linear form of this equation is seen in the results plotted in Figures 3.9.b and 3.10.b. For example, when the sensitivity  $Se = 1$ , the bias is zero as shown in Figures 3.9.b. Moreover, when  $Se = 0$  the bias plotted is simply the underlying prevalence  $E[p(t)]$ . Figure 3.10 explores the sensitivity of the test for varying prevalence; this demonstrates how disease dynamics can interact with test sensitivity in determining surveillance efficacy. As prevalence increases bias increases linearly as predicted by the above linear equation.



**Figure 3.10: Effect of disease dynamics and test sensitivity in Badgers** Plot 3.10.a shows the probability of detection  $PD$  versus actual prevalence  $E[p]$  for Badgers. Plots are generated by varying  $\beta$  with samples size  $m = 10$  and  $\alpha = 0.5$  and  $\mu_I = 0.165$  for three sensitivity levels: yellow (0.2); red (0.6); green (1.0). Plot 3.10.b shows the bias  $E[p(t)] - E[\hat{p}_{surv}]$ , based on a fixed sample size  $m = 10$ , of prevalence versus the expected value of true disease prevalence in the system,  $E[p(t)]$ . Data are shown for three different sensitivity levels as in 3.10.a. Results shown are based on 1000 surveillance bouts using the Gillespie implementation of the model (see text) run using the set of parameter values described in Table S.2.15 in Appendix 2.2.

Figure 3.10 shows that the bias in prevalence estimate increases as the prevalence increases for sensitivity levels  $< 1.0$  and the lower the sensitivity level, the greater the increase in bias as prevalence goes up. At any sensitivity  $< 1.0$  there is a percentage of test results that return as a false negative. This percentage also depends on how many infected individuals there are available to be diagnosed as false positive. As prevalence increases, the amount of false positive results also increases, hence the bias in prevalence estimate increases as the results get less accurate. This is consistent with binomial theory that states if true prevalence =  $p$  and sensitivity is  $se$  with specificity = 1 then estimated prevalence =  $p \cdot se$  and thus bias =  $p - p \cdot se$ . The probability of detection is seen to be depressed away from the binomial theory estimate for higher sensitivities (especially sensitivity 1.0) which has been shown in Chapter 2. However for lower sensitivity levels, the probability of detection seems to follow the binomial line very well. Thus as noted above one effect of fluctuations is that improvements in test sensitivity may not lead to the gains in detection rates predicted by constant prevalence theory. Moreover the effects of fluctuations on disease surveillance, which are the focus of this chapter and Chapter 2, may be masked by relatively poor tests with low sensitivity.

### 3.5 Discussion

In Chapter 2 we characterised surveillance of wildlife disease in terms of a handful of parameters describing the key components of such systems, namely host demography, disease dynamics and the nature of the surveillance itself. A simple but generic stochastic modelling framework was used to study interactions between these elements in homogeneous/well mixed populations. The results obtained from analysis and simulation of this model showed that combined fluctuations in population and disease dynamics bias the surveillance estimate of prevalence with respect to expected true prevalence of the underlying wildlife disease system. In addition the precision of these estimates and the probability that surveillance will detect disease are both lower than would be expected based on standard arguments which ignore such fluctuations. Thus our results suggest that power calculations or analyses which ignore fluctuations may lead to under powered wildlife disease surveillance programmes and/or over confidence in the results obtained. Such results would lead to incorrect characterisation of the risks posed by a given wildlife disease system.

In this chapter we have shown how the theoretical framework and models presented in Chapter 2 can be further developed and applied as tools to assess the outcomes of wildlife disease surveillance in real systems. Chapter 2 explored a broad range of wildlife disease systems i.e. a range of host-pathogen combinations where as in this chapter we asked if such effects are likely to be a feature of particular wildlife disease systems. Our aim was to build on the theoretical developments presented in Chapter 2 by parameterising stochastic models to represent two natural wildlife host populations (badgers and rabbits) for a range of disease dynamics (including TB and paraTB respectively). In so doing we have quantified the effects seen in Chapter 2 for these wildlife disease systems and can conclude that they are of practical concern for real systems and a range of diseases including the two specific examples which are of current interest. For example, such effects potentially undermine management strategies for TB in badgers because current surveillance designs may be under powered leading to poor characterisation of the risk posed to sympatric cattle populations.

In addressing the above wildlife disease systems we have expanded our theoretical framework to accommodate sources of bias related to aspects of the disease (i.e. disease induced mortality), diagnostic tests (reduced test sensitivity) and host behaviour (biases in trappability). The research presented in this chapter characterised and quantified the

impact of these currently acknowledged sources of bias in the context of wildlife disease systems subject to demographic and disease related fluctuations. The fluctuations demonstrated in badger and rabbit populations are not at the extreme end of the scale, and could be viewed as similar to the fluctuations found in livestock and managed populations. The results shown have the potential to be amplified when considering populations with larger and more extreme population fluctuations which could be demonstrated in source-sink scenario.

The effects of disease induced mortality are complicated by the fact higher rates depress both prevalence and size of infected populations. This alters the stability of both the disease and the populations and the resulting fluctuations reduce the efficacy of surveillance. For a particular wildlife host (identified by birth and background/natural mortality rates), a given level of prevalence results from the interplay between disease transmission and disease induced mortality. When both are relatively high resulting fluctuations in demography and disease dynamics can lead to biased and low precision estimates of prevalence and reduced rate of disease detection from surveillance programmes. This may, for example, be a feature of Rabbit Haemorrhagic Disease which is documented to have both a large disease induced mortality rate and be highly transmissible (Cooke 2002; Calvete 2006).

Bias in the trappability of an infected host, within different surveillance and disease prevalence scenarios, increases the bias seen in the prevalence estimate. This can happen in two ways; there will be a tendency towards positive bias (i.e. the surveillance over-estimates the prevalence) if individuals are trap happy. There are examples in the literature whereby animals become more likely to enter traps as they become weaker with the burden of disease and the food often used to bait traps becomes more attractive (Tuytens *et al.* 1999; Coltherd *et al.* 2010; Byrne *et al.* 2012a) . Conversely, there will be a tendency towards negative bias (i.e. the surveillance under-estimates the prevalence) if the individuals are trap shy (Tuytens *et al.* 1999; Wilkinson *et al.* 2000). For example TB can affect social aspects of badger behaviour and infected individuals become more isolated from the social group which may make it harder for surveillance to capture them if the effort is concentrated on the main sett (as is often the case). This relationship between trappability and bias is intuitive as, taking the case of positive bias, infected animals are more likely to be trapped and there will be a disproportionate number of infected individuals being caught by surveillance leading to an over-estimate of prevalence. The error (standard deviation) associated with prevalence estimates is also affected by

trappability status in the infected population. Standard deviation is reduced at low prevalence for trap shy individuals and increased at high prevalence compared to the neutral case. The converse is true for trap happy individuals. The probability of detection decreased for infected individuals exhibiting trap shy behaviour, and this is to be expected as a reduction in the likelihood of an infected individual being caught will lower the chance infection will be detected. The converse, again, is true for trap happy individuals. This explicit source of bias has the potential to counteract or inflate the impacts on surveillance induced by population fluctuations which we studied above and in Chapter 2.

The effect of sensitivity on the prevalence estimate is well described by standard binomial theory which ignores fluctuations. Surveillance detection rates rise with sensitivity and bias in the surveillance estimates of prevalence reduces. In the presence of fluctuations, at lower sensitivity levels the true probability of detections is quite accurately predicted by the constant prevalence estimate. However, at higher sensitivity levels the probability of detection is lower than the constant prevalence theory prediction (previously we had examined perfect high sensitivity tests). As the sensitivity of the test increases, the probability of detection from surveillance is lower than the constant prevalence theory predicts, and this is particularly marked at lower transmission rates (i.e. at lower prevalence's, all else being equal). For the case of perfect tests the message from both this chapter and Chapter 2 was that fluctuations in prevalence reduce the true probability of detection relative to the constant prevalence theory prediction. However, here we see that such effects may be masked by imperfect tests, such that they are not apparent at low sensitivities but may become discernible as sensitivity increases. Therefore as better tests become available the effects of fluctuations on wildlife disease surveillance described in Chapter 2 (and above) may be more noticeable in practice. If predictions for the improvement of disease detection rates (probability of detection) with enhanced (more sensitive) tests are based on constant prevalence theory then our results suggest that actual gains in the performance of surveillance may not be as great as predicted.

In summary we have demonstrated that the results of surveillance are determined by complex interactions between surveillance strategy, host demography and pathogen transmission dynamics. We have also explored the impact of a range of aspects of the ecology of wildlife disease (the impact of the disease on host mortality and behaviour) and the tools used for surveillance (sensitivity of diagnostic tests). In so doing we have illustrated how the models and framework we introduced in Chapter 2 and extended here

might be used as computational tools to design better surveillance programmes which take account of ecological and other effects. For example, rather than rely on binomial type calculations based on assuming fixed population size and prevalence, surveillance design for specific systems could be based on the simulated outputs from models like those used here if they can be suitably tailored to the wildlife disease system of interest. Moreover in this chapter and the previous one we have shown how to explore the efficacy of surveillance for a range of population and disease dynamics. The information obtained from these more generic studies could be used to inform surveillance when knowledge of host demography and/or pathogen characteristics is uncertain, or when the goal is to design generalised surveillance suitable for a range of hosts and pathogens.

The work of this and the previous chapter has established the role of temporal heterogeneities induced by stochasticity in well mixed spatially homogeneous local populations. However, the impact of spatial heterogeneity on the spread and persistence of disease has been well documented in the literature and given the results described here this is likely to have significant impacts on surveillance in wildlife. In the next chapter we will turn our attention to spatially extended meta-populations and the impact of spatial heterogeneity on the design and efficacy of surveillance.

## Chapter 4

Spatio-Temporal modelling of habitat composition, ecology and wildlife disease surveillance: impacts of emergent disease.

## 4.1 Abstract

Spatial heterogeneity is known to be an influential factor in the spread and persistence of disease suggesting it is likely to impact on the efficacy of surveillance. The generic stochastic modelling framework introduced in Chapter 2 and further developed in Chapter 3 was extended to incorporate spatial heterogeneity. In order to study surveillance over an extended region we developed a stochastic and spatially explicit meta-population model that describes demography, disease transmission and key aspects of surveillance. The impact of various aspects of design on the ability of spatially distributed surveillance networks to detect emergent disease at a regional scale is assessed by the level of disease present in the system at the point of first detection. In particular we use the extended framework to explore key spatial aspects of surveillance design including within location effort, number of surveillance locations, and choice of such locations according to habitat suitability for the host species. In line with current practice we find that increasing effort/number of locations under surveillance and spatial stratification according to the habitat suitability improves the ability of surveillance networks to detect emerging outbreaks. In addition we evaluate dynamic designs and find that, given the ability to conduct surveillance in a set number of locations, switching effort between locations is expected to lead to more rapid detection of disease than static designs. We conclude that spatial heterogeneity, disease dynamics and surveillance strategy should all be considered when assessing the ability of surveillance to detect emerging disease risk in a spatially extended wildlife population.

## 4.2 Introduction

In Chapters 2 and 3, it was shown that natural fluctuations in a target host population drive the efficacy of wildlife disease surveillance in terms of both estimating prevalence and the probability of detection. In many disease systems (host and pathogen combinations) fluctuations in host population size and disease prevalence lead to bias and reduced precision of estimates of prevalence and also lower the probability of disease detection when compared to standard binomial theory which ignores such effects. These results suggest that current surveillance programmes may be underpowered if designed using power calculations based on binomial theory. The

impact of the design of surveillance strategies (i.e. sample size and capture rate/effort) was also shown to impact surveillance efficacy in ways not predicted by standard theory. In particular capture rate can minimise bias in prevalence estimates and increasing sample size does not increase the probability of disease detection or the precision of prevalence estimates as fast as would be expected from standard arguments. These results were all shown in a non-spatial context and for endemic (stable and unstable) disease scenarios. However, there are a range of spatial effects that have been shown to impact on both the host population size and stability and on disease persistence and spread e.g. intrinsic and extrinsic spatial heterogeneity driven by stochasticity and habitat composition (Tilman and Karieva, 1997; Keeling, 1999; Keeling et al 2001). In this chapter we aim to explore how spatial heterogeneity affects surveillance efficacy in the context of emerging infectious diseases.

Emerging disease outbreaks are of particular importance when considering surveillance efficacy in a spatial context. Epidemiological modelling has shown that increased contact among populations can trigger epidemics (Hess 1996) and spatial composition of the environment (habitat) influences emergent disease risk (Ostfeld *et al.* 2005). Emerging (and re-emerging) disease outbreaks have the potential to be detrimental to wildlife (Bengis *et al.* 2004), including conservation efforts (Daszak *et al.* 1999, 2000), as well as livestock (Rhyan & Spraker 2010) and human health (Epstein 1995; Murphy 2008). In an emerging disease situation e.g. a disease incursion to a naïve population, the main goal of surveillance is to detect the disease as quickly as possible to limit spatial spread and risks to wildlife, livestock and human health. In order to do this effectively the premise of this chapter is that it is important to consider the spatial composition and structure of the environment (habitat quality, distribution, connectivity etc.) and the demography (births, deaths, immigration etc.) of the host population and how these are likely to affect the efficacy of surveillance.

It has been repeatedly shown that spatial heterogeneity impacts on the spread and persistence of disease (Sattenspiel & Simon 1988; Cliff 1995; Keeling & Grenfell 1998). Spatial effects have been found to be important in many types of systems, for example crops (Antle *et al.* 2003), human populations (Smith *et al.* 2002) and predator-prey systems (Hastings 1977) etc. In this chapter we will focus on spatial effects in wildlife meta-populations. However, there is little literature addressing how intrinsic and extrinsic spatial heterogeneities would impact on the ability to carry out

effective surveillance and on the reliability of the results obtained from such efforts. Intrinsic spatial heterogeneity is driven by the dynamics of the population and is an emergent property of the system. Extrinsic spatial heterogeneities are driven by external conditions to which the population responds, such as habitat suitability. In terms of surveillance design, efforts to account for extrinsic spatial heterogeneity have been made, for example tailoring surveillance to knowledge of wildlife habitat locations (i.e. extrinsic heterogeneity) (Walsh & Miller 2010). However, the effects of intrinsic spatial heterogeneity on surveillance have yet to be explored.

In this chapter we will explore how the interaction between both intrinsic and extrinsic heterogeneity affect surveillance efficacy within an emerging infectious disease context. We aim to assess how the spatial distribution of habitat quality, disease dynamics and surveillance strategy can impact the amount of disease present within the system, at the point of first detection following an incursion event. In doing so the aim is to develop insights, general recommendations and tools which contribute to the design of better strategies for wildlife disease surveillance compared with current understanding and practice.

## 4.3 Methods

### 4.3.1 Stochastic Model Description

The model represents a finite space, on which  $L$  so called patches are randomly distributed. Each patch supports a host population subject to demographic fluctuations (births, deaths and immigration) and the transmission of a single pathogen. At each point in time  $t$  and for each patch  $i=1,\dots,L$ , the state-space represents the total population size  $N_i(t)$ , with  $I_i(t)$  of these infected and the remainder,  $S_i(t) = N_i(t) - I_i(t)$  susceptible. In addition the prevalence in patch  $i$  is then given by  $p_i(t) = I_i(t)/N_i(t)$ . The birth rate of individuals is logistic,  $rN_i(1 - N_i/k_i)$ , with intrinsic growth rate  $r$  and a carrying capacity  $k_i$  reflecting the assumptions that population growth is resource limited. The carrying capacity  $k_i$  is determined by the suitability rating,  $\sigma_i$ , assigned to each patch multiplied by a maximum potential carrying capacity equal over all patches  $k_{max}$  i.e.  $k_i = \sigma_i k_{max}$ . The suitability rating (or level)  $\sigma_i$  represents how suitable the habitat is on patch  $i$  for the population and ranges from 0 to 1. Favourable patches are defined as those where

$\sigma = 1$  and the carrying capacity is  $k_{max}$ , whereas less favourable patches refer to locations where  $\sigma < 1$  and the carrying capacity is less than  $k_{max}$ . It is therefore meaningful to talk about varying the suitability of less favourable patches in the interval  $[0,1)$ . In the limit  $\sigma \rightarrow 1$  less favourable patches become/would be classified as favourable. Individuals have a per capita death rate  $\mu$  and secondary transmission within patch (contact with already infected individuals from the population) occurs at rate  $\beta_0 S_i(t) I_i(t)$ .

Immigration into each patch occurs at a constant rate  $\nu$  and a proportion,  $\gamma$ , of immigrants are infected, but otherwise all individuals enter the population (through birth or immigration) as susceptible as there is no vertical or pseudo-vertical transmission. As well as within patch secondary transmission described above, susceptible individuals become infected through between patch secondary transmission. Between patch transmission is defined as an infected individual transmitting disease to a susceptible individual in a different patch. The distance by which the disease can transmit from one patch to another is controlled by the distance kernel, i.e. the probability of disease transmitting between patches decreases as the distance between them increases. Although other forms of distance dependence e.g. power-law could easily be accommodated, in the model considered here between patch transmission is governed by an exponential kernel  $e^{-\theta d_{kj}}$  where  $d_{kj}$  is the Euclidean distance between infected individuals in patch  $k$  and the susceptible individuals in patch  $j$  and  $\theta$  controls the rate at which infection pressure decays with this distance. This means that at time  $t$  the rate at which new infections occur in patch  $j$  due to infection by the  $I_k(t)$  infected individuals in patch  $k$ , given there are  $S_j(t)$  susceptible individuals in patch  $j$ , is

$$\beta_1 S_j(t) I_k(t) e^{-\theta d_{jk}}$$

In the spatial context, surveillance is characterised by the set of locations (patches) currently under surveillance, the effort applied in each of these patches and the manner in which surveillance can switch from one location or set of locations to another. Within patches currently under surveillance we adopt the model used in earlier chapters, namely that disease surveillance is incorporated into the model in the form of capture, testing and release at per capita rate  $\alpha$ , for both susceptible and infected individuals. All surveillance testing is undertaken assuming perfect tests, which means that our measures of the performance of surveillance reflect a best case

scenario. The set of patches where this capture rate is applied is determined by random allocation of a proportion of patches where surveillance is activated at any one time. In some cases this allocation will be weighted according to habitat quality as described in the text e.g. when we only sample favourable habitats (see above). As indicated in some cases the set of patches under surveillance is allowed to evolve over time. When this is allowed an initial set of patches is chosen as described above and then this initial set remains under surveillance for a period of time after which surveillance is ended in that set and started in another set currently not under surveillance but otherwise chosen at random using criterion consistent with those used to identify the initial set of patches under surveillance. In the model used here we assume a fixed time to switching for surveillance sets. In the text below we describe this interchangeably as surveillance switching rates and times.

