The role of population structure and size in determining bat pathogen richness

Tim C. D. Lucas

A thesis submitted in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy of

University College London

2016

Primary supervisor:
Prof. Kate E. Jones

Secondary supervisor:
Dr Hilde M. Wilkinson-Herbots

I, Tim C. D. Lucas, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The huge number of pathogen species strongly affects human health and ecological systems. I examine the role of host population structure and size in maintaining pathogen species richness in an important reservoir host for zoonotic viruses, bats (Order, Chiroptera). Firstly I test whether population structure is associated with high viral richness across wild bat species within a comparative phylogenetic analysis. I find evidence that bat species with more structured populations have more virus species. As this type of study cannot distinguish between specific mechanisms, I then formulate epidemiological models to test whether more structured host populations may allow invading pathogens to avoid competition. These models show that population structure does not affect the rate of pathogen invasion by this mechanism. Rather, in these models only the disease dynamics within the local group matter. As both global host population structure and local group size appear to be important for disease invasion, I use the same modelling framework to compare the importance of host group size and number of groups. I find that host group size has a stronger affect than number of groups. There are very few population size estimates for bats to directly test the importance of host population size on pathogen richness. Therefore I develop a method for estimating bat population sizes from acoustic surveys to assist future research. Overall in this thesis, I show that the structure and size of host bat populations can affect their ability to maintain many pathogen species and I provide a method to measure population sizes of bats. These findings increase our understanding of the ecological process of pathogen community construction and can help optimise host surveillance for zoonotic pathogens.

Acknowledgements

Firstly and most importantly I would like to thank my wife, Katrina, for helping me beyond measure throughout my PhD and my son, Dylan, for making my life tiring and brilliant for the last two years. You are

Secondly, I would like to thank my supervisors Kate and Hilde.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Pathogen richness and the impacts of zoonotic diseases

Over 60% of newly emerged diseases are zoonotic (acquired from animals) with wild animals being the predominant source (Jones et al. 2008, Taylor et al. 2001, Woolhouse & Gowtage-Sequeria 2006). Zoonotic diseases can be extremely virulent with viruses such as Nipah and Ebola having case fatality rates over 50% (Lefebvre et al. 2014, Luby et al. 2009). Furthermore these pathogens can have large economic costs. For example SARS is estimated to have cost $40 billion (Knobler et al. 2004). In particular these impacts can have huge effects on developing economies. In the case of the 2014 Ebola epidemic caused both Liberia and Guinea to experience negative per capita growth rates of -2% (World Bank 2014, World Bank 2015). More generally, death rates per 1,000 people living with AIDS are up to ten times higher in developing countries than in Europe and North America (Granich et al. 2015).

Surveillance of zoonotic diseases is crucial to reducing the health impacts of these diseases. In particular it is important to categorise and describe diseases before they spill over into humans, for example, SARS was not identified until months into the pandemic (Drosten et al. 2003). It is also important to improve our ability to predict when outbreaks will occur. For example, if it is known that there is *i*) a disease prevalence in a given host species than normal, or *ii*) a greater-than-usual abundance of a species that is a known reservoir of a high risk zoonotic disease, or *iii*) increased contacts between humans and a pathogen reservoir, preparations can be made for a potential outbreak in that area.

However, funds for zoonotic disease surveillance are limited and so efforts must be optimised. Knowing which species are likely to have many pathogens allows us to sample and identify potentially zoonotic viruses efficiently. Suggested factors that might control pathogen richness include individual, environmental and population level traits. Individual traits that have been studied include body mass and longevity. Increased body mass is expected to increase pathogen richness as large bodies provide more resource for pathogens to consume and potentially more niches for them to occupy (Arneberg 2002, Bordes et al. 2008, Gómez-Rodríguez et al. 2015, Kamiya et al. 2014, Poulin 1995). Increased longevity is also expected to increase pathogen richness by increasing the number of pathogens a host encounters in its lifetime (Ezenwa et al. 2006, Luis et al. 2013, Nunn et al. 2003). Environmental factors may also play a role. Latitude has been studied