#### 4.3.2 Model Implementation

The model is implemented as a continuous-time discrete-state space Markov process (see Chapter 2), which is simulated using Gillespie's exact algorithm (Gillespie, 1976). In contrast to the stochastic differential equation approach predominantly used in the previous two chapters, the Gillespie algorithm is computationally more intensive but exact. The discrete nature of the state-space under the Gillespie algorithm is a more direct implementation of the model described above, and as discussed previously it therefore represents a more natural description of the population and the processes that affect it. Moreover, it provides a more accurate representation of population dynamics for populations which are important when considering spatial heterogeneity of meta-populations (Keeling & Ross 2008). All events with their corresponding rate and effect on population size for the Gillespie implementation can be seen in Table 4.1 below.

Event	Rate	Effect
<b>Birth</b>	$rN_i(1 - N_i/k)$	$S_i \rightarrow S_i + 1$
<b>Death of Susceptible</b>	$\mu S_i$	$S_i \rightarrow S_i - 1$
<b>Death of Infected</b>	$\mu I_i$	$I_i \rightarrow I_i - 1$
<b>Susceptible Immigration</b>	$(1 - \gamma) \nu$	$S_i \rightarrow S_i + 1$
<b>Infected Immigration</b>	$\gamma \nu$	$I_i \rightarrow I_i + 1$
<b>Between Patch Transmission</b>	$\beta_1 I_j S_i e^{-\theta d_{jk}}$	$S_i \rightarrow S_i - 1$ $I_i \rightarrow I_i + 1$
<b>Within Patch Transmission</b>	$\beta_0 I_i S_i$	$S_i \rightarrow S_i - 1$ $I_i \rightarrow I_i + 1$
<b>Susceptible Capture and Release</b>	$\alpha S_i$	$S_i \rightarrow S_i$
<b>Infected Capture and Release</b>	$\alpha I_i$	$I_i \rightarrow I_i$

**Table 4.1: Event, Rate and Effect on the State Space of the model.** Conceptually the effect of each event affects an individual and this is reflected in the discrete nature of the corresponding changes in the state space. However, given this underlying conception of the model there are a number of different implementations which can be considered including via the Gillespie algorithm and stochastic differential equations (see text for details).

In Chapter 2 we explored a range of wildlife host species and their pathogens and in Chapter 3 we focused on two exemplar host species with contrasting population dynamics. In both chapters we showed that temporal heterogeneities induced by stochasticity in demographic and disease dynamics impacted on the efficacy of wildlife disease surveillance (see above for details). Since such stochasticity is known to play an important role in spatial systems (see e.g. Tilman and Kareiva, 1997) we anticipate that these impacts on surveillance will also be apparent in spatially extended systems.

In this chapter we therefore focus our efforts on studying the impacts on surveillance of spatial (rather than temporal) heterogeneities. To do so we parameterise the model to represent a relatively stable host population. A range of simulations were then run to assess the impact of surveillance design i.e., targeting surveillance on different habitat types, the number of patches under surveillance, the effort applied in each patch and different switching rates. In addition the structure of habitat composition within the modelled area was explored by varying both the spatial habitat composition (varying the ratio of favourable to less favourable habitat type

patches in the space) and habitat suitability (varying the suitability index of less favourable habitat type from 0 to 1). Simulations were also run to explore the impact of within and between patch transmission rates. The results from all these simulations are described in detail below. The parameterisation of the model and details of the simulations run can be found in the relevant tables in Appendix 3 (indicated in Figure descriptions in the results section).

#### 4.3.3 Statistics generated from the model

In this chapter we are focussed on emergent disease and therefore we do not collect long term averages, but rather focus on *ensemble expectations* (approximated by averages over many realisations of the process) of *out of equilibrium* incursion events corresponding to the introduction of the disease and its spatial spread up to the point where it is detected by surveillance. Each simulation is run for a period of time to allow the population to reach equilibrium before a disease incursion is introduced into a random patch. A surveillance bout continues until a detection event occurs or until a maximum time,  $t_{\max}$ , is reached. Statistics are calculated and averaged over the many realisations in which disease detection occurred. The calculations include; average number of infected at time of detection, the average number infected in infected patches, the average time of detection and the number of infected patches.

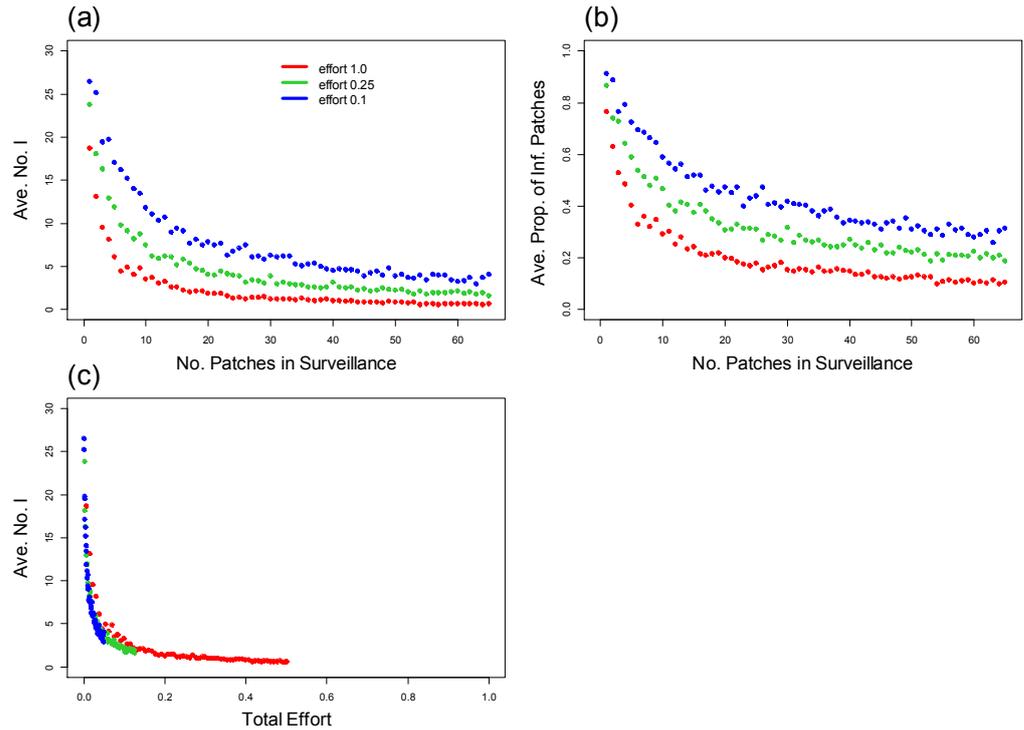
Here a surveillance bout corresponds to a period starting with a disease incursion event where we conduct surveillance until an infected individual is detected, or until some large upper time limit is reached (this just covers cases where disease dies out prior to detection). Individuals within patches under surveillance are captured at per capita rate  $\alpha$ , and patches under surveillance are switched at switching rate  $\tau$ , as described in detail above. The total number of infected individuals caught per patch are recorded until one of the totals reaches 1, i.e. all totals were previously 0 and this is the first infected individual to be caught in any patch. At this point, disease has been detected for the first time. Note this could be easily extended to account for imperfect disease diagnostics by recording the number testing positive. When surveillance ends, either because a detection event has occurred or because a time limit has been reached, the number of infected individuals and patches is recorded. In addition, the time taken for disease to be detected is also recorded. Therefore over repeated surveillance bouts it is straightforward to estimate the average amount of disease present in the spatial system at the point of first detection.

## 4.4 Results

### 4.4.1 Static surveillance: spatial distribution of effort

When disease first enters a naive population, surveillance aims to detect the disease before a significant risk has developed. The risk associated with a given level of spread will depend on factors including host species under surveillance, potential host range, pathogen virulence, the economic and cultural values attached to these species and many others. We measure how well surveillance performs by calculating the amount of disease in the space at the point of first detection. For any given wildlife disease system (host and pathogen characteristics) the way surveillance is distributed within the spatial system, and how much surveillance is undertaken, both in breadth of patches covered and within patch intensity (capture rate), will affect the amount of disease present at the point of first detection. Figure 4.1 demonstrates how the amount of disease present in the spatially extended system reduces at the point of first detection with increased number of patches concurrently included in the surveillance sample at three different capture rates. To explore the effect of the distribution of surveillance effort (i.e. the difference between spreading surveillance thinly with low effort compared to concentrating high effort on a small number of patches) the total effort was calculated as follows:

$$\textit{Total Effort} = \frac{\textit{Number of Patches in Surveillance}}{\textit{Total Number of Patches}} \times \textit{Capture Rate}$$

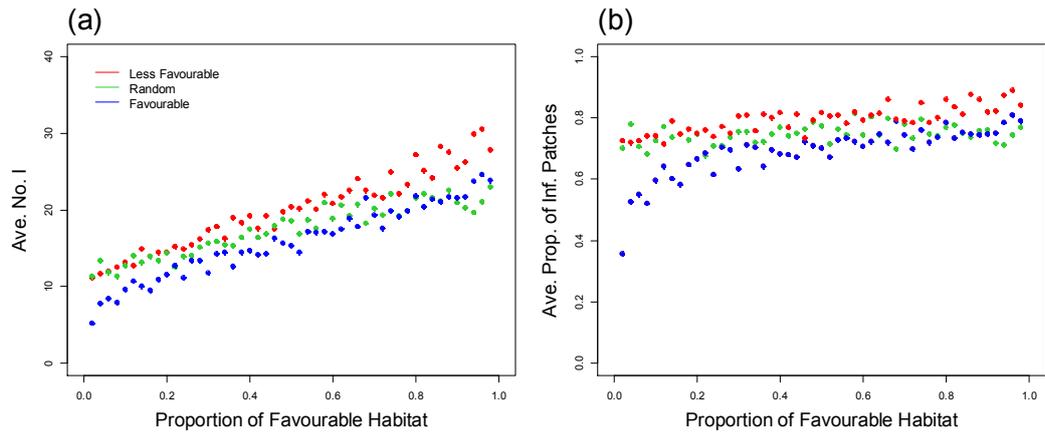


**Figure 4.1: Effect of the distribution of surveillance effort between patches.** Data are shown for three different values of the capture rate  $\alpha$  (1.0-red,0.25-green,0.1-blue) each for a varying range from 1 to 65 patches under surveillance (with no switching). Plot 4.1.a shows that the average number of infected individuals and plot 4.1.b shows the average number of patches with infection both decrease as the number of patches in the surveillance sample increases. This improvement happens more quickly as effort increases (capture rate). Plot 4.1.c shows the average number of infected individuals for the “total effort” from the simulation values used in Plot 4.1.a. The plots shown were run using the set of parameter values described in Table S.3.1 in Appendix 3.1.

Figure 4.1 shows, as the number of patches included in the surveillance sample increases, the less disease is present on average at time of first detection. This is an intuitive result, as the more space that can be covered in surveillance, the higher the chance of detecting disease. However, these results also indicate that, somewhat less intuitively, the distribution of surveillance effort between patches does not impact significantly on the average size of the disease incursion at time of first detection.

#### 4.4.2 Static surveillance: stratified designs

Environmental factors, such as habitat suitability, can also influence the efficacy of surveillance. As discussed previously, habitat suitability is modelled in terms of the carrying capacity of each patch. This is characterised by a suitability level  $\sigma_i$  which ranges from zero to 1 and interpolates linearly between the carrying capacity of favourable patches and zero. In this chapter we consider two habitat types at any one time. Favourable habitat has the maximum possible carrying capacity  $\sigma_i=1$  and less favourable habitat has  $\sigma_i < 1$  i.e. carrying capacity ranges from 0 to that of the favourable habitat. As previously discussed in the introduction, habitat distribution within the spatial system is integral to the spread and persistence of disease. Figure 4.2 demonstrates the effect, on the amount of disease present at first detection, of varying proportions of favourable to less favourable habitat at three different capture rates for 1% patch surveillance set. Here less favourable habitat has a suitability index of  $\sigma = 0.5$ . Three different surveillance strategies are also compared: 1) only selecting more favourable habitat for the surveillance sample; 2) only selecting less favourable; and 3) selecting both with equal probability.

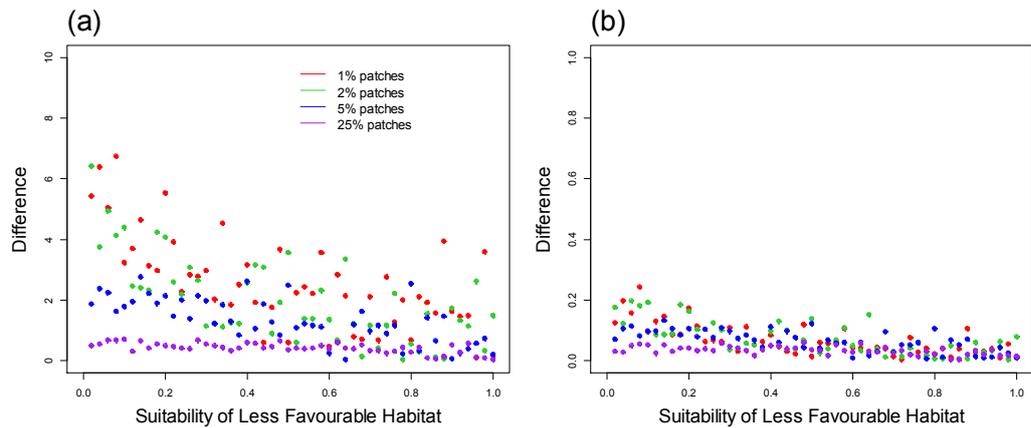


**Figure 4.2: The effect of habitat quality on stratified sampling schemes.** Amount of disease present at time of first detection for three differently stratified sampling strategies (surveillance in: Less favourable only – red; Favourable and less favourable equally – green; and Favourable only - blue) as a function of the proportion of favourable habitat. Data are shown for effort level 1.0 in 1% of patches with a random disease incursion relative to the proportion of favourable and less favourable habitat in (a) the average number of infected individuals in the spatial system at the time of first detection and (b) the corresponding figure for the average number of infected patches. The plots shown were run using the set of parameter values described in Tables S.3.2 in Appendix 3.1.

Figure 4.2 shows that as the proportion of favourable habitat increases the amount of disease in the system at first detection also increases, regardless of sampling strategy. This simply reflects the fact that the system as a whole is more susceptible to disease as there are more hosts to infect when the proportion of good quality habitat increases. In comparing the three sampling strategies Figure 4.2 shows that sampling only less favourable habitat patches leads to the highest level of disease in the system at the time of first detection. In contrast sampling only favourable habitat patches is the best performing strategy while the mixed strategy falls between the other two. When the proportion of favourable habitat patches approaches 1 the mixed strategy is equivalent to sampling only favourable habitat because there are fewer and fewer less favourable habitat patches. Similarly when the proportion of good habitat patches approaches zero the mixed strategy of sampling both habitat type patches is equivalent to sampling only less favourable patches (because, in the limit, there are only poor quality patches).

Sampling good quality habitat patches is the superior strategy because hosts are more abundant in favourable habitat which is therefore more favourable for the pathogen and more likely to shelter infected individuals who can be sampled by surveillance. It is also evident from Figure 4.2 that this effect becomes more pronounced as the proportion of good quality habitat patches shrinks. Initial disease incursion events occur into a randomly chosen patch and therefore when favourable habitat is rare incursion is most likely to occur in a poor quality patch. In such circumstances disease is less likely to spread far, but sampling in only less favourable habitat is most likely to detect disease. However, when disease incursion does (rarely) happen in a good quality patch not only will it be initially missed less favourable habitat sampling it is also likely to be a larger outbreak. This inflates the average outbreak size under this (and the mixed) sampling strategy. In contrast, in the limit of a low proportion of favourable habitat, when sampling in favourable only habitat the majority of outbreaks are small (because most patches are poor) but remain undetected. However, surveillance will detect disease quickly when an initial incursion occurs in, or an outbreak reaches, a favourable patch. In this case the size of outbreak on detection is likely to be small. These effects are amplified as the proportion of favourable patches shrinks and thus it makes most sense to develop stratified surveillance designs (based on habitat quality) when good quality habitat is rare.

Figure 4.3 demonstrates the effect of varying the suitability level of the less favourable habitat on the amount of disease in the system at the point of first detection, when the proportion of suitable and less suitable patches are equal i.e., 0.5. We consider 4 levels of patch inclusion in surveillance ranging from 1% of patches to 25%. How hostile the environment is for the host species will not only impact on the population size but also on the disease and its ability to spread. This has implications for surveillance efficacy.

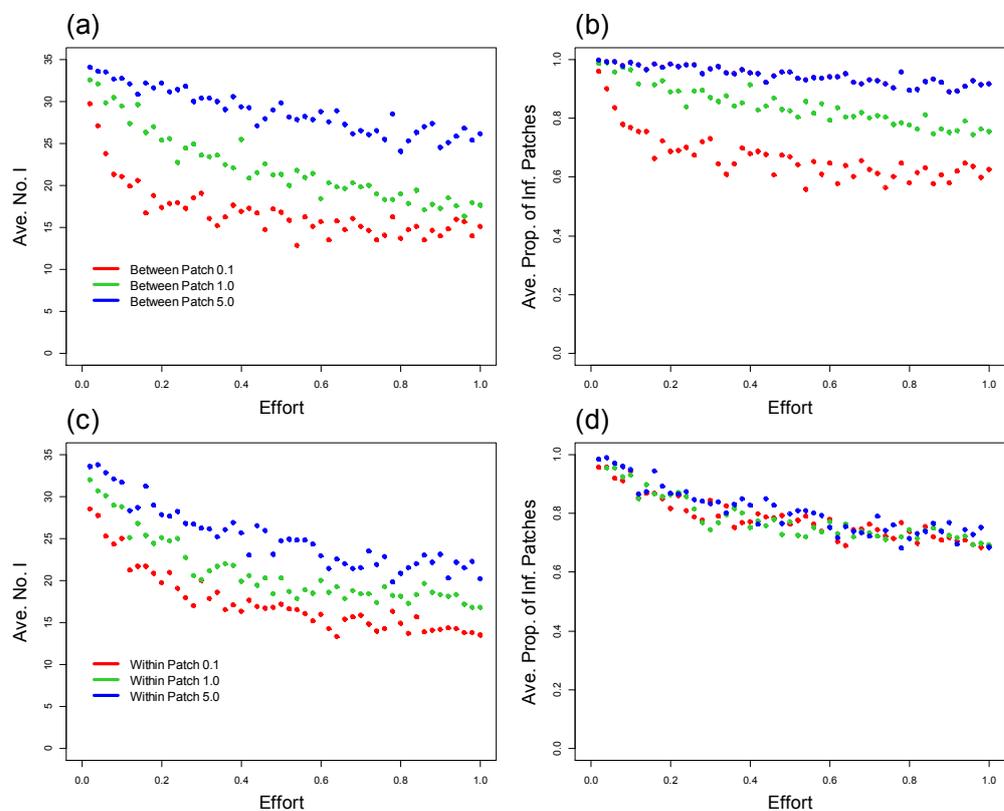


**Figure 4.3 The effect of the suitability of less favourable on surveillance** The proportion of favourable habitat patches is 0.5 i.e. 1-0.5 have less favourable habitat. Data are shown for the difference between targeting surveillance on favourable habitat patches and random sampling of any habitat patch type (a positive difference indicates lower levels of disease at detection in the targeted case). In (a) the average number of infected individuals in the spatial system at the time of first detection and (b) the corresponding figure for the average number of infected patches. Each graph shows four different percentages of patches under surveillance (1%-red, 2%-green, 5%-blue, 25%-purple). The capture rate was set to 1.0 for any patch under surveillance. The suitability rating of the less favourable habitat type was varied range from 0 to 1 (as shown). The plots shown were run using the set of parameter values described in Tables S.3.3 in Appendix 3.1.

Figure 4.3 shows that, generally, targeting favourable habitat patches is effective in reducing the amount of disease in the spatial system at the point of first detection, especially at low i.e. 1% patch inclusion rates. When considering only a small proportion of patches in the surveillance sample, it will generally take longer to detect disease and therefore there is likely to be more disease in the system at the point of first detection. As the habitat becomes more suitable for the host population infection levels will increase, as the disease is able to stabilise in the population and spread more easily within patches. When  $\sigma_i$  is low (i.e. when the less favourable habitat is really poor), sampling favourable habitat is a sensible approach; Figure 4.3 demonstrates a clear reduction in disease levels. However as less favourable habitat improves (i.e. as  $\sigma_i$  increases) this benefit is reduced. This difference is greater when there are a smaller proportion of patches under surveillance. Finally as  $\sigma_i$  tends to 1, there is little difference between favourable and less favourable patches, and it

therefore makes no difference which is targeted as such targeting equates to random sampling of undifferentiated patches.

Figure 4.4 shows the impact of both between patch and within patch transmission with varying surveillance effort. We can see from Figure 4.4 that the higher between patch transmission, the more effort required to decrease the level of disease at time of detection. From other simulations explored, it would appear that at fixed between patch transmission, the within patch contact rate does not have as big an impact on the amount of disease at time of detection. However we anticipate that this would have a greater effect at very low levels where the disease was on the limit of being locally stable.



**Figure 4.4: The effect of surveillance effort and secondary transmission.** Data are shown for three different between patch transmission rates (0.1-red, 1.0-green, 5.0-blue) for varying surveillance effort from 0 – 1 (plots 4.4.a and 4.4.b) and three different within patch transmission rates (0.1-red, 1.0-green, 5.0-blue) for varying surveillance effort from 0 – 1 (plots 4.4.c and 4.4.d). The percentage of patches under surveillance is set at 1% and disease incursion is introduced into a random patch. The

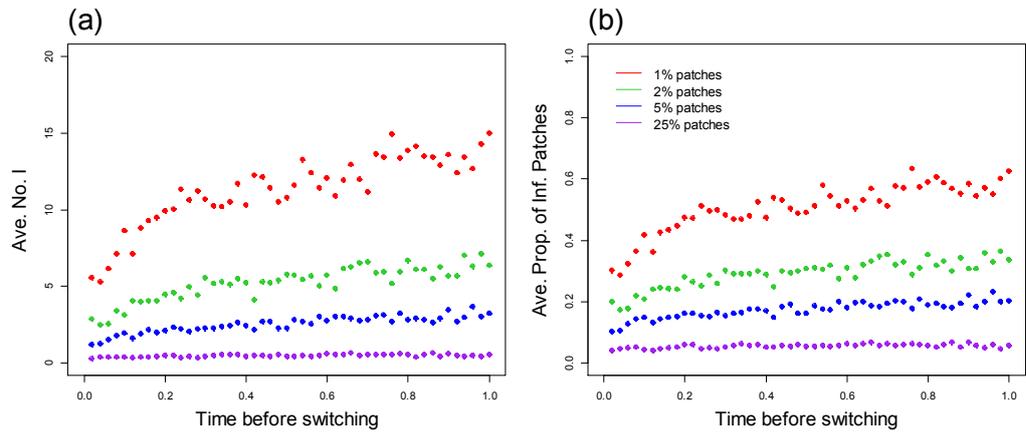
plots shown were run using the set of parameter values described in Tables S.3.4 in Appendix 3.1.

The results above in Figure 4.4 show that the total effort and the distribution of that effort play a role in limiting the spread of emerging disease prior to detection. The design of surveillance is particularly important when habitat is heterogeneous with respect to its suitability for the focal host species. In such cases we have seen that stratifying sampling design with respect to habitat type results in more efficient surveillance. We now consider the effect of disease transmission on the ability to conduct surveillance.

#### 4.4.3 Dynamic designs

Although our results above showed that increasing the number of patches under surveillance improved the ability to detect emerging disease outbreaks, in real surveillance situations, resources are restricted and there will be a limit to how many patches can be brought under surveillance at any one time. A potential alternative strategy to increasing the number of patches in surveillance is fixing the number of patches under surveillance at any one time but to move this effort around by switching surveillance to a new set of patches. The rationale behind this approach is that if the disease is slow at transmitting over the spatial system, staying in a fixed spot could severely hinder the chance of detecting the disease early, whereas if surveillance is moving around the space at a faster rate than the disease, there is potential for earlier detection. We do not consider the costs associated with such dynamic surveillance design (switching) here, but rather focus on quantifying the potential gains in terms of outbreak size at time of first detection.

Figure 4.5 demonstrates the potential effects of switching on the amount of disease present at time of detection. Results are shown for different lengths of time (switching times/rates) surveillance effort is deployed in a patch before this effort is redeployed in another patch. As described earlier the patches to be subjected to such surveillance are chosen at random, but potentially weighted according to some selection criteria e.g. using a stratified design. These results are also shown for different numbers of patches under surveillance at any given time (percentage of patches under surveillance).

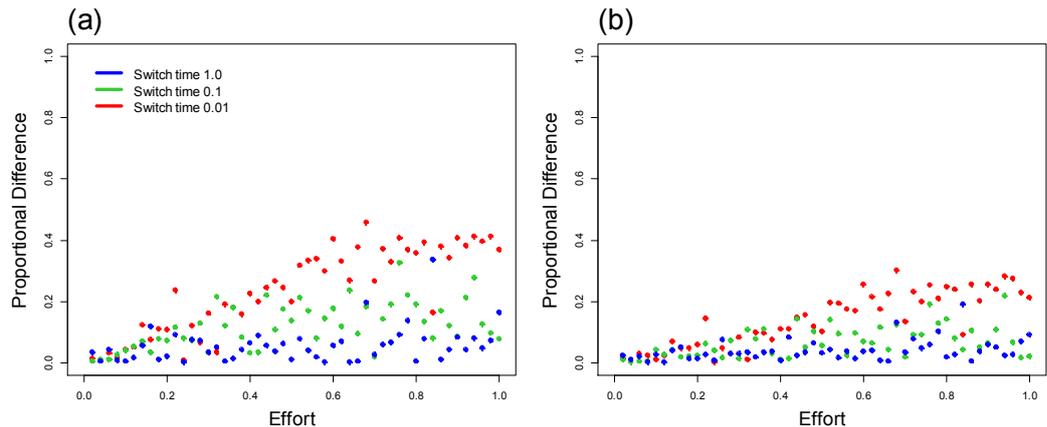


**Figure 4.5: Effect of switching:** Data are shown for four different percentage patches under surveillance (1%-red, 2%-green, 5%-blue, 25%-purple) each for a varying range from 0 to 1 of time spent in each combination of surveillance patch(es) before switching. Amount of disease is calculated at time of first detection for infected (a) individuals and (b) patches. Patches are split 50/50 into favourable and less favourable habitat. Between patch transmission was set to 0.1 and within patch transmission rate set to 0.5. The plots shown were run using the set of parameter values described in Tables S.3.5 in Appendix 3.1.

Figure 4.5 shows switching can reduce the average amount of disease in the global population at time of first detection i.e. allows surveillance to detect outbreaks faster. Note that large switching times approach the limit of no switching whereas smaller ‘time before switching’ corresponds to increasingly fast switching. Therefore the potential gains from dynamic surveillance are measured to a good approximation from Figure 4.5 by comparing the average number infected for switching times 0 and 1. The shorter the time spent in the current surveillance sample before switching, the less prevalent is disease at time of first detection. Figure 4.5 also shows that there is more to be gained from switching when a lower percentage of patches are under surveillance. If the number of patches surveillance sample is large, there will be no more to gain from switching the patches under surveillance periodically. Thus switching is potentially most useful when surveillance is relatively under resourced. This effect is amplified for low between patch transmission rates, i.e. when the disease is less likely to spread between patches.

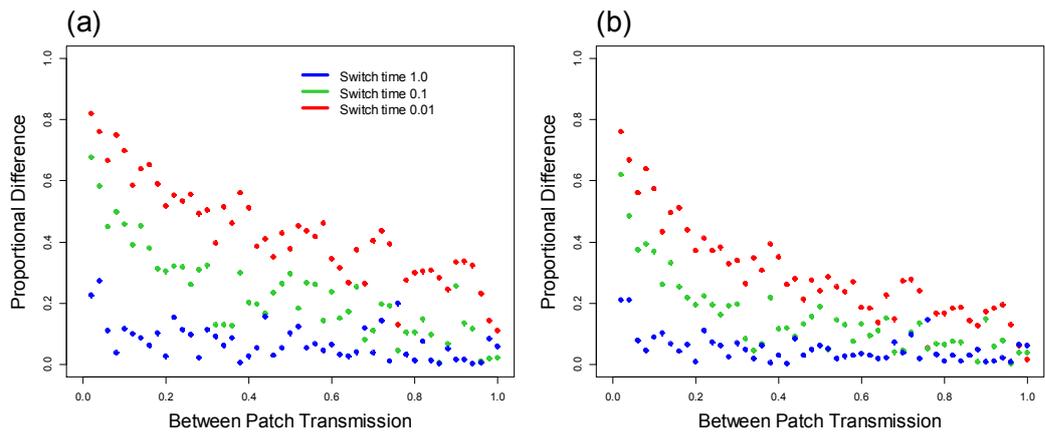
Figure 4.6 considers the proportional difference between switching and not switching for varying effort and three switching rates. The proportional difference (actual gain) of switching compared with not switching is defined as:

$$\text{Proportional difference} = (\text{no switching} - \text{switching rate}) / \text{no switching}$$



**Figure 4.6: Switching and deployed effort.** Percentage improvement obtained from switching compared with not switching, measured in terms of the amount of disease present at time of first detection for infected (a) individuals and (b) patches. Data are shown for three different switching rates (0.01 switching-red, 0.1 switching-green, 1.0 switching-blue) as a function of surveillance effort from 0 – 1. Within and between patch transmission was set to 0.5 and all patches are favourable habitat. The plots shown were run using the set of parameter values described in Tables S.3.6 in Appendix 3.1.

Figure 4.6 shows that the biggest gain seen in terms of reduction of disease occurs when both the switching level is fast (i.e. the surveillance stays in each patch for a very short time before moving to the next patch) and the surveillance effort is highest. Figure 4.7 shows how between patch transmission and the rate at which surveillance switches between patches, affects the percentage difference in disease in the system at point of first detection compared to not switching at all.



**Figure 4.7: The effect of between patch transmission for different switching levels.**

Data are shown for the difference between three different switching levels (0.01 switching-red, 0.1 switching-green, 1.0 switching-blue) and not switching for amount of disease at time of first detection for infected (a) individuals and (b) patches. Between patch transmission rate is varied from 0 – 1 for 1% of patches under surveillance. Within patch transmission was set to 0.5 and capture rate was set to 1.0. The plots shown were run using the set of parameter values described in Tables S.3.7 in Appendix 3.1.

Figure 4.7 shows that by switching surveillance around, the amount of disease present at first detection can be reduced (quite considerably in some cases). At low between patch transmission, the quicker switching between patches occurs, the greater the percentage reduction in the amount of disease at the point of first detection that can be seen compared with no switching. Figure 4.7 also shows that switching at rate 0.01 gives reduction in disease of up to 80% compared to not switching. This is likely to be because in this case surveillance is moving around the region under observation faster than the disease is spreading and is therefore able to detect the outbreak before it becomes widespread.

## 4.5 Discussion

Spatial heterogeneity is a big influencing factor on the spread, prevalence and persistence of disease (Keeling 1999; Fulford *et al.* 2002; Ostfeld *et al.* 2005). In this chapter we have explored the consequences of these effects on the efficacy of disease surveillance applied to disease incursions into pathogen free host meta-populations. Spatial structure and composition of the environment are important components in emerging infections (Favier *et al.* 2005; Suzán *et al.* 2008) as they will help determine the probability of the disease establishing within the population. The surveillance strategy (e.g. distribution of effort) will also determine the time to detection of a disease incursion and thus how much and how widespread the disease is at the point of first detection (Morse 1995; Blanchong *et al.* 2008). If the surveillance effort is too low, or is deployed in an uninfected area for too long then it will be harder to detect disease before the infection has established widely within the modelled area.

In this chapter we considered a relatively stable host population in order to more clearly explore the impact of spatial heterogeneity of habitat suitability on the performance of surveillance. However, in earlier chapters (2 & 3) we learned the importance of population fluctuations in determining the efficacy of surveillance. It would therefore be interesting to consider a wider range of host demographics in the context of surveillance in spatially extended systems. Moreover, the fact that, in a range of systems, spatial heterogeneity has been shown to interact with and typically increase the importance of local stochastic population fluctuations (Tilman & Kareiva 1997) it is likely that the results of chapters 2 and 3 will be equally as important for spatially extended surveillance. Therefore, in addition to the effects shown here, spatially extended surveillance of wildlife systems are likely to lead to biased and lower precision estimates of prevalence and a reduction in the ability to detect disease compared with what would be expected under standard binomial assessments that ignore population and disease fluctuations. We have already explored these effects, driven by temporal heterogeneities in a non-spatial setting and fully expect them to carry over to the spatial case. However, in this chapter we have applied a spatial model to explore the effects of spatial heterogeneities on surveillance rather than revisit the effect of temporal heterogeneities or consider

their interaction with spatial heterogeneity. Nonetheless this would be an interesting avenue for further research.

The task of surveillance is to detect disease before infection becomes a major risk i.e. to limit the spread of infection prior to detection such that disease control remains logistically and financially possible and/or cost effective. Firstly, we have explored how the number of habitat patches included in surveillance affects the amount of disease at first detection. It was shown that by increasing the number of patches that are under surveillance at any one time, the amount of disease at first detection reduces rapidly, even at low efforts. Increasing the area that surveillance can cover will always improve efficacy; however it is usually lack of resources (funding, manpower, knowledge etc) that prevent this from being a practical solution.

As expected the number of patches infected in the space at time of first detection is highly sensitive to the between patch rate of transmission a pathogen. The further the pathogen can spread spatially, the greater number of sub populations it can infect and potentially colonise. In this spatial context, it was shown to have a generally larger impact on the amount of disease in the system as a whole compared to within patch transmission. Within patch transmission is only able to affect the level of disease found in an infected patch and the probability that when infection reaches a susceptible patch that it will be able to sustain.

As mentioned previously, the composition of the environment (i.e. suitability and spread of the habitat) can impact surveillance efficacy. Factors which affect the stability and size of the population will also affect the stability and spread of an emerging disease. It has been demonstrated in this chapter that when the habitat is more favourable for the population, this can increase the amount of disease in the space at the point of first detection. If the population is stable, it will be easier for the pathogen to infect individuals and establish itself in the population. This effect can be partially mitigated against when surveillance is targeted in the most favourable habitat type. Utilising knowledge about the population and environment by targeting areas which are preferable to the population (i.e. due to food sources available, flora, and land type) or areas which are known to contain the population because of previous sightings for example, enables the surveillance strategy to gain the most from the resources available. This result supports the stratification of surveillance by habitat type used by Nusser (2008).

While the biggest reduction in the amount of disease at time of first detection was arguably seen by increasing the number of patches under surveillance, this, as previously discussed, would very rarely be a viable option. Most national and regional surveillance strategies would only be able to cover a small percentage of the total area. Where resources limit the area under surveillance to only a small fraction of the region (1% in the simulations considered here) we have shown in this chapter the potential benefits of dynamically switching the surveillance from patch to patch. If the time between switching is sufficiently small, this can decrease the amount of disease in the system at first detection (i.e. improve the ability of surveillance to detect emerging outbreaks), when compared to not switching, by a significant percentage (in some cases this was shown to be over 80%). This effect is most apparent when the effort rate is high and when the between patch transmission is low.