as a proxy for environmental factors (Kamiya et al. 2014, Poulin 2010). It is predicted that warmer climates promote species richness via metabolic mechanisms or by increasing the rate of evolution (Brown et al. 2004, Dunn et al. 2010, Rohde 1992). Furthermore, population level traits that affect the dynamics of disease spread have also been studied. Animal density (Arneberg 2002, Kamiya et al. 2014, Nunn et al. 2003), sociality (Altizer et al. 2003, Bordes et al. 2007, Ezenwa et al. 2006, Vitone et al. 2004) and population structure (Gay et al. 2014, Maganga et al. 2014, Nunes et al. 2006, Turmelle & Olival 2009) have both been predicted to increase pathogen richness by increasing the rate of spread of new pathogens. Finally, species with larger range sizes are expected to have higher pathogen richness as they experience a wider range of environments and have more sympatric host species (Kamiya et al. 2014, Nunn et al. 2003)). These relationships provide a basis for predicting which species will have high pathogen richness and should be prioritised for sampling and surveillance. However, without a better mechanistic understanding of how pathogen richness is created and maintained it is difficult to predict how pathogen richness, and therefore zoonotic disease risk, will respond to global change.

The global richness of pathogens is large but mostly unknown (Poulin 2014). Recent large studies have found tens (Anthony et al. 2013) or even hundreds (Anthony et al. 2015) of virus species in a single host species. This suggests that the global number of mammalian virus species is of the order of hundreds of thousands (Anthony et al. 2013) while only 3,000 virus species, across all host groups, are currently described (King et al. 2011). Recent large databases include nearly 2,000 pathogens from approximately 400 wild animal hosts (Wardeh et al. 2015). Given that there are over 5,000 named mammal species (Wilson & Reeder 2005), the number of undiscovered pathogens is likely to be huge.

Competition between pathogens can occur by different mechanisms: immunological mechanisms such as cross-immunity or shared immune response (Fenton & Perkins 2010) and ecological mechanisms such as removal of susceptible hosts by death (Rohani et al. 2003) or competition for internal host resources (Griffiths et al. 2014). As in ecological systems, competition leads us to the expectation that competitive exclusion occurs (Ackleh & Allen 2003, Ackleh & Salceanu 2014, Bremermann & Thieme 1989, Martcheva & Li 2013, Turner & Garnett 2002). Therefore, the large number of coexisting parasite species needs an explanation.

1.2 Influence of population structure and size on pathogen richness

The roles of population size and density in disease dynamics are well established (Anderson & May 1979, Heesterbeek 2002, Lloyd-Smith et al. 2005, May & Anderson 1979). Broadly, larger populations can maintain diseases more easily by having a larger pool of susceptible individuals (individuals without acquired immunity) and having a greater number of new susceptible individuals enter the population by birth or immigration (Anderson & May 1979, May & Anderson 1979). High density populations are expected to have a greater number of contacts between individuals and so promote disease spread. However, there is much discussion on if and when the number of contacts might scale independently of density (McCallum et al. 2001)).

There is a large literature on the role of population structure on disease dynamics (see review by Pastor-Satorras et al. (2015)) driven by applications to health applications as well as computer viruses (Pastor-Satorras & Vespignani 2001) and the social spread of information (Goffman & Newill 1964). In particular, work has concentrated on how structure affects the basic reproduction number, R_0 (Barthélemy et al. 2010, Colizza & Vespignani 2007, May & Lloyd 2001, Pastor-Satorras & Vespignani 2001, Wu et al. 2013). This value combines relevant parameters to yield a threshold above which a disease is expected to infect a significant proportion of the population (Anderson & May 1979, May & Anderson 1979). Below the threshold, only small outbreaks that quickly die out are expected.

However, the majority of theoretical work considers single pathogens with models examining whether a pathogen can spread and persist in a population, ignoring all other pathogens. A number of studies have considered the case where two pathogens spread concurrently and examine which pathogen infects more individuals. These studies have found that increased population structure reduces dominance of the more competitive strain (Poletto et al. 2013, Poletto et al. 2015, van de Bovenkamp et al. 2014). However, this again reveals little about how pathogen communities form and what factors control total pathogen richness. Far fewer papers explicitly study the longer term competition between two or more pathogens. Those that do commonly find that competitive exclusion is likely (Ackleh & Allen 2003, Ackleh & Salceanu 2014, Bremermann & Thieme 1989, Castillo-Chavez et al. 1995, Martcheva & Li 2013, Turner & Garnett 2002).

Mechanisms that have been shown to allow pathogen coexistence include superinfection (Li et al. 2010, May & Nowak 1994), density-dependent deaths (Ackleh & Allen 2003, Kirupaharan & Allen 2004) and differing transmission routes (Allen et al. 2003).