We have shown that there are alternative strategies to improving the efficacy of surveillance than simply increasing the area that surveillance covers. We have also demonstrated that, like non-spatial demographic factors, the spatial ecology of host populations in relation to heterogeneous distributions of habitat resources affects the efficacy of disease surveillance. It is important to understand the ecology, demography and other factors that influence population and disease dynamics (i.e. habitat quality, quantity, structure and location). We have identified and characterised methods to improve the performance of surveillance when resources and thus effort are limited, as they always will be. Two main strategies were identified to improve the ability of surveillance networks to detect emerging and re-emerging disease threats, namely stratification by habitat suitability and dynamic reallocation of surveillance effort. The literature on wildlife disease surveillance makes limited reference to the first of these (Witmer 2005; Nusser *et al.* 2008; Walsh & Miller 2010), but we are unaware of any literature that discusses switching. Our characterisation of these aspects of surveillance design can be used to inform management strategies, as in many applications it will only be possible to cover a very small percentage of the total area over which surveillance is required. Therefore stratified designs and switching could be deployed to improve surveillance efficacy in these instances. The work presented here is a first step towards exploring spatial heterogeneity and how this impacts on the efficacy wildlife disease surveillance systems. There are many more attributes of ecology that could have been explored,

for example fragmentation of habitat (e.g. clumping, corridors), source-sink dynamics and seasonality. As noted above we expect that the interaction of spatial and temporal heterogeneities will reduce the efficacy of surveillance in line with the results of Chapters 2 and 3. In addition, costs, including switching from one location to another, could also be accounted for in assessing efficacy of surveillance design. These aspects would be interesting next steps to take the research area forward.

# Chapter 5

## General Discussion: Towards a new approach to wildlife disease surveillance

### 5.1 Background

The overall aim of this thesis was to explore how population and disease ecology in wildlife affects surveillance efficacy in a spatial and non-spatial context. The first step of which was to seek out known theory and practice used in designing surveillance schemes. The literature on wildlife disease surveillance is somewhat lacking in terms of protocols used, how they are implemented in the field, approaches used for the analysis of results, and documentation describing what equations are used in the design of surveillance schemes. From the literature available, it seems that to a large extent, practice for wildlife disease surveillance (Artois *et al.* 2009a) is based on ideas developed for livestock systems. This includes both calculation of sample sizes needed for accurate prevalence estimation (Grimes & Schulz 1996; Fosgate 2005) and detection of disease within a population (Dohoo *et al.* 2005). Fluctuations in host populations and disease prevalence are ignored in these methods, and while constant population size and prevalence may be reasonable assumptions for the analysis of livestock systems, they are less appropriate in wildlife disease systems that are characteristically subject to much greater fluctuations in host population density and disease prevalence. Fosgate (2009) reviews current approaches to sample size calculations in livestock systems and emphasises the importance of basing analyses on realistic assumptions about the system under surveillance. The above discussion strongly suggests that current design and analysis for wildlife disease surveillance is not based on realistic assumptions.

Despite a long history of research which addresses how stochasticity impacts disease prevalence and persistence (Anderson 1991; Renshaw 1991; Marion *et al.* 2000; Smith *et al.* 2005) we are not aware of literature explicitly addressing the implications for surveillance design or outcome. Similarly a great deal of work has explored the effects of spatial heterogeneity (Sattenspiel & Simon 1988; Cliff 1995; Tilman & Kareiva 1997; Keeling & Grenfell 1998; Keeling & Rohani 2007) on disease prevalence and persistence. However, for the case of intrinsically generated spatial heterogeneity (which emerges purely from system dynamics even in a homogeneous environment) again there is nothing to our knowledge that specifically addresses how such a fundamental characteristic of spatial systems affects the outcome of surveillance. On the other hand there have been papers which mention spatial stratification of samples based on extrinsic sources of heterogeneity, i.e. biasing surveillance effort towards favourable habitats where the host species will be more likely to be found (Nusser *et al.* 2008; Walsh & Miller 2010).

As noted above current methods for assessing required sample size in wildlife disease surveillance also ignore temporal fluctuations in population size and disease prevalence. In addition, spatial heterogeneity is often a major factor in wildlife disease systems which are typically characterised by significant fluctuations in both space and time. This contrast between the spatial and temporal heterogeneity of real wildlife disease systems and the assumptions of homogeneity that underpin current approaches to the design and analysis of surveillance in such systems is the gap in knowledge which this thesis has sought to address.

## 5.2 Fluctuations undermine wildlife disease surveillance

This thesis has addressed the issues outlined above by firstly creating a stochastic model representing wildlife disease systems and comparing the results with what little common practice has been published in regards to surveillance design. This simple but generic model represents interactions between the three key elements of wildlife disease surveillance systems, i.e. host demographics, pathogen dynamics, and the surveillance effort itself. The efficacy of such surveillance was assessed in terms of the following measures: the probability with which the surveillance system detects disease, the bias of the prevalence estimate and the standard deviation (the inverse of precision) of the prevalence estimate. In Chapter 2 this framework was then used to address generic questions about the efficacy

of wildlife disease surveillance, by considering a range of parameterisations of host demography, representing a broad range of wildlife species, alongside a wide spectrum of different pathogens. Plotting the above measures of surveillance efficacy as a function of prevalence reveals the range of possible outcomes and allows the impact of key aspects, e.g. surveillance effort, sample size and pathogen transmission, to be explored.

For a simplified surveillance scenario analytical results obtained show that unless the covariance of the population and prevalence fluctuations is zero (which is true e.g. if the population or prevalence are constant), there would be bias in the surveillance estimate of prevalence. This result was confirmed for more realistic surveillance scenarios by stochastic simulation of the model. In addition such results were compared with predictions based on standard theory which as noted above relies on binomial arguments and assumptions of constant population size and prevalence. The simulations showed the bias and standard deviation in the estimate of prevalence can be severely underestimated by such standard theory.

Fluctuations in demography and disease dynamics were also shown to compromise the ability of surveillance programmes to detect disease. The results in Chapter 2 demonstrated that standard binomial theory which ignores such fluctuations can also under estimate the power of surveillance programmes to detect disease, i.e. the probability of detecting disease is predicted to be much higher than it actually is in reality. This misinformation gives rise to over-confidence in the ability of the surveillance scheme which can only be detrimental to the design of surveillance, the interpretation of the results and ultimately to the actions taken based on them. This thesis has both highlighted such concerns and begun the development of a more complete theoretical underpinning describing how surveillance statistics are affected by key characteristics of wildlife systems. This work highlights reductions in the efficacy of wildlife disease surveillance caused by population fluctuations and stochasticity in disease dynamics. For example, we have highlighted how power calculations based on fixed prevalence calculations of the probability of disease detection would lead to under powered studies. Moreover the models presented in this chapter and the next could be used as the basis for more reliable power calculations. One area that deserves attention is the inverse problem, namely how to assess, say true prevalence, given only results from surveillance and some knowledge of the underlying wildlife disease system. However, this is beyond the scope of this thesis.

### 5.3 Tools for assessing wildlife disease surveillance in real systems

In Chapter 3 we demonstrate how the modelling framework introduced in Chapter 2 may be adapted and employed as a tool to assess the efficacy of surveillance in real systems. Such an approach naturally provides an aid to the design of surveillance programmes if the efficacy of multiple surveillance designs are evaluated and compared. We sought to quantify the effects identified in Chapter 2 for two specific wildlife-disease systems to both illustrate the power of the model and also to better understand surveillance in badger and rabbit populations. This work showed that the generic findings of Chapter 2 are relevant to real wildlife disease systems. Focusing on disease induced mortality as a natural source of bias found in wild populations, we have demonstrated that the interaction between disease dynamics and population stability impacts the ability to both detect disease and accurately estimate the prevalence in the population, as was shown for a wide class of systems in Chapter 2.

We extended the theory quantified from Chapter 2 to explore other known, previously acknowledged, sources of bias to assess the effect they have on the efficacy of surveillance. Bias in the trappability of an infected host, biases the surveillance prevalence estimate. This can happen in two ways; there will be a positive bias (i.e. the surveillance over-estimates the prevalence) if individuals are trap happy, and there will be negative bias (i.e. the surveillance under-estimates the prevalence) if the individuals are trap shy. The probability of detection decreased for infected individuals showing trap shy behaviour, and this is to be expected as this lowers the likelihood of an infected individual being caught and lowers the chance the infection will be detected.

The impact of the sensitivity of diagnostic tests of disease used in surveillance was also explored and showed some interesting results. Imperfections in diagnostic tests would appear to mask the effect of fluctuations on the ability of surveillance to detect disease in that for poor tests the resulting probability of detection (based on simulations that account for fluctuations) is close to the constant prevalence theory estimate. As test sensitivity increases the difference between the true probability of detection and the constant prevalence theory prediction widens. We can conclude that gains in the sensitivity of diagnostics are likely to be reduced in the field by the effects of fluctuations. On the other hand these results also suggest that as test sensitivities increase the inadequacies of standard theory should become more transparent.

In summary we can assert that there are clear differences between the constant prevalence theory estimates and our simulated results. The latter, account for fluctuations in host population and disease dynamics, and clearly show impacts on the efficacy of surveillance. These effects have the potential to impact management strategies for current disease challenges and should be assessed through simulation tools such as those introduced here and explored further in the field

## 5.4 Spatial heterogeneity and the design of wildlife disease surveillance

The generic framework was extended to incorporate spatial heterogeneity. The impact of design on the ability of spatially distributed surveillance networks to detect emergent disease at a regional scale was then assessed. Moving on from the non-spatial models used in the previous two chapters, chapter four aimed to explore how the ecology of populations in space would affect the efficacy of surveillance in an incursion scenario. We paid particular attention to how surveillance would be distributed and how habitat quality affecting the population would then go on to affect the surveillance outcome. By calculating the amount of disease in the system at the point of first detection it was clear that the more “patches” in space the surveillance was able to visit at one time greatly reduced the amount of disease seen in the system. In practice active surveillance can be prohibitively expensive and funding is often a limiting factor. Furthermore, the amount of time/manpower required to cover a large enough area of the habitat under surveillance may be unfeasible.

To address this issue, we explored the idea of only visiting a select percentage of the spatial system, but switch periodically around the space using different strategies (i.e. only focusing on good habitat areas). When the time between each switching was very small and the effort (capture rate) of surveillance was large enough, there was a reduction in the amount of disease present at time of first detection. The biggest effect seen in switching the surveillance around the spatial system happened at high effort (capture rate), low between patch transmission and sampling the favourable habitat patches only. The first factor, high effort, is intuitively important in the efficacy of surveillance, and has been demonstrated to impact on surveillance in Chapter 2. The more effort that is put into surveillance, the quicker the disease will be detected, and therefore the less disease will be present at the time of first detection. Switching is more likely to be advantageous when the between patch transmission rate is low, as the spread between patch to patch will be limited and

surveillance (if the effort is high enough) will have a better chance of detecting disease before it has developed into a major risk.

We have seen instances before of strategies targeting surveillance in order to improve the probability of collecting samples (Nusser *et al.* 2008; Walsh & Miller 2010) discussed in Chapter 2, here we have explored the effect of this strategy further. Comparing the reduction in disease at the point of detection when targeting favourable habitat to choosing patches randomly to be under surveillance, we have shown that there is an advantage in targeting habitats in which the population thrives. Again, this is intuitive as a surveillance strategy which targets areas in which the population lives (and preferably is stable) will have a higher probability of detecting infection if the population has been exposed to an incursion. Knowledge about the population and environment (i.e. food sources available, flora, land type, behaviours of host, social structure) should always be incorporated into management strategies as this enables the surveillance strategy to gain the most from the resources available.

## 5.5 Future Work

This thesis has demonstrated the importance of understanding demographic fluctuations and disease dynamics in the host population when undertaking surveillance. The potential for both implementations (SDE and Gillespie) of the stochastic model used in Chapter 2 have been highlighted in Chapter 3, but there are still many other attributes of demography, surveillance, habitat and environmental factors that are still to be explored.

In terms of surveillance, throughout this thesis we have only considered a type of active surveillance. However, in terms of national wildlife disease surveillance, passive surveillance is most often used around the world, as discussed previously, as resources for active surveillance can be limited. It would be interesting to explore the impact demographic fluctuations and disease dynamics have on passive surveillance, especially in cases of high disease induced mortality. Since passive surveillance is predominantly the collection of deceased animals (road kill, death by natural causes, hunting etc), there is a large scope for bias in the sample, especially when trying to estimate the prevalence. We have shown in Chapter 2 and Chapter 3 how much bias can be seen when dealing with active surveillance,

and as passive surveillance has an associated bias of its own, the results found previously have the potential to be magnified quite considerably. The demographic components associated with passive surveillance (decay of deceased individuals, scavenge rate of deceased individuals, probability of detecting deceased individuals, probability of public reporting a finding for potential testing etc) could be easily incorporated into the existing model as additional rates. The decay aspect for example could be included as an exponential rate increasing with time. This would multiply by the number of deceased individuals to determine the probability of full decomposition and removal from the deceased population. This approach would reflect that the probability of a dead body decomposing completely increases with the amount of time spent deceased.

In an effort to keep the complexity of the stochastic model at a manageable level, and to leave scope for several different applications, there are common population traits that have not been included. It would be possible to distinguish between attributes such as, sex, age, dominance, added disease states etc, and to explore the effect of these on the efficacy of surveillance. For example, if there was a “recovered” state included in the disease dynamics, it would be interesting to explore how this would affect the bias seen previously in the prevalence estimate and probability of detection. This could be particularly significant when disease induced mortality is included in the disease dynamics. In addition inclusion of such a disease related state would complicate the modelling of surveillance since it is likely that different diagnostic tests would detect individuals in latent and infective states and the recovered state. The former would be targeted by pathogen recovery e.g. bacterial culture or DNA based tests for the presence and abundance of pathogens, whereas the latter would most likely be detected by tests based on serological analysis indicating exposure to a pathogen e.g. detecting antibodies (Teunis *et al.* 2002; Bidet *et al.* 2008; López-Olvera *et al.* 2010). The contrasting nature of these diagnostic tools would require identification of appropriate sensitivity and specificity levels associated with testing animals in different disease states. This would add to the complexity of modelling but such an analysis may reveal further sources of bias and uncertainty in wildlife disease surveillance.

An external influence that has not been considered is climate; seasonality is known to affect the spread of disease (Dowell 2001; Altizer *et al.* 2006); and climate change is predicted to have significant impacts on disease (Anderson *et al.* 2004; Trenberth 2008). Seasonality could impact the stochastic fluctuations of the population as many natural populations have specific breeding seasons. Limiting the growth of the population to certain seasons could, in

some cases, decrease the overall system stability if a disease had a high enough disease induced mortality rate. We have seen previously in this thesis that unstable population dynamics have a big impact on the bias and uncertainty of the prevalence estimate as well as in the probability of detection. Seasonal reproduction drives the total numbers of animals on the ground to dramatically increase in spring and summer compared to the autumn/winter population size. Furthermore the disease susceptibility profile of the population changes as disease-resistant adult dominated populations are diluted with disease susceptible young. As well as this, winter is often associated with starvation and the adults surviving winter frequently emerge with a compromised immune system. There are well known and already characterised drivers of population fluctuation and disease dynamics, and the results obtained in this thesis strongly suggest that such factors need to be considered when designing wildlife disease surveillance. There is also scope for including such things as extreme weather events, for example. Climate change and extreme weather events have the potential to increase the prevalence of disease within a population (D'Amato *et al.* 2013) which could well have adverse consequences for the efficacy of surveillance i.e. if the influx of disease was rapid, surveillance might be too slow at detecting disease which would lead to increased disease risk.

Spatial heterogeneity has been explored in Chapter 4 in terms of the ability of surveillance to detect disease incursion in spatially explicit wildlife disease system, but there are many more avenues to explore. Habitat distribution is known to influence both the host population demography and persistence of disease (Fahrig 2003). We have touched upon this by considering suitability of habitat and proportion of suitable habitat within the space. However another aspect we have yet to explore is clumping of habitat, i.e. the tendency of habitat patches of similar types be closer together than typical distances to habitats of a different type. If the habitat is clumped, this can mean that if disease is introduced into a clump then it will have a strong chance of spreading and sustaining itself in nearby habitats. However, other clumps of habitat patches of the same type (e.g. habitat suitable for a given host species) may be at less risk if the distance between such clumps is too far for the disease to travel. Clumping of habitats can lead to populations becoming isolated and to preserve species diversity, often corridors are introduced to allow animals to move from habitat clumps. While this can be good for the overall population level, it can also lead to an increase in the spread and persistence of disease. This has an obvious potential to impact on the efficacy of surveillance and it would be a worthwhile area to explore.

This thesis has been a first step in characterising how key ecological features of natural populations impact on surveillance efficacy. We have shown that currently standard assumptions made when designing surveillance strategies and design over-look important stochastic fluctuations and biases found in estimating prevalence. Taking this research forward will hopefully benefit surveillance in the field by aiding more informed design, and interpretation of results. As the recent example of the West Africa Ebola outbreak has highlighted (Nishiura & Chowell 2014; PHE 2014), there is still a clear need for better disease surveillance strategies. The findings of this research should be investigated in the field to assess how important the relationships we have characterised are in practice. We have shown there are several factors which affect the efficacy of surveillance and at the very least should be considered when designing surveillance programs. Although wildlife disease surveillance has been our motivating example, there is no reason why the effects discussed in this thesis or related phenomena should not be applicable to human and livestock populations e.g. that are subject to large population fluctuations. It is difficult to give specific recommendations for wildlife disease surveillance, as each host-pathogen system is unique. However we have presented results (e.g. specific graphs) that span a broad spectrum of wildlife disease systems, and these perhaps be used to design relatively robust surveillance systems (e.g. Figure 2.4) . In addition Chapter 3 demonstrated that we have developed tools in the form of our stochastic simulation models and the wider framework within which we have used them to assess the efficacy of different surveillance systems. This research demonstrated the potential to apply this framework to specific host populations and diseases of current interest, In addition the model we have developed has the scope to be adapted and used in many different situations. We conclude that host demography, disease dynamics and spatial heterogeneity do impact on the efficacy of surveillance and must be considered when undertaking/designing surveillance of wildlife disease systems.

# Appendices

# Appendix 1

## 1.1 Relationship between discrete and continuous (SDE) state-space model implementations

In this appendix we describe the relationship between the continuous time discrete state-space Markov process and the stochastic differential equation (SDE) implementations of the model described in the main text.

Our starting point is the SI model described in Table 2.1 (main text) implemented as a continuous time discrete state-space Markov process in which the number of infected individuals  $I(t)$  and total population size  $N(t) = S(t)+I(t)$ , are represented as integer variables. The Gillespie algorithm exploits the fact that the time between events is distributed exponentially with parameter  $R(t)$  given by the sum of all the event rates in Table 2.1 and the probability that a given event occurs is given by the associated event rate divided by  $R(t)$ .

However, under this implementation one can also consider the expectation and variance-covariance of the change in the state-space variables  $I(t)$  and  $N(t)$  during a small time interval. For convenience denote the state of the system at time  $t$  by  $X(t)=\{I(t),N(t)\}$ . Then for example, conditional on the state of the system at time  $t$ , the expected change in the population size associated with birth events from time  $t$  to  $t+\delta t$  is given by  $E_B[\delta N(t)|X(t)] = rN(t) (1 - N(t)/k)\delta t$ . Similarly, the variance in  $\delta N$  associated with birth events is  $\text{Var}_B[\delta N(t)]= rN(t) (1 - N(t)/k)\delta t + O(\delta t^2)$ , and henceforth we will assume  $\delta t$  is sufficiently small to ignore the higher order terms. In the model described in the main text (see Table 2.1 and surrounding text) all individuals are born susceptible and therefore birth does not affect the infective population size  $I(t)$  i.e.  $E_B[\delta I(t)|X(t)] = 0$ ,  $\text{Var}_B[\delta I(t)]=0$ , and  $\text{Cov}_B [\delta I(t),\delta N(t)|X(t)]=0$ . However, migration of infectives affects both  $I(t)$  and  $N(t)$  and to first order in  $\delta t$  we find that  $E_{mi}[\delta N(t)|X(t)]= \gamma v \delta t$ ,  $\text{Var}_{mi}[\delta N(t)]= \gamma v \delta t$ ,  $E_{mi}[\delta I(t)|X(t)] = \gamma v \delta t$ ,  $\text{Var}_{mi}[\delta I(t)]= \gamma v \delta t$  and  $\text{Cov}_{mi} [\delta I(t),\delta N(t)|X(t)]= \gamma v \delta t$ . The full set of first- and second-order statistics describing changes in the state-space associated with each event type are given (up to first order in  $\delta t$ ) in Table S.1.1.

E-type	Event	$E[\delta N X(t)]$	$E[\delta I X(t)]$	$\text{Var}[\delta N X(t)]$	$\text{Var}[I X(t)]$	$\text{Cov}[\delta N, \delta I X(t)]$
B	Birth	$rN(1 - N/k)\delta t$	0	$rN(1 - N/k)\delta t$	0	0
DS	Death of S	$-\mu S\delta t$	0	$\mu S\delta t$	0	0
DI	Death of Infected	$-\mu I\delta t$	$-\mu I\delta t$	$\mu I\delta t$	$\mu I\delta t$	$\mu I\delta t$
mS	S Immigration	$(1 - \gamma)v\delta t$	0	$(1 - \gamma)v\delta t$	0	0
mI	Infected Immigration	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$	$\gamma v\delta t$
1ry	Primary Transmission	0	$\beta_0 S\delta t$	0	$\beta_0 S\delta t$	0
2ry	Secondary Transmission	0	$\beta I S\delta t$	0	$\beta_0 S\delta t$	0

**Table S.1.1:** Expectations and variance-covariances in changes (during the time interval  $t$  to  $t+\delta t$ ) to the state space  $\{I(t), N(t)\}$  associated with each event type in the discrete state-space model described in the main text (see Table 1.1). All such quantities are shown to first order in  $\delta t$ . Note: capture and release events are omitted since they affect neither  $I(t)$  or  $N(t)$ .

We now show how to construct a continuous time, continuous state-space (diffusion) version of the model which is consistent with above implementation in that it preserves the means and variance-covariance statistics shown in Table S.1.1. To do so we construct a set of stochastic differential equations (SDEs) which we later solve numerically in discrete time steps (Mao 1997; Higham 2001). The following Itô stochastic differential equations represent the change in the system state variables during an infinitesimally small time interval  $dt$

$$\begin{aligned}
dN(t) &= \left( f_{N,B}(X(t)) + f_{N,DS}(X(t)) + f_{N,DI}(X(t)) + f_{N,mS}(X(t)) \right. \\
&\quad \left. + f_{N,mI}(X(t)) + f_{N,1ry}(X(t)) + f_{N,2ry}(X(t)) \right) dt \\
&\quad + g_{N,B}(X(t))dB_B(t) + g_{N,DS}(X(t))dB_{DS}(t) + g_{N,DI}(X(t))dB_{DI}(t) \\
&\quad + g_{N,mS}(X(t))dB_{mS}(t) + g_{N,mI}(X(t))dB_{mI}(t) + g_{N,1ry}(X(t))dB_{1ry}(t) \\
&\quad + g_{N,2ry}(X(t))dB_{2ry}(t)
\end{aligned}$$

$$\begin{aligned}
dI(t) &= \left( f_{I,B}(X(t)) + f_{I,DS}(X(t)) + f_{I,DI}(X(t)) + f_{I,mS}(X(t)) + f_{I,mI}(X(t)) \right. \\
&\quad \left. + f_{I,1ry}(X(t)) + f_{I,2ry}(X(t)) \right) dt \\
&\quad + g_{I,B}(X(t))dB_B(t) + g_{I,DS}(X(t))dB_{DS}(t) + g_{I,DI}(X(t))dB_{DI}(t) \\
&\quad + g_{I,mS}(X(t))dB_{mS}(t) + g_{I,mI}(X(t))dB_{mI}(t) + g_{I,1ry}(X(t))dB_{1ry}(t) \\
&\quad + g_{I,2ry}(X(t))dB_{2ry}(t)
\end{aligned}$$

Here the quantities  $B_B(t)$ ,  $B_{DS}(t)$ ,  $B_{DI}(t)$ ,  $B_{mS}(t)$ ,  $B_{mI}(t)$ ,  $B_{1ry}(t)$ ,  $B_{2ry}(t)$  are independent Brownian motions corresponding to each of the seven event types and the correct interpretation of these equations requires consideration of associated stochastic integrals (REF). For small but finite  $dt$  the quantities  $dB_B(t)$ ,  $dB_{DS}(t)$ ,  $dB_{DI}(t)$ ,  $dB_{mS}(t)$ ,  $dB_{mI}(t)$ ,  $dB_{1ry}(t)$ ,  $dB_{2ry}(t)$  can be interpreted as independent draws from a zero mean Gaussian with variance  $dt$  for each event type and each time point  $0, dt, 2dt, \dots, T \in (0, T)$ . Thus e.g.  $E[dB_B(t)] = 0$ ,  $E[dB_B(t)dB_B(t)] = 0$  and  $E[dB_B(t)dB_{DS}(t)] = 0$ . This discretisation is the basis for the numerical simulation of these SDEs used in this paper.

The so-called drift,  $f_{N,B}(X(t))$ ,  $f_{N,DS}(X(t))$ ,  $f_{N,DI}(X(t))$ ,  $f_{N,mS}(X(t))$ ,  $f_{N,mI}(X(t))$ ,  $f_{N,1ry}(X(t))$ ,  $f_{N,2ry}(X(t))$  and diffusion,  $g_{N,B}(X(t))$ ,  $g_{N,DS}(X(t))$ ,  $g_{N,DI}(X(t))$ ,  $g_{N,mS}(X(t))$ ,  $g_{N,mI}(X(t))$ ,  $g_{N,1ry}(X(t))$ ,  $g_{N,2ry}(X(t))$ , terms representing changes in the variable  $N(t)$  and the corresponding quantities representing changes in  $I(t)$  are deterministic functions of the state-space  $X(t)$  determined as follows.

Given the nature of the Brownian motions taking the expectation of the above equations yields

$$\begin{aligned}
E[dN(t)|X(t)] &= \left( f_{N,B}(X(t)) + f_{N,DS}(X(t)) + f_{N,DI}(X(t)) + f_{N,mS}(X(t)) + f_{N,mI}(X(t)) \right. \\
&\quad \left. + f_{N,1ry}(X(t)) + f_{N,2ry}(X(t)) \right) dt
\end{aligned}$$

$$\begin{aligned}
E[dI(t)|X(t)] &= \left( f_{I,B}(X(t)) + f_{I,DS}(X(t)) + f_{I,DI}(X(t)) + f_{I,mS}(X(t)) + f_{I,mI}(X(t)) \right. \\
&\quad \left. + f_{I,1ry}(X(t)) + f_{I,2ry}(X(t)) \right) dt
\end{aligned}$$

Which suggests that for each event type  $Etype$   $f_{N,Etype}(X(t))$  and  $f_{I,Etype}(X(t))$  should be interpreted as the mean update shown in Table S1 for  $N(t)$  and  $I(t)$  respectively. For example,  $f_{N,1ry}(X(t))$  and  $f_{N,2ry}(X(t))$  are both zero since only birth, death and migration change the population size, i.e. neither primary nor secondary infection changes the population size.

The variance in the update for  $N(t)$  is given by

$$\text{Var}[dN(t)|X(t)] = E[dN(t)^2|X(t)] - E[dN(t)|X(t)]^2$$

However, we have just shown that  $E[dN(t)|X(t)]$  is of order  $dt$  and therefore to first order in  $dt$  we can write

$$\begin{aligned} \text{Var}[dN(t)|X(t)] = E[dN(t)^2|X(t)] = \\ g_{N,B}(X(t))^2 dt + g_{N,DS}(X(t))^2 dt + g_{N,DI}(X(t))^2 dt + g_{N,mS}(X(t))^2 dt \\ + g_{N,mI}(X(t))^2 dt + g_{N,1ry}(X(t))^2 dt + g_{N,2ry}(X(t))^2 dt \end{aligned}$$

and

$$\begin{aligned} \text{Var}[dI(t)|X(t)] = E[dI(t)^2|X(t)] = \\ g_{I,B}(X(t))^2 dt + g_{I,DS}(X(t))^2 dt + g_{I,DI}(X(t))^2 dt + g_{I,mS}(X(t))^2 dt \\ + g_{I,mI}(X(t))^2 dt + g_{I,1ry}(X(t))^2 dt + g_{I,2ry}(X(t))^2 dt \end{aligned}$$

Here we have made use of the independent nature of the Brownian motions described above.

These last two equations therefore suggest that for each event type  $Etype$ ,  $g_{N,Etype}(X(t))^2$  and  $g_{I,Etype}(X(t))^2$  should be interpreted as the variance in update shown in Table S.1.1 for  $N(t)$  and  $I(t)$  respectively.

The above calculations are summarised in Table S.1.2. Comparison with Table S.1.1 allows the functional form for each drift and diffusion term to be identified.

Finally, the covariance

$$\text{Cov}[dN(t)dI(t)|X(t)] = E[dN(t)dI(t)|X(t)] - E[dN(t)|X(t)]E[dI(t)|X(t)]$$

to first order in  $dt$  is given by

$$\begin{aligned} \text{Cov}[dN(t)dI(t)|X(t)] &= E[dN(t)dI(t)|X(t)] = \\ &+ g_{N,DI}(X(t))g_{I,DI}(X(t))dt + g_{N,mI}(X(t))g_{I,mI}(X(t))dt \end{aligned}$$

where we have shown only the non-zero terms. Comparison with the functional forms for the diffusion terms described above shows that this expression is consistent with the covariance terms shown in Table S.1.1.

E-type	Event	$E[\delta N X(t)]$	$E[\delta I X(t)]$	$\text{Var}[\delta N X(t)]$	$\text{Var}[I X(t)]$
B	$f_{N,B}(X(t))dt$	$f_{I,B}(X(t))dt$	$g_{N,B}(X(t))^2 dt$	$g_{I,B}(X(t))^2 dt$	0
DS	$f_{N,DS}(X(t))dt$	$f_{I,DS}(X(t))dt$	$g_{N,DS}(X(t))^2 dt$	$g_{I,DS}(X(t))^2 dt$	0
DI	$f_{N,DI}(X(t))dt$	$f_{N,DI}(X(t))dt$	$g_{N,DI}(X(t))^2 dt$	$g_{I,DI}(X(t))^2 dt$	$g_{N,DI}(X(t)) \times g_{I,DI}(X(t))dt$
mS	$f_{N,mS}(X(t))dt$	$f_{I,mS}(X(t))dt$	$g_{N,mS}(X(t))^2 dt$	$g_{I,mS}(X(t))^2 dt$	0
mI	$f_{N,mI}(X(t))dt$	$f_{I,mI}(X(t))dt$	$g_{N,mI}(X(t))^2 dt$	$g_{I,mI}(X(t))^2 dt$	$g_{N,mI}(X(t)) \times g_{I,mI}(X(t))dt$
1ry	$f_{N,1ry}(X(t))dt$	$f_{I,1ry}(X(t))dt$	$g_{N,1ry}(X(t))^2 dt$	$g_{I,1ry}(X(t))^2 dt$	0
2ry	$f_{N,2ry}(X(t))dt$	$f_{I,2ry}(X(t))dt$	$g_{N,2ry}(X(t))^2 dt$	$g_{I,2ry}(X(t))^2 dt$	0

**Table S.1.2:** Expectation and variance-covariances in changes (during the time interval  $t$  to  $t+dt$ ) to the state space  $\{I(t), N(t)\}$  associated with each event type in the SDE model as described in Appendix 1. All such quantities are shown to first order in  $dt$ . Comparison with Table S.1.1 enables both drift e.g.  $f_{N,B}(X(t))$  and diffusion e.g.  $g_{N,B}(X(t))$  functions to be identified. Note: capture and release events are omitted since they affect neither  $I(t)$  or  $N(t)$ .

## 1.2 Parameterisations used

This section of the appendix describes in detail the parameter combinations used to produce the graphs in the main text. Values of the form: a,b,c,d etc refer to discrete values used for different lines shown on the Figures. Values of the form a;b;c refer to smallest value; largest value; step size describing the range of values (e.g. of the death rate) simulated to produce the Figures. Values of the form a – b refer to the range of values covered with a non-constant step size. All other parameters with single values are held constant in simulations.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.1, 0.04, 0.01
Carrying Capacity	$k$	120
Growth Rate	$r$	0.5
Death Rate	$\mu$	0.1;0.5;0.1
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0
Susceptible Active Capture	$\alpha$	0.1
Infected Active Capture	$\alpha$	0.1
Sample Target	$m$	10.0

**Table S.1.3:** Parameter values are shown for Figure 2.1 in the main text which demonstrates the effect of the death rate and transmission rate on the bias and variance of the prevalence estimate as well as the effect of the death rate on the population size and variance. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.1
Carrying Capacity	$k$	120
Growth Rate	$r$	0.5
Death Rate	$\mu$	0.4, 0.43
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0
Susceptible Active Capture	$\alpha$	0 - 10
Infected Active Capture	$\alpha$	0 - 10
Sample Target	$m$	10.0

**Table S.1.4:** Parameter values are shown for Figure 2.2 in the main text which demonstrates the effect of the capture rate on the bias and variance of the prevalence estimate. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.1
Carrying Capacity	$k$	120
Growth Rate	$r$	0.5
Death Rate	$\mu$	0.4, 0.43
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0
Susceptible Active Capture	$\alpha$	0.1
Infected Active Capture	$\alpha$	0.1
Sample Target	$m$	1 - 10000

**Table S.1.5:** Parameter values are shown for Figure 2.2 in the main text which demonstrates the effect of the sample size on the bias and variance of the prevalence estimate. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.1, 0.04, 0.01
Carrying Capacity	$k$	120
Growth Rate	$r$	0.5
Death Rate	$\mu$	0.1;0.5;0.01
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0
Susceptible Active Capture	$\alpha$	10, 1.0, 0.1, 0.01
Infected Active Capture	$\alpha$	10, 1.0, 0.1, 0.01
Sample Target	$m$	10

**Table S.1.6:** Parameter values are shown for Figure 2.3 in the main text which demonstrates the effect of the death rate and transmission rate, as well as the sample size and capture rate, on the probability of detecting disease. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

<b>Rate Name</b>	<b>Rate</b>	<b>Value</b>
<b>Secondary Transmission Rate</b>	$\beta$	0.01,0.05,0.09,0.2,0.6,1.0,2.0,5.0
<b>Carrying Capacity</b>	$k$	1;36.0;3.5
<b>Growth Rate</b>	$r$	0.5;23;2.5
<b>Death Rate</b>	$\mu$	0.25;14.0;1.25
<b>Immigration</b>	$\nu$	1.0
<b>Infected Immigration Proportion</b>	$\gamma$	0.01
<b>Primary Transmission Rate</b>	$\beta_0$	0.01
<b>Susceptible Active Capture</b>	$\alpha$	0.5
<b>Infected Active Capture</b>	$\alpha$	0.5
<b>Sample Target</b>	$m$	10.0, 20.0

**Table S.1.7:** Parameter values are shown for Figure 2.4 in the main text which demonstrates the effect of the transmission, death rate, birth rate, carrying capacity, as well as the sample size, on the probability of detecting disease. 1000 simulations were run per parameter combination with a time limit of 45. If the simulation did not reach the sample target within the time limit, the run is discarded and not used in the statistical calculations. If out of 1000 realisations a parameter combination ceases to reach the sample target at least 15 times, that parameter combination is discarded totally as the results are deemed to be unreliable. Increasing the time limit bears little to no effect on the amount simulations which reach the target sample, so the precise value of the time limit does not affect the results obtained from the model. These parameters were implemented using the Gillespie version of the model.

# Appendix 2

## Appendix 2.1

### 2.1.1 Estimating Badger Parameters

The paper on which the badger parameter exploration was based, Shirley *et al* 2003, describes an individual based spatially explicit model, using discrete probabilities for all event types. The model obtains data from each badger at 6 month intervals to determine the life history. The model accounts for sex, three separate age classes (cub, juvenile and adult), TB status and breeding structure where only one dominant breeding pair produce offspring. This creates additional difficulties when deriving parameters for the models used in this thesis since not only do we need to translate from discrete to continuous time formulations we also need to define average rates in our model using information describing the various classes defined in Shirley *et al*. Information was also used from Rogers *et al* 2003 and Kruuk and Parish 1982 for the carrying capacity estimate. Below we discuss how we translate between the parameters given in Shirley *et al* 2003 and those required for the models used in Chapter 2.