The specific role of density on the ability of pathogens to coexist has not been theoretically studied though it is commonly found to promote pathogen richness in comparative empirical studies (Arneberg 2002, Kamiya et al. 2014, Nunn et al. 2003). The few papers that have directly studied how coexistence of pathogens responds to population structure have found that population structure can allow pathogens to coexist even though competitive exclusion would occur in a fully mixed population (Allen et al. 2004, Nunes et al. 2006, Qiu et al. 2013). Furthermore, genetic diversity has been shown to be maximised at intermediate levels of population structure (Campos & Gordo 2006). The roles of population structure and social group size have been examined in comparative studies (Altizer et al. 2003, Bordes et al. 2007, Ezenwa et al. 2006, Gay et al. 2014, Maganga et al. 2014, Rifkin et al. 2012, Turmelle & Olival 2009, Vitone et al. 2004). There is much disagreement between these studies; population structure being shown to promote (Maganga et al. 2014, Turmelle & Olival 2009) and inhibit pathogen richness (Gay et al. 2014) and similarly, group size has been shown to promote (Bordes et al. 2007, Rifkin et al. 2012) and inhibit (Ezenwa et al. 2006) pathogen richness. While increased group size should generally decrease population structure, the literature is rarely clear on the relationships between these variables.

1.3 Bats as reservoirs of zoonotic diseases

In recent decades bats have been implicated in a number of high profile zoonotic outbreaks such as Nipah (Field et al. 2001, Halpin et al. 2011), Ebola (Leroy et al. 2005), SARS (Li et al. 2005) and Hendra (Field et al. 2001). These outbreaks have lead to much research on whether bats are a particular source of zoonotic disease (Luis et al. 2013, Olival et al. 2015, Wang et al. 2011) and examinations of factors, such as flight, social living and longevity, that might predispose them to being reservoirs of zoonotic viruses (Calisher et al. 2006, Dobson 2005, Kuzmin et al. 2011, O'Shea et al. 2014, Racey 2015). Given that bats are the second largest order of mammals (Wilson & Reeder 2005), we may expect them to be the source of many viruses simply through weight of numbers (Luis et al. 2013). The broad conclusions are that while bats do host more zoonotic viruses than other groups

(Luis et al. 2013) they do not host more virus species per host species (Olival et al. 2015).

Many factors of bat populations make them epidemiologically interesting. They have highly varied and sometimes complex social structures (Kerth 2008). While some species are largely solitary or live in very small groups (e.g., *Lasiurus borealis* (Shump & Shump 1982)) some species live in colonies of millions of individuals (e.g., *Pteropus scapulatus* (Birt et al. 2008)). These groups can be very stable (Kerth et al. 2011, McCracken & Bradbury 1981). Further complexity arises due to their propensity for seasonal migration (Cryan et al. 2014, Fleming et al. 2003, Richter & Cumming 2008) and seasonally changing social organisation such as maternity roosts, hibernation roosts and swarming sites (Kerth 2008). Finally, their ability to fly means that populations can be well mixed across large distances (Peel et al. 2013, Petit & Mayer 1999), though this is highly variable with some species having limited dispersal (Wilmer et al. 1994).

However, the population density of many bat species, particularly tree roosting species, is unknown (Clement & Castleberry 2013). As they are small, nocturnal and difficult to identify in flight, estimating their density is incredibly difficult without disruptive and time-consuming roost surveys (Humphrey 1971, Kloepper et al. 2016, Sabol & Hudson 1995). Furthermore, bat densities are generally estimated by counting bats in roosts and dividing this number by area which assumes all roosts have been surveyed (Moreno-Valdez et al. 2004, Speakman et al. 1991, Zahn et al. 2006). As density is tightly linked to pathogen richness (Kamiya et al. 2014) and central to epidemiological models (Anderson & May 1979, May & Anderson 1979) this leaves large gaps in our understanding of disease processes in this group.

1.4 Thesis overview

In this thesis I examined the role of population structure and density on pathogen richness. I used bats as a case study throughout due to their interesting social structure and importance as zoonotic reservoirs. I combined empirical, comparative studies with simulation models. This allowed me to study specific mechanisms while linking my theoretical insights to real-world, empirical tests of hypotheses.

First, in Chapter 2, I empirically tested the hypothesis that population structure is associated with pathogen richness in wild bat populations. I used two

measures of population structure — the number of subspecies and gene flow — and a larger data set than previous studies to ensure robust results and used viral richness as a proxy for overall pathogen richness. For both measures I found that bat species with more structured populations have more known viruses. This relationship is still present after controlling for study bias and phylogenetic nonindependence. I also tested for relationships between body mass and pathogen richness, and range size and pathogen richness, and found strong support for larger bodied bats carrying more viruses and mixed support for range size promoting pathogen richness.

In Chapter 3, I examined one specific mechanism by which population structure may promote increased pathogen richness. I tested whether increased population structure can allow newly evolved pathogen strains to invade and persist more easily. I modelled bat populations as individual-based, stochastic metapopulations and examined the competition dynamics of two identical pathogen strains. I tested two factors related to population structure: dispersal rate and the number of links between subpopulations. I found that increased dispersal rate significantly increased the probability of a newly evolved pathogen invading and persisting in the population. However, this was only the case at intermediate transmission rates. I did not find a significant difference in invasion probability due to the number of linked between subpopulations.

Next, I examined the relationships between a number of elements of population structure (Chapter 4). I clarified the interdependence between range size, population size and density. I also noted that population size can be decomposed into colony size and the number of colonies. Using the same model as in Chapter 3, I then tested which of these factors are most important in promoting pathogen richness. Specifically I tested which factor most strongly promotes the invasion and establishment of newly evolved pathogens. I found that population size is more important than population density and that colony size is the important component of population size.

Given the importance of population size on pathogen richness it is important to have good population estimates for wild bat populations. However, there are currently very few measurements of bat population size due to their small size, nocturnal habit and difficulties in identification. Therefore I aimed to develop a method for estimating bat population size from acoustic data, specifically data collected by the iBats project (Jones et al. 2011). In Chapter 5 I developed a generally applicable method — based on random encounter models (Rowcliffe et al.

2008, Yapp 1956) — for estimating population sizes of animal populations using camera traps or acoustic detectors. I used spatial simulations to test the method for biases and to assess its precision. I found that the method is unbiased and precise as long as a reasonable amount of data is collected.

Finally, in Chapter 6, I discuss broader conclusions, applications and implications of my results. I also discuss potential future directions for research.

2 A comparative test of the role of population structure in determining pathogen richness

This work was conducted in collaboration with Kate Jones and Hilde Wilkinson-Herbots.

2.1 Abstract

Zoonotic diseases make up the majority of human infectious diseases and are a major drain on healthcare resources and economies. Species that host many pathogen species are more likely to be the source of a novel zoonotic disease than species with few pathogens. However, the factors that influence pathogen richness in animal species are poorly understood. The pattern of contacts between individuals (i.e. population structure) can be influenced by habitat fragmentation, sociality and dispersal behaviour. Epidemiological theory suggests that increased population structure can promote pathogen richness by reducing competition between pathogen species. Conversely, it is often assumed that as greater population structure slows the spread of a new pathogen, less structured populations should have greater pathogen richness. Previous studies have had contradictory results and different measures of population structure have been used, complicating the interpretation. Here I used comparative data across 203 bat species, controlling for body mass, range size, study effort and phylogeny, to test whether increased population structure correlates with viral richness. Bats, as a group, make a useful case study because they have been associated with a number of important, recent zoonotic outbreaks. Unlike previous studies, I used two measures of population structure: the number of subspecies and effective levels of gene flow. I find that both measures are positively associated with pathogen richness. My results add more robust support to the hypothesis that increased population structure promotes viral richness in bats. The results support the prediction that increased population structure allows greater pathogen richness by reducing competition between pathogens The prediction that factors that increase R_0 should increase pathogen richness is not supported. Although my analysis implies that increased population structure does promote pathogen richness in bats, the weakness of the relationship and the difficulty in obtaining some measurements means that this is probably not a useful, predictive factor on its own for optimising zoonotic surveillance.

2.2 Introduction

Zoonotic pathogens make up the majority of newly emerging diseases and have profound consequences for public health, economics and international development (Jones et al. 2008, Smith et al. 2014, World Bank 2014). Better models for

predicting which wild host species are potential reservoirs of zoonotic diseases would allow us to optimise zoonotic disease surveillance and anticipate how the risks of disease spillover might change with global change. The chance that a host species will be the source of an outbreak depends on a number of factors, such as its proximity and interactions with humans, and the prevalence and the number of pathogen species it carries (Wolfe et al. 2000). However, the factors that control the number of pathogen species in a host species remain poorly understood.