#### 2.1.1.1 Discrete to continuous time

Suppose under the discrete time model there is a time step which represents  $\Delta t$  units of time (e.g. 6 months in the above example) and a given event type happens with probability  $P$  during this interval. For example, this event might correspond to a transition from one age class to another (i.e. survival to a given age). Starting at time  $t=0$  the probability that this event has yet to occur after  $n$  time steps i.e. by time  $t=n\Delta t$  is given by:

$$(1 - P)^n$$

On the other hand consider an analogous continuous time Markov process where the event rate is  $\alpha$ , then as we saw in Chapter 1, the probability that the event has yet to occur by time  $t=n\Delta t$  is:

$$e^{-\alpha n\Delta t}$$

Equating the last two equations, taking logs and noting that the number of time steps  $n$  cancels we obtain:

$$\alpha = -\frac{\ln(1-P)}{\Delta t},$$

which enables translation between discrete and continuous time parameterisations. Also note that when  $P$  is small  $-\ln(1-P) \sim P$  and this can be approximated by  $\alpha = P/\Delta t$ .

Considering the event described above as the survival of an individual through a stage class (e.g. relating to an age range or a developmental stage) the average time spent in the stage is given by the length of each time step multiplied by the average number of steps before the event occurs, i.e.  $\Delta t * 1/P$  in the discrete time model. This results from the fact that the probability of making a transition to the next stage class at time step  $n$  is given by a geometric distribution with a probability  $P$  of success per trial – which has mean  $1/P$  (Consul *et al.* 2006). In the continuous time model the time to the next event is an exponential distribution (as shown above) which has mean  $1/\alpha$ . Equating these expected times from the discrete and continuous time models leads to:

$$\alpha = \frac{P}{\Delta t},$$

which as we saw is equivalent to the previous formula when  $P$  is sufficiently small.

### 2.1.1.2 Translating from multi- to single-stage models

Where parameters are available for multiple stages e.g. age or developmental stages, one approach to parameterising a model with only a single stage is to focus on the expected time spent transiting all the stages. For example Shirley *et al* 2003 consider cubs, juveniles and adults and therefore if we wish to consider mortality we need to estimate the expected lifespan i.e. the average time taken to transit all three of these stages from birth to death.

Suppose in a discrete time stage structured model, the length of the first two stages are both a single time step and the probability of surviving the cub stage is  $A$  and the probability of surviving the juvenile stage is  $B$ . Thereafter if the probability of surviving each time step as an adult is  $C$ , then as we saw above the average number of time steps spent as an adult is  $1/C$  i.e. time spent as an adult =  $\Delta t * 1/C$ . However, since it takes two time steps to reach the adult stage then the expected life time would be  $\Delta t * (2 + 1/C)$  if the survival of

cubs to juveniles and juveniles to adults were both assured i.e where  $A=B=1$ .

In cases where  $A, B < 1$  and only a fraction of cubs and juveniles survive we calculate the expected life span in terms of the contribution from three groups, namely those that die as cubs, those that die as juveniles or those that die as adults. To do so, we consider the probability that an individual is in each category multiplied by the length of life in each category. However, note that in the discrete time model we only count in full time steps and therefore individuals that die as cubs do not contribute to the expected lifespan.

The probability that an individual dies as a juvenile is  $A*(1-B)$  i.e. they survived the cub stage but died during the juvenile stage. Individuals in this category live for one but less than two time steps and therefore each contributes  $\Delta t*1$  to the expected lifespan.

The probability that an individual dies as an adult is  $A*B*1$  since any individual that reaches adulthood will die as an adult. The fact that such individuals have survived both the cub and juvenile stages means their age is at least  $\Delta t*2$ . However, once an individual reaches adulthood its chance of survival per time step is  $C$ , and as we saw above this means that the expected number of time steps it spends as an adult is  $1/C$ . Therefore, on average, individuals that die as adults will have lived  $(2+1/C)*\Delta t$ . Putting these calculations together leads to an expected lifespan  $\tau$  of:

$$\tau = A(1 - B)\Delta t + AB \left( 2 + \frac{1}{C} \right) \Delta t$$

Which simplifies to:

$$\tau = A \left( 1 + B \left( 1 + \frac{1}{C} \right) \right) \Delta t$$

As discussed above, this expected lifespan can be used to define the death rate in an *unstructured* continuous time model as  $1/\tau$ .

We have utilised the methods above to estimate the badger parameters as follows.

- Carrying Capacity used in the continuous time model  $K = 20$

European badger social groups documented in the literature range from 1 or 2 (Kruuk & Parish 2009) up to 27 (Rogers *et al.* 1997). Using this information the total carrying capacity used in the simulations presented here was chosen to be 20 to simulate a group large enough (around 18) to calculate meaningful statistics but not so large as to be extreme. We

tested the parameter choice for robustness and found that using a larger carrying capacity had little effect on the overall results.

- Death rate used in the continuous time model  $d = 0.313$

This rate was derived using information from (Shirley *et al.* 2003) and calculating the life expectancy (LE) for Males and Females separately using a geometric distribution, as the model uses discrete time. Female badgers born have a probability 0.76 of dying while in the cub stage per 6 month period, probability 0.771 of dying in the adolescent stage and probability 0.122 of dying in the adult stage, the corresponding probabilities for Males are 0.76, 0.704 and 0.161 respectively. Using information we can calculate the LE for females and males as follows:

$$\text{Female: } \frac{1}{2 \left( 1 + 0.76 \left( 1 + 0.771 \left( 1 + \frac{1}{0.122} \right) \right) \right)} = 3.5745$$

$$\text{Male: } \frac{1}{2 \left( 1 + 0.76 \left( 1 + 0.704 \left( 1 + \frac{1}{0.161} \right) \right) \right)} = 2.8091$$

Averaged over sex, the average lifespan is calculated to be 3.1918. Therefore the annual mortality rate using a geometric distribution with mean  $1/p$  is  $1/3.1918 = 0.313$ .

- Per capita growth rate used in the continuous time model  $r = 5$

Again, using information from (Shirley *et al.* 2003) we find that the average litter size for a breeding pair is 2.97 and the average number of litters per pair is 1.1206. We can deduce therefore that the breeding pair will contribute 3.328 to the population per year. However, in our continuous time model we do not represent breeding structure and therefore need to translate this average number of offspring produced per year into an average annual growth rate per capita. This of course is dependent on the typical population size which in turn depends on the birth rate (given the death rate defined above). We therefore ran a set of simulations in which the birth rate was varied until we obtained a relatively stable population with an average population of 18 which lead to the choice of  $r = 5$ .

- Disease induced mortality rate used in the continuous time model  $\mu_i = 0.165$

A small percentage of badgers infected with TB will become super-excretors. These animals shed a lot more virus and also have a higher mortality rate associated with them. In the Shirley *et al* paper, only super-excretors had added life history consequences, whilst both excretors and super excretors are infectious. Shirley *et al* 2003 quote figurers describing the average length of time individuals spend as excretors and super-excretors which reveals that as a fraction of the time spent infectious 0.6308 was spent as a super-excretor experiencing higher levels of mortality. The mean super excretor mortality over both genders was calculated to be a probability of 0.2415 and therefore we estimate the probability of death for infections individuals to be  $0 \times (1 - 0.6308) + 0.2415 \times 0.6308$ . Therefore, using the formula for conversion between discrete and continuous time models described above our disease induced mortality rate is 0.165

- Disease transmission rate used in the continuous time model  $\beta = 0.057$

From Shirley *et al* 2003, transmission rate for both excretors and super excretors are said to “vary”. Therefore, through simulation of the stochastic model transmission rate for TB has been set at 0.057 to get an average prevalence around 48%. There are various reports of prevalence from badger studies and we we’re aiming for an average between 40% – 60%. Considering information from Zijerveld (2012), the estimated overall death rate = 0.4 but the author did not look at disease induced mortality separately. Therefore the average death rate for the parameters considered in this research should equal  $(n \times 0.313 + I \times 0.165) / n = 0.4 \Rightarrow 0.313 + (I/n) \times 0.165 = 0.4$ , assuming  $n$  individuals of whom  $I$  are infected. This then leads to a prevalence of  $(I/n) = (0.4 - 0.313) / 0.165 = 0.5272 \sim 0.53$ . This is approximately consistent with the average prevalence levels simulated in this research.

### 2.1.2 Estimating Rabbit Parameters

The paper on which the rabbit parameter exploration was based, Judge *et al* 2007, describes an individual based continuous time, stochastic process with state-space defined by sex, age, disease status and location of each animal.

- Carrying Capacity used in the continuous time model  $K = 115$

Using information from Judge *et al* 2007, rabbit social groups have a typical maximum population size of around 90 individuals. We set the carrying capacity based on the birth and death rate so that the population can peak around this number but the equilibrium is below (around 65).

- Per capita death rate used in the continuous time model  $\mu = 5.7$

Using monthly death rates for rabbits at different life stages (i.e. adult, adolescent, infant) we calculated  $((0.0909+0.667+0.25+0.66)/4)*12 = 5.0$  as an approximate death rate not taking into account time spent in each stage. Using simulation results we increased this slightly to 5.7 to produce realistic population fluctuations.

- Intrinsic growth rate used in the continuous time model  $r = 13.5$

From the adult female birth rate in Judge *et al* 2007, a straight forward calculation of  $1.67*12$  (i.e. birth rate multiplied by 12 months) would give approximately 20.0 as the birth rate. However, as the model in Judge *et al* 2007 differentiated sex and age, we might expect this estimate to be a little crude and indeed simulations where we employ the above death rates and carrying capacity, but vary the birth rate to target a typical population size of around 65 unsurprisingly do suggest a lower birth rate. This process led to the selection of a birth rate of 13.5 as this gave a more realistic depiction of rabbit population dynamics.

- Disease transmission rate interval  $\beta = 0.156 - 0.565$

Judge *et al* 2007 employed a statistical estimate of the disease transmission rate per susceptible-infective pair used in the range 0.013-0.046 per month. Therefore, based on these values, the range was calculated as  $0.013*12$  and  $0.046*12 = 0.156 - 0.565$ . Rates chosen were within this calculated interval.

## Appendix 2.2

### 2.2.1 Parameterisations used

This section of the appendix describes in detail the parameter combinations used to produce the graphs in the main text. Values of the form: a,b,c,d etc refer to discrete values used for different lines shown on the Figures. Values of the form a;b;c refer to smallest value; largest value; step size describing the range of values (e.g. of the death rate) simulated to produce the Figures. Values of the form a – b refer to the range of values covered with a non-constant step size. All other parameters with single values are held constant in simulations.

#### Badgers

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.057
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0.165
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1

**Table S.2.1:** Parameter values are shown for part a in both sections of Figure 1 in the main text which demonstrates 1 realisation of the badger population through time using the Gillespie and SDE implementation. The Gillespie example is shown from time 0 until time 25 and the SDE model is shown from time 60 until time 80 to allow a longer burn in period.

## Rabbits

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.156, 0.552
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	0
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1

**Table S.2.2:** Parameter values are shown for part b and c in both sections of Figure 1 in the main text which demonstrates 1 realisation of the rabbit population through time using the Gillespie and SDE implementation. The Gillespie example is shown from time 0 until time 25 and the SDE model is shown from time 60 until time 80 to allow a longer burn in period.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.5, 0.057
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0;2.5;0.05
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10.0

**Table S.2.3:** Parameter values are shown for Figure 3.2 in the main text which demonstrates the effect of the infected death rate and transmission rate on the bias and variance of the prevalence estimate as well as the effect of the death rate on the population size and variance in a badger population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1.0, 0.552, 0.225
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	0.0;7.4;0.2
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10.0

**Table S.2.4:** Parameter values are shown for Figure 3.3 in the main text which demonstrates the effect of the infected death rate and transmission rate on the bias and variance of the prevalence estimate as well as the effect of the death rate on the population size and variance in a rabbit population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.005;1.5;0.005
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0,1,2
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10

**Table S.2.5:** Parameter values are shown for Figure 3.4.a and 3.4.b in the main text which demonstrates the effect of the transmission rate at three fixed level of disease induced mortality in a badger population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.05;1.5;0.05
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	0,2,3
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10

**Table S.2.6:** Parameter values are shown for Figure 3.4.c and 3.4.d in the main text which demonstrates the effect of the transmission rate at three fixed level of disease induced mortality in a rabbit population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	1.75
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1;10;0.1
Sample Target	$m$	10

**Table S.2.7:** Parameter values are shown for Figure 3.5 in the main text which demonstrates the effect of the capture rate on surveillance efficacy a badger population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	4.4
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1;10;0.1
Sample Target	$m$	10

**Table S.2.8:** Parameter values are shown for Figure 3.6 in the main text which demonstrates the effect of the capture rate on surveillance efficacy in a rabbit population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1,0.5,0.057
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0;2.5;0.05
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10

**Table S.2.9:** Parameter values are shown for Figure 3.7.a in the main which demonstrates the effect of the transmission rate and disease induced mortality on the probability of detection in a badger population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.5
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0;2.5;0.05
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1, 1, 2
Sample Target	$m$	10

**Table S.2.10:** Parameter values are shown for Figure 3.7.b in the main text which demonstrates the effect of the capture rate and disease induced mortality on the probability of detection in a badger population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	1, 0.552, 0.225
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	0.0;7.4;0.2
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1
Sample Target	$m$	10

**Table S.2.11:** Parameter values are shown for Figure 3.7.c in the main text which demonstrates the effect of the transmission rate and disease induced mortality on the probability of detection in a rabbit population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.552
Carrying Capacity	$k$	115
Growth Rate	$r$	13.5
Death Rate	$\mu$	5.7
Infected Death Rate	$\mu_i$	0.0;7.4;0.2
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1, 1, 2
Sample Target	$m$	10

**Table S.2.12:** Parameter values are shown for Figure 3.7.d in the main text which demonstrates the effect of the capture rate and disease induced mortality on the probability of detection in a rabbit population. 100,000 simulations are run of each combination and terminate when the sample target is reached, i.e. there is no time limit imposed. These parameters were implemented using the SDE version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.0025;0.75;0.0025
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0.165
Immigration	$\nu$	0.1
Infected Immigration Proportion	$\gamma$	0.1
Primary Transmission Rate	$\beta_0$	0.1
Active Capture	$\alpha$	0.1;1;0.1
Sample Target	$m$	10
Trappability Rate	$\tau$	0.5,1, 2

**Table S.2.13:** Parameter values are shown for Figure 3.8 in the main text which demonstrates the effect of trappability and transmission rate, as well as the sample size and capture rate, on the probability of detecting disease. These parameters were implemented using the Gillespie version of the model. 1000 simulations were run per parameter combination with a time limit of 45. If the simulation did not reach the sample target within the time limit, the run is discarded and not used in the statistical calculations. If out of 1000 realisations a parameter combination ceases to reach the sample target at least 15 times, that parameter combination is discarded totally as the results are deemed to be unreliable. Increasing the time limit bears little to no effect on the amount simulations which reach the target sample, so the precise value of the time limit does not affect the results obtained from the model. These parameters were implemented using the Gillespie version of the model.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.25,0.057,0.025
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0.165
Immigration	$\nu$	1.0
Infected Immigration Proportion	$\gamma$	0.01
Primary Transmission Rate	$\beta_0$	0.01
Active Capture	$\alpha$	0.5
Sample Target	$m$	10.0
Test Sensitivity	$s$	0;1;0.01

**Table S.2.14:** Parameter values are shown for Figure 3.9 in the main text which demonstrates the effect of the sensitivity of the test at three different transmission rates on the probability detecting disease and the bias in estimating prevalence in a badger population. Specificity is 1.0 in all simulations. 1000 simulations were run per parameter combination with a time limit of 45. See description of Table S13 for simulation details.

Rate Name	Rate	Value
Secondary Transmission Rate	$\beta$	0.0025;0.75;0.0025
Carrying Capacity	$k$	20
Growth Rate	$r$	5
Death Rate	$\mu$	0.313
Infected Death Rate	$\mu_i$	0.165
Immigration	$\nu$	1.0
Infected Immigration Proportion	$\gamma$	0.01
Primary Transmission Rate	$\beta_0$	0.01
Susceptible Active Capture	$\alpha$	0.5
Sample Target	$m$	10.0
Test Sensitivity	$s$	0.2,0.6,1.0

**Table S.2.15:** Parameter values are shown for Figure 3.10 in the main text which demonstrates the effect of the transmission rate at three different test sensitivities on the probability detecting disease and the bias in estimating prevalence in a badger population. 1000 simulations were run per parameter combination with a time limit of 45. See description of Table S13 for simulation details.

# Appendix 3

## 3.1 Parameterisations used

This section of the appendix describes in detail the parameter combinations used to produce the graphs in the main text. Values of the form: a,b,c,d etc refer to discrete values used for different lines shown on the Figures. Values of the form a;b;c refer to smallest value; largest value; step size describing the range of values (e.g. of the death rate) simulated to produce the Figures. Values of the form a – b refer to the range of values covered with a non-constant step size. All other parameters with single values are held constant in simulations.

The total number of patches,  $L$ , is set constant at 130. The simulation has a “burn in” period of 10, at which point a disease incursion event is triggered in a randomly selected patch of any suitability type. After a single infected individual is introduced into a random patch; the statistics are collected until time of first detection or time  $t_{max}$ . The proportion of favourable to less favourable habitat is set at 50/50 unless otherwise stated and the habitat suitability level of the favourable habitat type is fixed at 1.0.

Rate Name	Rate	Value
<b>Within Patch Transmission Rate</b>	$\beta_0$	0.5
<b>Maximum Carrying Capacity</b>	$k$	50
<b>Growth Rate</b>	$r$	20
<b>Death Rate</b>	$\mu$	1.5
<b>Immigration</b>	$\nu$	1.0
<b>Infected Immigration Proportion</b>	$\gamma$	0.1
<b>Between Patch Transmission Rate</b>	$\beta_1$	0.1
<b>Capture rate</b>	$\alpha$	0.1, 0.25, 1.0
<b>Switching Rate</b>	$\tau$	0
<b>Less Favourable habitat Suitability index</b>	$\sigma$	0.5
<b>Less favourable habitat proportion</b>	$\rho$	0.5
<b>Number of patches under surveillance</b>	$n$	1;65;1

**Table S.3.1:** Parameter values are shown for Figure 4.1 in the main text which demonstrates how the number of patches included in the surveillance set affects the level of disease in the spatial system at the point of first detection.