A number of species traits that might control pathogen richness have been studied. These traits can be at the level of the individual (e.g., body mass and longevity) or the level of the population (e.g., population density, sociality and species range size). Large bodied animals have been shown to have high pathogen richness with large bodies providing more resources for pathogens (Arneberg 2002, Bordes et al. 2008, Kamiya et al. 2014, Luis et al. 2013, Poulin 1995). Long lived species are expected to have high pathogen richness, because the number of pathogens a host encounters in its lifetime will be higher (Ezenwa et al. 2006, Luis et al. 2013, Nunn et al. 2003). Animal density (Arneberg 2002, Kamiya et al. 2014, Nunn et al. 2003) and sociality (Altizer et al. 2003, Bordes et al. 2007, Ezenwa et al. 2006, Vitone et al. 2004) are both predicted to increase pathogen richness by increasing the rate of spread, R_0, of a new pathogen. Finally, widely distributed species have high pathogen richness, potentially because they experience a wider range of environments or because they are sympatric with more species (Kamiya et al. 2014, Luis et al. 2013, Nunn et al. 2003).

A further population level factor that may affect pathogen richness is population structure. Population structure can be defined as the extent to which interactions between individuals in a population are non-random. The role of population structure on human epidemics has been studied in depth and it has been shown that decreased population structure increases the speed of pathogen spread and makes establishment of a new pathogen more likely (Colizza & Vespignani 2007, Vespignani 2008). In comparative studies of pathogen richness in wild animals, this relationship with R_0 is often taken as a prediction that decreased population structure will increase pathogen richness relative to other host species (Altizer et al. 2003, Morand 2000, Nunn et al. 2003, Poulin 2014, Poulin & Morand 2000). However, epidemiological models of highly virulent pathogens have shown that increased population structure can allow persistence of a pathogen where a well-mixed population would experience a single, large epidemic followed by pathogen extinction (Blackwood et al. 2013, Plowright et al.

2011). Furthermore, the assumption that high R_0 leads to high pathogen richness ignores inter-pathogen competition. Simple epidemiological models of competition between multiple pathogens show that, in completely unstructured populations, a competitive exclusion process occurs but that adding population structure makes coexistence possible (Allen et al. 2004, Nunes et al. 2006, Qiu et al. 2013).

There is a lack of large, comparative studies of the role of population structure on pathogen richness. Sociality, which is one constituent part of population structure, has been well studied. However, in primates only a weak positive association between sociality and pathogen richness was found (Vitone et al. 2004). Furthermore, a negative association was found in rodents (Bordes et al. 2007) and in even and odd-toed hoofed mammals (Ezenwa et al. 2006). Finally, two studies tested for an association between group size and parasite richness in bats (Bordes et al. 2008, Gay et al. 2014). Amongst 138 bat species, Bordes et al. (2008) found no relationship between group size (coded into four classes) and bat fly species richness. Gay et al. (2014) a negative relationship between colony size and viral richness but a positive relationship between colony size and ectoparasite richness. While sociality is an important component of population structure it does not capture fully how connected the population is globally.

Three studies have used comparative data to test for an association between global population structure and viral richness in bats. A study on 15 African bat species found a positive relationship between the extent of distribution fragmentation and viral richness (Maganga et al. 2014). Conversely, a study on 20 South-East Asian bat species found the opposite relationship (Gay et al. 2014). These studies used the ratio between the perimeter and area of the species' geographic range as their measure of population structure. However, range maps are very coarse for many species. Furthermore, range maps are likely to be more detailed (and therefore have a greater perimeter) in well studied species.

A global study on 33 bat species found a positive relationship between F_{ST} — a measure of genetic structure — and viral richness (Turmelle & Olival 2009). However, this study included measures using mtDNA which only measures female dispersal which may have biased the results as many bat species show female philopatry (Hulva et al. 2010, Kerth et al. 2002). Furthermore, this study used measures of F_{ST} irrespective of the spatial scale of the study including studies covering from tens (McCracken & Bradbury 1981) to thousands (Petit & Mayer 1999) of kilometres. As isolation by distance has been shown in a number of bat

species (Burland et al. 1999, Hulva et al. 2010, O'Donnell et al. 2015, Vonhof et al. 2015), this could bias results further. Finally, when a global F_{ST} value is not given, Turmelle & Olival (2009) used the mean of all pairwise F_{ST} values between sites. This is not correct as pairwise and global F_{ST} values have different relationships with effective migration rates.