Rate Name	Rate	Value
Within Patch Transmission Rate	$\beta_0$	0.5
Maximum Carrying Capacity	$k$	50
Growth Rate	$r$	20
Death Rate	$\mu$	1.5
Immigration	$\nu$	1.0
Infected Immigration Proportion	$\gamma$	0.1
Between Patch Transmission Rate	$\beta_1$	0.1
Capture rate	$\alpha$	1.0
Switching Rate	$\tau$	0
Less Favourable habitat Suitability index	$\sigma$	0.02;1.0;0.02
Less favourable habitat proportion	$\rho$	0.5
Number of patches in surveillance set	$n$	1, 3, 7, 33

**Table S.3.2:** Parameter values are shown for Figure 4.2 in the main text which demonstrates how the suitability index affects the level of disease in the spatial system at the point of first detection.

Rate Name	Rate	Value
Within Patch Transmission Rate	$\beta_0$	0.5
Maximum Carrying Capacity	$k$	50
Growth Rate	$r$	20
Death Rate	$\mu$	1.5
Immigration	$\nu$	1.0
Infected Immigration Proportion	$\gamma$	0.1
Between Patch Transmission Rate	$\beta_1$	0.1
Capture rate	$\alpha$	1.0
Switching Rate	$\tau$	0
Favourable habitat Suitability index	$\sigma$	1.0
Favourable habitat Suitability index	$\sigma$	0.5
Less favourable habitat proportion	$\rho$	0.02;1.0;0.02
Number of patches in surveillance set	$n$	1

**Table S.3.3:** Parameter values are shown for Figure 4.3 in the main text which demonstrates how the proportion of good and bad habitat, as well as the type of patch targeted by surveillance, affects the level of disease in the spatial system at the point of first detection.

Rate Name	Rate	Value
Within Patch Transmission Rate	$\beta_0$	1.0, 2.0, 5.0
Maximum Carrying Capacity	$k$	50
Growth Rate	$r$	20
Death Rate	$\mu$	1.5
Immigration	$\nu$	1.0
Infected Immigration Proportion	$\gamma$	0.1
Between Patch Transmission Rate	$\beta_1$	1.0, 2.0, 5.0
Capture rate	$\alpha$	0.02;1.0;0.02
Switching Rate	$\tau$	0
Favourable habitat Suitability index	$\sigma$	1.0
Favourable habitat Suitability index	$\sigma$	0.5
Less favourable habitat proportion	$\rho$	0.5
Number of patches in surveillance set	$n$	1

**Table S.3.4:** Parameter values are shown for Figure 4.4 in the main text which demonstrates how the effort (capture rate) of surveillance and the within patch and between patch transmission rate affects the level of disease in the spatial system at the point of first detection.

<b>Rate Name</b>	<b>Rate</b>	<b>Value</b>
<b>Within Patch Transmission Rate</b>	$\beta_0$	0.5
<b>Maximum Carrying Capacity</b>	$k$	50
<b>Growth Rate</b>	$r$	20
<b>Death Rate</b>	$\mu$	1.5
<b>Immigration</b>	$\nu$	1.0
<b>Infected Immigration Proportion</b>	$\gamma$	0.1
<b>Between Patch Transmission Rate</b>	$\beta_1$	0.5
<b>Capture rate</b>	$\alpha$	0.1, 0.25, 1.0
<b>Switching Rate</b>	$\tau$	0.02;1.0;0.02
<b>Favourable habitat Suitability index</b>	$\sigma$	1.0
<b>Favourable habitat Suitability index</b>	$\sigma$	0.5
<b>Less favourable habitat proportion</b>	$\rho$	0.5
<b>Number of patches in surveillance set</b>	$n$	1, 3, 7, 33

**Table S.3.5:** Parameter values are shown for Figure 4.5 in the main text which demonstrates how the time before switching of surveillance and percentage number of patches in the surveillance set affects the level of disease in the spatial system at the point of first detection.

<b>Rate Name</b>	<b>Rate</b>	<b>Value</b>
<b>Within Patch Transmission Rate</b>	$\beta_0$	0.5
<b>Maximum Carrying Capacity</b>	$k$	50
<b>Growth Rate</b>	$r$	20
<b>Death Rate</b>	$\mu$	1.5
<b>Immigration</b>	$\nu$	1.0
<b>Infected Immigration Proportion</b>	$\gamma$	0.1
<b>Between Patch Transmission Rate</b>	$\beta_1$	0.1
<b>Capture rate</b>	$\alpha$	0.02;1.0;0.02
<b>Switching Rate</b>	$\tau$	0, 0.01, 0.1, 1.0
<b>Favourable habitat Suitability index</b>	$\sigma$	1.0
<b>Favourable habitat Suitability index</b>	$\sigma$	0.5
<b>Less favourable habitat proportion</b>	$\rho$	0.5
<b>Number of patches in surveillance set</b>	$n$	1

**Table S.3.6:** Parameter values are shown for Figure 4.6 in the main text which demonstrates the difference between switching (at different rates) and not switching and the effect of the effort put into surveillance has on the level of disease in the spatial system at the point of first detection.

<b>Rate Name</b>	<b>Rate</b>	<b>Value</b>
<b>Within Patch Transmission Rate</b>	$\beta_0$	0.5
<b>Maximum Carrying Capacity</b>	$k$	50
<b>Growth Rate</b>	$r$	20
<b>Death Rate</b>	$\mu$	1.5
<b>Immigration</b>	$\nu$	1.0
<b>Infected Immigration Proportion</b>	$\gamma$	0.1
<b>Between Patch Transmission Rate</b>	$\beta_1$	0.02;1.0;0.02
<b>Capture rate</b>	$\alpha$	1.0
<b>Switching Rate</b>	$\tau$	0, 0.01, 0.1, 1.0
<b>Favourable habitat Suitability index</b>	$\sigma$	1.0
<b>Favourable habitat Suitability index</b>	$\sigma$	0.5
<b>Less favourable habitat proportion</b>	$\rho$	0.5
<b>Number of patches in surveillance set</b>	$n$	1

**Table S.3.7:** Parameter values are shown for Figure 4.7 in the main text which demonstrates the difference between switching (at different rates) and not switching and the effect of between patch transmission has on the level of disease in the spatial system at the point of first detection.

# References

1.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M. & Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecol. Lett.*, 9, 467–84.

2.

Amanfu, W. (2006). The situation of tuberculosis and tuberculosis control in animals of economic interest. *Tuberculosis (Edinb.)*, 86, 330–5.

3.

Anderson, P.K., Cunningham, A. a, Patel, N.G., Morales, F.J., Epstein, P.R. & Daszak, P. (2004). Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol. Evol.*, 19, 535–44.

4.

Anderson, R. (1991). Populations and infectious diseases: ecology or epidemiology? *J. Anim. Ecol.*, 60, 1–50.

5.

Anderson, R. & May, R. (1979). Population biology of infectious diseases: Part I. *Nature*, 280, 361–367.

6.

Antle, J., Capalbo, S., Mooney, S., Elliott, E. & Paustian, K. (2003). Spatial heterogeneity, contract design, and the efficiency of carbon sequestration policies for agriculture. *J. Environ. Econ. Manage.*, 46, 231–250.

7.

Armstrong, L.R., Zaki, S.R., Goldoft, M.J., Todd, R.L., Khan, A.S., Khabbaz, R.F., *et al.* (1995). Hantavirus Pulmonary Syndrome Associated with Entering or Cleaning Rarely Used, Rodent-Infested Structures. *J. Infectious Dis.*, 172, 1166.

8.

Artois, M., Bengis, R., Delahay, R.J., Duchêne, M.-J., Duff, J.P., Ferroglio, E., *et al.* (2009a). *Management of disease in wild Mammals*. Springer Japan, Tokyo.

9.

Artois, M., Bicolout, D., Doctrinal, D., Fouchier, R., Gavier-Widen, D., Globig, a, *et al.* (2009b). Outbreaks of highly pathogenic avian influenza in Europe: the risks associated with wild birds. *Rev. Sci. Tech.*, 28, 69–92.

10.

Artois, M., Delahay, R., Guberti, V. & Cheeseman, C. (2001). Control of infectious diseases of wildlife in Europe. *Vet. J.*, 162, 141–52.

11.

Awais, D., Siegel, C. a & Higgins, P.D.R. (2009). Modelling dysplasia detection in ulcerative colitis: clinical implications of surveillance intensity. *Gut*, 58, 1498–503.

12.

Beard, P.M., Rhind, S.M., Buxton, D., Daniels, M.J., Henderson, D., Pirie, a, *et al.* (2001). Natural paratuberculosis infection in rabbits in Scotland. *J. Comp. Pathol.*, 124, 290–9.

13.

Belant, A.R. & Deese, J.L. (2010). The importance of wildlife disease surveillance. *Human-Wildlife Interact.*, 4, 165–169.

14.

Béneult, B., Ciliberti, A. & Artois, M. (2014). A Generic Action Plan against the Invasion of the EU by an Emerging Pathogen in Wildlife-A WildTech Perspective. *Planet@ Risk*, 2, 174–181.

15.

Bengis, R.G., Leighton, F.A., Fischer, J.R., Artois, M. & Mörner, T. (2004). The role of wildlife in emerging and re-emerging zoonoses Recent emerging zoonoses Viral zoonoses, 23, 497–511.

16.

Bidet, P., Liguori, S., De Lauzanne, A., Caro, V., Lorrot, M., Carol, A., *et al.* (2008). Real-time PCR measurement of persistence of *Bordetella pertussis* DNA in nasopharyngeal secretions during antibiotic treatment of young children with pertussis. *J. Clin. Microbiol.*, 46, 3636–8.

17.

Blanchong, J. a, Samuel, M.D., Scribner, K.T., Weckworth, B. V, Langenberg, J. a & Filcek, K.B. (2008). Landscape genetics and the spatial distribution of chronic wasting disease. *Biol. Lett.*, 4, 130–3.

18.

Van den Brom, R., Luttikholt, S.J.M., Lievaart-Peterson, K., Peperkamp, N.H.M.T., Mars, M.H., van der Poel, W.H.M., *et al.* (2012). Epizootic of ovine congenital malformations associated with Schmallenberg virus infection. *Tijdschr. Diergeneeskd.*, 137, 106–11.

19.

Butler, D. (2006). Disease surveillance needs a revolution. *Nature*, 440, 6–7.

20.

Byrne, A.W., O’Keeffe, J., Green, S., Sleeman, D.P., Corner, L. a L., Gormley, E., *et al.* (2012a). Population estimation and trappability of the European badger (*Meles meles*): implications for tuberculosis management. *PLoS One*, 7, e50807.

21.

Byrne, A.W., Sleeman, D.P., O’Keeffe, J. & Davenport, J. (2012b). The Ecology of the European Badger (*Meles Meles*) in Ireland - A review. *Biol. Environ.*, 112, 105 – 132.

22.

Caffrey, J.P. (1994). Status of bovine tuberculosis eradication programmes in Europe. *Vet. Microbiol.*, 40, 1–4.

23.

Calvete, C. (2006). Modeling the Effect of Population Dynamics on the Impact of Rabbit Hemorrhagic Disease. *Conserv. Biol.*, 20, 1232–1241.

24.

Carter, S.P., Delahay, R.J., Smith, G.C., Macdonald, D.W., Riordan, P., Etherington, T.R., *et al.* (2007). Culling-induced social perturbation in Eurasian badgers *Meles meles* and the management of TB in cattle: an analysis of a critical problem in applied ecology. *Proc. Biol. Sci.*, 274, 2769–77.

25.

Chua, K.B., Goh, K.J., Wong, K.T., Kamarulzaman, A., Seow, P., Tan, K., *et al.* (1999). Fatal encephalitis due to Nipah virus among pig-farm e rs in M a l a y s i a, 354, 1257–1259.

26.

Cliff, A.D. (1995). *Incorporating spatial components into models of epidemic spread, in Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press.

27.

Coltherd, J.C., Morgan, C., Judge, J., Smith, L. a. & Hutchings, M.R. (2010). The effects of parasitism on recapture rates of wood mice ( *Apodemus sylvaticus* ). *Wildl. Res.*, 37, 413.

28.

Condy, J.B., Herniman, K.A.J. & Hedger, R.S. (1969). Foot-and-Mouth Disease in wildlife in Rhodesia and other African territories. *J. Comp. Pathol.*, 79, 27–31.

29.

Conraths, F.J., Peters, M. & Beer, M. (2013). Schmallenberg virus, a novel orthobunyavirus infection in ruminants in Europe: potential global impact and preventive measures. *N. Z. Vet. J.*, 61, 63–7.

30.

Consul, P.C., Kotz, S. & Famoye, F. (2006). *Lagrangian Probability Distributions*. Springer Science & Business Media.

31.

Cooke, B.D. (2002). Rabbit haemorrhagic disease: field epidemiology and the management of wild rabbit populations. *Rev. Sci. Tech.*, 21, 347–58.

32.

Cottam, E.M., Wadsworth, J., Shaw, A.E., Rowlands, R.J., Goatley, L., Maan, S., *et al.* (2008). Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathog.*, 4, e1000050.

33.

Cunningham, A.A. (1996). Translocations of Wildlife Disease Risks. *Conserv. Biol.*, 10, 349–353.

34.

D'Amato, G., Baena-Cagnani, C.E., Cecchi, L., Annesi-Maesano, I., Nunes, C., Ansotegui, I., *et al.* (2013). Climate change, air pollution and extreme events leading to increasing prevalence of allergic respiratory diseases. *Multidiscip. Respir. Med.*, 8, 12.

35.

Daniels, M.J., Ball, N., Hutchings, M.R. & Greig, A. (2001). The grazing response of cattle to pasture contaminated with rabbit faeces and the implications for the transmission of paratuberculosis. *Vet. J.*, 161, 306–13.

36.

Daszak, P., Berger, L., Cunningham, a a, Hyatt, a D., Green, D.E. & Speare, R. (1999). Emerging infectious diseases and amphibian population declines. *Emerg. Infect. Dis.*, 5, 735–48.

37.

Daszak, P., Cunningham, a a & Hyatt, a D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science*, 287, 443–9.

38.

Davidson, R.S., Marion, G. & Hutchings, M.R. (2008). Effects of host social hierarchy on disease persistence. *J. Theor. Biol.*, 253, 424–33.

39.

DEFRA. (2011). Bovine TB eradication programme for England. *Vet. Rec.*, 169, 689.

40.

DEFRA. (2013). Defra, UK - Badger control – culling of badgers [WWW Document]. *www.defra.gov.uk*. URL <http://www.defra.gov.uk/animal-diseases/a-z/bovine-tb/badgers/culling/>.

41.

DEFRA. (2013). Bovine TB (tuberculosis) [WWW Document]. URL <http://www.defra.gov.uk/animal-diseases/a-z/bovine-tb/>.

42.

Delahay, R.J., Langton, S., Smith, G.C., Clifton-Hadley, R.S. & Cheeseman, C.L. (2000). The spatio-temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density badger population. *J. Anim. Ecol.*, 69, 428–441.

43.

Department for Environment, F. and R.A. (Defra) [webmaster@defra.gsi.gov.uk](mailto:webmaster@defra.gsi.gov.uk). (n.d.). Defra, UK - Schmallenberg virus.

44.

Dohoo, I., Martin, W. & Stryhn, H. (2005). Veterinary Epidemiologic Research. *Prev. Vet. Med.*, 68, 289–292.

45.

Donnelly, C. a, Woodroffe, R., Cox, D.R., Bourne, F.J., Cheeseman, C.L., Clifton-Hadley, R.S., *et al.* (2006). Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature*, 439, 843–6.

46.

Donnelly, C., Woodroffe, R. & Cox, D. (2003). Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature*, 426.

47.

Donnelly, C.A. & Hone, J. (2010). Is there an association between levels of bovine tuberculosis in cattle herds and badgers ? *Stat. Commun. Infect. Dis.*

48.

Dórea, F.C., Sanchez, J. & Revie, C.W. (2011). Veterinary syndromic surveillance: Current initiatives and potential for development. *Prev. Vet. Med.*, 101, 1–17.

49.

Dowell, S.F. (2001). Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg. Infect. Dis.*, 7, 369–74.

50.

ECDC. (2012). European Food Safety Authority publishes its second report on the Schmallenberg virus.

51.

Eckert, J. & Deplazes, P. (2004). Biological , Epidemiological , and Clinical Aspects of Echinococcosis , a Zoonosis of Increasing Concern. *Clin. Microbiol. Rev.*, 17, 107–135.

52.

EFSA. (2012). “ Schmallenberg ” virus : Analysis of the Epidemiological Data and Assessment of Impact.

53.

Elbers, A.R.W., Loeffen, W.L.A., Quak, S., Boer-luijtz, E. De, Spek, A.N. Van Der, Bouwstra, R., *et al.* (2012). Seroprevalence of Schmallenberg Virus Antibodies among Dairy, 18, 1065–1071.

54.

Epps, C.W., Palsbøll, P.J., Wehausen, J.D., Roderick, G.K., Ramey, R.R. & McCullough, D.R. (2005). Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecol. Lett.*, 8, 1029–1038.

55.

Epstein, P.R. (1995). Emerging diseases and ecosystem instability: new threats to public health. *Am. J. Public Health*, 85, 168–72.

56.

Evenson, D. (2008). Wildlife disease can put conservation at risk. *Nature*, 452.

57.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 34, 487–515.

58.

FAO. (2011). *Challenges of animal health information systems and surveillance for animal diseases and zoonoses*.

59.

Favier, C., Schmit, D., Müller-Graf, C.D.M., Cazelles, B., Degallier, N., Mondet, B., *et al.* (2005). Influence of spatial heterogeneity on an emerging infectious disease: the case of dengue epidemics. *Proc. Biol. Sci.*, 272, 1171–7.

60.

Firdessa, R., Tschopp, R., Wubete, A., Sombo, M., Hailu, E., Erenso, G., *et al.* (2012). High prevalence of bovine tuberculosis in dairy cattle in central ethiopia: implications for the dairy industry and public health. *PLoS One*, 7, e52851.

61.

Fosgate, G.T. (2005). Modified exact sample size for a binomial proportion with special emphasis on diagnostic test parameter estimation. *Stat. Med.*, 24, 2857–66.

62.

Fosgate, G.T. (2009). Practical Sample Size Calculations for Surveillance and Diagnostic Investigations. *J. Vet. Diagnostic Investig.*, 21, 3–14.

63.

Frölich, K., Thiede, S., Kozikowski, T. & Jakob, W. (2002). A Review of Mutual Transmission of Important Infectious Diseases between Livestock and Wildlife in Europe. *Ann. N.Y. Acad. Sci.*, 4–13.

64.

Fulford, G.R., Roberts, M.G. & Heesterbeek, J. a P. (2002). The metapopulation dynamics of an infectious disease: tuberculosis in possums. *Theor. Popul. Biol.*, 61, 15–29.

65.

Genovesi, P. & Shine, C. (2002). Convention on the conservation of european wildlife and natural habitats. In: *Eur. Strateg. Invasive Alien Species*. Strasbourg.

66.

Gillespie, D.T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22, 403–434.

67.

Girard, M.P., Tam, J.S., Assossou, O.M. & Kieny, M.P. (2010). The 2009 A (H1N1) influenza virus pandemic: A review. *Vaccine*, 28, 4895–902.

68.

Gortázar, C., Ferroglio, E., Höfle, U., Frölich, K. & Vicente, J. (2007). Diseases shared between wildlife and livestock: a European perspective. *Eur. J. Wildl. Res.*, 53, 241–256.

69.

Greig, A., Stevenson, K. & Perez, V. (1997). Paratuberculosis in wild rabbits (*Oryctolagus cuniculus*). *Vet. Rec.*, 140, 141 – 143.

70.

Grimes, D.A. & Schulz, K.F. (1996). Determining sample size and power in clinical trials: the forgotten essential. *Semin. Reprod. Endocrinol.*, 14, 125–31.

71.

Hadorn, D.C. & Stärk, K.D.C. (2008). Evaluation and optimization of surveillance systems for rare and emerging infectious diseases. *Vet. Res.*, 39–57.

72.

Hagenaars, T.J., Donnelly, C. a & Ferguson, N.M. (2004). Spatial heterogeneity and the persistence of infectious diseases. *J. Theor. Biol.*, 229, 349–59.

73.

Harris, K. a, Eglin, R.D., Hayward, S., Milnes, a, Davies, I., Cook, a J.C., *et al.* (2014). Impact of Schmallenberg virus on British sheep farms during the 2011/2012 lambing season. *Vet. Rec.*, 175, 172.

74.

Hastings, A. (1977). Spatial heterogeneity and the stability of predator-prey systems. *Theor. Popul. Biol.*, 48, 37–48.

75.

Hawkins, C., McCallum, H., Mooney, N., Jones, M. & Holdsworth, M. (2008). *Sarcophilus harrisi* (Tasmanian Devil) [WWW Document]. *IUCN 2013. IUCN Red List Threat. Species*. URL <http://www.iucnredlist.org/details/full/40540/0>.

76.

Haydon, D.T. (2008). Cross-disciplinary demands of multihost pathogens. *J. Anim. Ecol.*, 77, 1079–81.

77.

Hess, G. (1996). Disease in metapopulation models: implications for conservation. *Ecology*, 77, 1617–1632.

78.

Higham, D.J. (2001). An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, 43, 525–546.

79.

Huffaker, C.B. (1958). Experimental Studies on Predation: Dispersion Factors and Predator-Prey Oscillations. *Hilgardia*, 27, 795 – 835.

80.

Jebara, K. Ben. (2004). Surveillance , detection and response : managing emerging diseases at national and international levels, 23, 709–715.

81.

Jones, K.E., Patel, N.G., Levy, M. a, Storeygard, A., Balk, D., Gittleman, J.L., *et al.* (2008). Global trends in emerging infectious diseases. *Nature*, 451, 990–3.

82.

Jones, M.E., Paetkau, D., Geffen, E. & Moritz, C. (2004). Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. *Mol. Ecol.*, 13, 2197–209.

83.

Judge, J., Davidson, R.S., Marion, G., White, P.C.L. & Hutchings, M.R. (2007). Persistence of *Mycobacterium avium* subspecies paratuberculosis in rabbits: the interplay between horizontal and vertical transmission. *J. Appl. Ecol.*, 44, 302–311.

84.

Judge, J., Greig, A., Kyriazakis, I. & Hutchings, M.R. (2005). Ingestion of faeces by grazing herbivores—risk of inter-species disease transmission. *Agric. Ecosyst. Environ.*, 107, 267–274.

85.

Judge, J., Kyriazakis, I., Greig, A., Davidson, R.S. & Hutchings, M.R. (2006). paratuberculosis in Rabbits (*Oryctolagus cuniculus*): a Field Study, 72, 398–403.

86.

Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theor. Popul. Biol.*, 67, 1–8.

87.

Keeling, M.J. (1999). The effects of local spatial structure on epidemiological invasions. *Proc. Biol. Sci.*, 266, 859–67.

88.

Keeling, M.J. & Grenfell, B.T. (1998). Effect of variability in infection period on the persistence and spatial spread of infectious diseases. *Math. Biosci.*, 147, 207–26.

89.

Keeling, M.J. & Rohani, P. (2007). *Modeling Infectious Diseases in Humans and Animals*. 1st edn. Princeton University Press.

90.

Keeling, M.J. & Ross, J. V. (2008). On methods for studying stochastic disease dynamics. *J. R. Soc. Interface*, 5, 171–81.

91.

Keeling, M.J., Wilson, H.B. & Pacala, S.W. (2000). Space , Reinterpreting Lags , Functional Responses Models Ecological, 290, 1758–1761.

92.

Kruse, H., Kirkemo, A.-M. & Handeland, K. (2004). Wildlife as source of zoonotic infections. *Emerg. Infect. Dis.*, 10, 2067–72.

93.

Kruuk, H. & Parish, T. (2009). Factors affecting population density, group size and territory size of the European badger, *Meles meles*. *J. Zool.*, 196, 31–39.

94.

Kuiken, T., Ryser-Degiorgis, M.P., Gavier-Widen, D. & Gortázar, C. (2011). Establishing a European network for wildlife, 30, 755–761.

95.

Lancoua, J.B., Homelb, B.B.C. & Elottoc, A.B. (2005). Emerging or re-emerging bacterial zoonoses : factors of emergence , surveillance and control, 36, 507–522.

96.

Lederberg, J., Shope, R.E. & Oaks, S.C. (1992). Emerging Infections: Microbial Threats to health in the United States. *Natl. Acad. Press*, 298.

97.

Leisewitz, a L., Carter, a, van Vuuren, M. & van Blerk, L. (2001). Canine distemper infections, with special reference to South Africa, with a review of the literature. *J. S. Afr. Vet. Assoc.*, 72, 127–36.

98.

Lipkin, W.I. (2013). The changing face of pathogen discovery and surveillance. *Nat. Rev. Microbiol.*, 11, 133–41.

99.

Lloyd, A.L. & May, R.M. (1996). Spatial Heterogeneity in Epidemic Models. *J. Theor. Biol.*, 179, 1–11.

100.

López-Olvera, J.R., Falconi, C., Fernández-Pacheco, P., Fernández-Pinero, J., Sánchez, M.A., Palma, A., *et al.* (2010). Experimental infection of European red deer (*Cervus elaphus*) with bluetongue virus serotypes 1 and 8. *Vet. Microbiol.*, 145, 148–52.

101.

Lucio-Arias, D. & Scharnhorst, A. (2012). *Models of Science Dynamics. Understanding Complex Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg.

102.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press.

103.

Mao, X. (1997). *Stochastic Differential Equations and Applications*. 2nd edn. Horwood, New York.

104.

Marca, E. La, Lips, K.R., Lotters, S., Puschendorf, R., Ibanez, R., Rueda-Almonacid, J.V., *et al.* (2005). Catastrophic Population Declines and Extinctions in Neotropical Harlequin Frogs ( *Bufo* : *Atelopus* ) 1, 37, 190–201.

105.

Marion, G., Renshaw, E. & Gibson, G. (2000). Stochastic modelling of environmental variation for biological populations. *Theor. Popul. Biol.*, 57, 197–217.

106.

Martin, S.W., Meek, A.H. & Willeberg, P. (1987). *Veterinary epidemiology. Principles and methods*. Iowa State University Press.

107.

McCallum, H. (2008). Tasmanian devil facial tumour disease: lessons for conservation biology. *Trends Ecol. Evol.*, 23, 631–7.

108.

McCallum, H., Tompkins, D.M., Jones, M., Lachish, S., Marvanek, S., Lazenby, B., *et al.* (2007). Distribution and Impacts of Tasmanian Devil Facial Tumor Disease. *Ecohealth*, 4, 318–325.

109.

Mörner, T., Obendorf, D.L., Artois, M. & Woodford, M.H. (2002). Surveillance and monitoring of wildlife diseases. *Rev. Sci. Tech.*, 21, 67–76.

110.

Morse, S.S. (1995). Factors in the emergence of infectious diseases. *Emerg. Infect. Dis.*, 1, 7–15.

111.

Murphy, F. a. (2008). Emerging zoonoses: the challenge for public health and biodefense. *Prev. Vet. Med.*, 86, 216–23.

112.

Murray, J.D. (2002). *Mathematical Biology: I. An Introduction*. Springer.

113.

Neal, E. & Cheeseman, C. (1996). *Badgers*. Poyser Natural History.

114.

Nishiura, H. & Chowell, G. (2014). Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. *Euro Surveill.*, 19, 1–6.

115.

Nor, M.N.M., Gan, C.H. & Ong, B.L. (2000). Nipah virus infection of pigs in peninsular Malaysia Epidemiological findings, 19, 160–165.

116.

Nusser, S.M., Clark, W.R., Otis, D.L. & Huang, L. (2008). Sampling Considerations for Disease Surveillance in Wildlife Populations. *J. Wildl. Manage.*, 72, 52–60.

117.

OIE. (2008). *REPORT OF THE MEETING OF THE OIE AQUATIC ANIMAL HEALTH STANDARDS COMMISSION*.

118.

OIE. (2013). Terrestrial Animal Health Code [WWW Document]. URL [http://www.oie.int/fileadmin/Home/eng/Health\\_standards/tahc/2010/en\\_chapitre\\_1.1.4.htm](http://www.oie.int/fileadmin/Home/eng/Health_standards/tahc/2010/en_chapitre_1.1.4.htm).

119.

Ostfeld, R.S., Glass, G.E. & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol. Evol.*, 20, 328–36.

120.

Pagán, I., González-Jara, P., Moreno-Letelier, A., Rodelo-Urrego, M., Fraile, A., Piñero, D., *et al.* (2012). Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system. *PLoS Pathog.*, 8, e1002796.

121.

PHE. (2014). *Risk assessment of the Ebola outbreak in West Africa*.

122.

Prentice, J.C., Marion, G., White, P.C.L., Davidson, R.S. & Hutchings, M.R. (2014). Demographic processes drive increases in wildlife disease following population reduction. *PLoS One*, 9, e86563.

123.

Read, J.M. & Keeling, M.J. (2003). Disease evolution on networks: the role of contact structure. *Proc. Biol. Sci.*, 270, 699–708.

124.

Renshaw, E. (1991). *Modelling Biological Populations in Space and Time*. Cambridge University Press.

125.

Retallick, R.W.R., McCallum, H. & Speare, R. (2004). Endemic infection of the amphibian chytrid fungus in a frog community post-decline. *PLoS Biol.*, 2, e351.

126.

Rhyan, J.C. & Spraker, T.R. (2010). Emergence of diseases from wildlife reservoirs. *Vet. Pathol.*, 47, 34–9.

127.

Roelke-Parker, M.E., Munson, L., Packer, C., Kock, R., Cleaveland, S., Carpenter, M., *et al.* (1996). A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature*, 379, 441–445.

128.

Rogers, L., Cheeseman, C., Mallinson, P. & Clifton-Hadley, R. (1997). The demography of a high-density badger (*Meles meles*) population in the west of England. *J. Zool.*, 242, 705–728.

129.

Ryser-Degiorgis, M.-P. (2013). Wildlife health investigations: needs, challenges and recommendations. *BMC Vet. Res.*, 9, 223.

130.

Sattenspiel, L. & Simon, C.P. (1988). The spread and persistence of infectious diseases in structured populations. *Math. Biosci.*, 90, 341–366.

131.

Schmid, B., Balvanera, P., Cardinale, B.J., Godbold, J., Andrea, B.P., Raffaelli, D., *et al.* (2009). Ecosystem Functioning, and Human Wellbeing An Ecological and Economic Perspective of data from biodiversity experiments. In: *Biodiversity, Ecosyst. Funct. Hum. Wellbeing An Ecol. Econ. Perspect.* Oxford University Press, pp. 14–29.

132.

Scottish Government, S.A.H.R.R.E.E. 3DG T. 556 8400 ceu@scotland. gsi. gov. u. (2008). Foot and Mouth Disease Review 2007: Economic Impact in Scotland.

133.

Shaughnessy, L.J., Smith, L. a, Evans, J., Anderson, D., Caldow, G., Marion, G., *et al.* (2013). High prevalence of paratuberculosis in rabbits is associated with difficulties in controlling the disease in cattle. *Vet. J.*, 198, 267–70.

134.

Shirley, M.D.F., Rushton, S.P., Smith, G.C., South, A.B. & Lurz, P.W.W. (2003). Investigating the spatial dynamics of bovine tuberculosis in badger populations: evaluating an individual-based simulation model. *Ecol. Modell.*, 167, 139–157.

135.

Shoham, D. (2011). The modes of evolutionary emergence of primal and late pandemic influenza virus strains from viral reservoir in animals: an interdisciplinary analysis. *Influenza Res. Treat.*

136.

Smith, D.L., Lucey, B., Waller, L. a, Childs, J.E. & Real, L. a. (2002). Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl. Acad. Sci. U. S. A.*, 99, 3668–72.

137.

Smith, K.F., Acevedo-Whitehouse, K. & Pedersen, a. B. (2009). The role of infectious diseases in biological conservation. *Anim. Conserv.*, 12, 1–12.

138.

Smith, K.F., Dobson, A.P., McKenzie, F.E., Real, L. a, Smith, D.L. & Wilson, M.L. (2005). Ecological theory to enhance infectious disease control and public health policy. *Front. Ecol. Environ.*, 3, 29–37.

139.

Stallknecht, D.E. (2007). Impediments to wildlife disease surveillance, research, and diagnostics. *Curr. Top. Microbiol. Immunol.*, 315, 445–61.

140.

Suzán, G., Marcé, E., Giermakowski, J.T., Armién, B., Pascale, J., Mills, J., *et al.* (2008). The effect of habitat fragmentation and species diversity loss on hantavirus prevalence in Panama. *Ann. N. Y. Acad. Sci.*, 1149, 80–3.

141.

Swayne, D.E. (2009). *Avian Influenza*. John Wiley & Sons.

142.

Tam, J.S. (2002). Influenza A (H5N1) in Hong Kong: an overview. *Vaccine*, 20 Suppl 2, S77–81.

143.

Taylor, L.H., Latham, S.M. & Woolhouse, M.E. (2001). Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 356, 983–9.

144.

Teunis, P.F.M., van der Heijden, O.G., de Melker, H.E., Schellekens, J.F.P., Versteegh, F.G. a & Kretzschmar, M.E.E. (2002). Kinetics of the IgG antibody response to pertussis toxin after infection with *B. pertussis*. *Epidemiol. Infect.*, 129, 479–89.

145.

Tilman, D. & Kareiva, P. (1997). *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press.

146.

Trenberth, K. (2008). Climate change and extreme weather events. ... *Model. Forum Chging Clim. Dyn. ...*, 2, 90–104.

147.

Tuytens, F., Delahay, R.J., MacDonald, D.W., Cheeseman, C.L., Long, B. & Donnelly, C.A. (2000). Spatial Perturbation caused by a badger (*Meles meles*) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (*Mycobacterium bovis*). *J. Anim. ...*, 69, 815–828.

148.

Tuytens, F.A.M., Macdonald, D.W., Delahay, R., Rogers, L.M., Mallinson, P.J., Donnelly, C.A., *et al.* (1999). Differences in trappability of European badgers *Meles meles* in three populations in England. *J. Appl. Ecol.*, 1051–1062.

149.

Valiakos, G., Papaspyropoulos, K., Giannakopoulos, A., Birtsas, P., Tsiodras, S., Hutchings, M.R., *et al.* (2014). Use of Wild Bird Surveillance, Human Case Data and GIS Spatial Analysis for Predicting Spatial Distributions of West Nile Virus in Greece. *PLoS One*, 9, e96935.

150.

Vallat, B. (2008). *Improving wildlife surveillance for its protection while protecting us from the diseases it transmits: OIE - World Organisation for Animal Health.*

151.

Vicente, J., Delahay, R., Walker, N. & Cheeseman, C.L. (2007). Social organization and movement influence the incidence of bovine tuberculosis in an undisturbed high-density badger *Meles meles* population. *J. Anim. ...*, 76, 348–360.

152.

Walsh, D.P. & Miller, M.W. (2010). A weighted surveillance approach for detecting chronic wasting disease foci. *J. Wildl. Dis.*, 46, 118–35.

153.

Ward, M.P., Laffan, S.W. & Highfield, L.D. (2007). The potential role of wild and feral animals as reservoirs of foot-and-mouth disease. *Prev. Vet. Med.*, 80, 9–23.

154.

WHO. (2004). Nipah virus outbreak(s) in Bangladesh, January–April 2004. *Wkly. Epidemiol. Rec.*, 79, 161–172.

155.

WILDTECH Report Summary [WWW Document]. (2014). . URL [http://cordis.europa.eu/result/rcn/55284\\_en.html](http://cordis.europa.eu/result/rcn/55284_en.html).

156.

Wilkinson, D., Smith, G.C., Delahay, R.J., Rogers, L.M., Cheeseman, C.L. & Clifton-Hadley, R.S. (2000). The effects of bovine tuberculosis (*Mycobacterium bovis*) on mortality in a badger (*Meles meles*) population in England. *J. Zool.*, 250, 389–395.

157.

Wilson, H.B. & Hassell, M.P. (1997). Host – parasitoid spatial models: the interplay of demographic stochasticity and dynamics. *Proc. R. Soc. Lond. B*, 264, 1189–1195.

158.

Witmer, G. (2005). Wildlife population monitoring: some practical considerations. *Wildl. Res.*

159.

Young, J.C., Mills, J.N., Enria, D. a, Dolan, N.E., Khan, a S. & Ksiazek, T.G. (1998). New World hantaviruses. *Br. Med. Bull.*, 54, 659–73.

160.

Zijerveld, L.J.J. (2012). *Integrated modelling and Bayesian inference applied to population and disease dynamics in wildlife: M.Bovis in Badgers in Woodchaster Park.*