Here I used a phylogenetic comparative approach to test for a relationship between increased population structure and pathogen richness in the largest study of bats to date. I used phylogenetic linear models, controlling for the other life history characteristics known to impact pathogen richness, to quantify the relationship between viral richness (as a proxy for pathogen richness) and two measures of population structure: the number of subspecies and effective gene flow. I used two measures of population structure to increase the robustness of the analysis; this is particularly important as previous studies have had contradictory results (Gay et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009).

I found that increases in both measures of population structure are positively associated with viral richness and are included as explanatory variables in the best models for describing viral richness. Furthermore, I found that the role of phylogeny is very weak both in the models and in the distribution of viral richness amongst taxa.

2.3 Methods

2.3.1 Data Collection

2.3.1.1 Pathogen richness

To measure pathogen richness I used data from Luis et al. (2013). This data simply includes known infections of a bat species with a virus species. I have used viral richness as a proxy for pathogen richness more generally, but the analysis could also be considered as representative of viral richness only. Rows with host species that were not identified to species level according to Wilson & Reeder (2005) were removed. Many viruses were not identified to species level or their specified species names were not in the ICTV virus taxonomy (King et al. 2011). Therefore, I counted a virus if it was the only virus, for that host species, in the lowest taxonomic level identified (present in the ICTV taxonomy). For example, if a host is recorded as harbouring an unknown Paramyxoviridae virus, then it is logical to assume that the host carries at least one Paramyxoviridae virus. If a host carries

an unknown Paramyxoviridae virus and a known Paramyxoviridae virus, it is hard to confirm that the unknown virus is not another record of the known virus. In this case, the host would be counted as having one virus species.

2.3.1.2 Population structure data

I used two measures of population structure: the number of subspecies and the effective level of gene flow. The number of subspecies was counted using the taxonomy from Wilson & Reeder (2005). The effective level of gene flow was calculated from estimates of F_{ST} collated from the literature. The studies were from a wide range of spatial scales, from local ($\sim$ 10 km) to continental. As F_{ST} often increases with spatial scale (Burland et al. 1999, Hulva et al. 2010, O'Donnell et al. 2015, Vonhof et al. 2015) I controlled for this by only using data from studies where a large proportion of the species range was studied. I used the ratio of the furthest distance between F_{ST} samples (taken from the paper or measured with http://www.distancefromto.net/ if not stated) to the length of the IUCN species range (IUCN 2010) and only used studies if this ratio was greater than 0.2. This is an arbitrary value that was a compromise between retaining a reasonable number of data points and controlling for the bias in spatial scale. I only used global F_{ST} estimates as the mean of pairwise F_{ST} values is not necessarily equal to the global F_{ST} value. I converted all F_{ST} values to effective migration rates using $M = (1-F_{ST})/4F_{ST}$. This transforms the data from being bound by $(0, 1)$ to being in the range $[0, \infty)$ and is easier to interpret.

The two measures of population structure were analysed separately because the number of subspecies data set had 196 data points but there was only F_{ST} data for 24 bat species. For the subspecies analysis, all bat species in Luis et al. (2013) were used (i.e. all species with at least one known virus species). This was to avoid using the very large number of bat species that have simply never been sampled for viruses. However, for the gene flow analysis, all bat species with suitable F_{ST} estimates were used. As some bat species had suitable F_{ST} estimates but were not present in Luis et al. (2013), some bat species with zero known virus species were included. These bat species with no known viruses were included to make the greatest use of the F_{ST} data available and because the number of species with no known virus species was not unduly large (7 species).

After data cleaning there was data for 196 bat species in 11 families for the subspecies analysis. Due to the limited number of studies and the restrictive requirements imposed on study design, there was only data for 24 bat species in

7 families for the effective gene flow analysis. The raw data are included in Table A.1.

2.3.1.3 Other explanatory variables

To control for study bias I collected the number of PubMed and Google Scholar citations for each bat species name including synonyms from ITIS (ITIS 2015) via the *taxize* package (Chamberlain & Szöcs 2013). The counts were scraped using the *rvest* package (Wickham 2015). I log transformed these variables as they were strongly right skewed. I tested for correlation between these two proxies for study effort using phylogenetic least squares regression (pgls), using the best-supported phylogeny from Fritz et al. (2009), and likelihood ratio tests using the *caper* package (Orme et al. 2012) (Figures 2.1 and A.1). The log number of citations on PubMed and Google scholar were highly correlated (pgls: $t = 19.32$, df $= 194$, $p < 10^{-5}$). As the correlation between citation counts is strong, I only used Google Scholar reference counts in subsequent analyses.

A number of other factors that have previously been found to be important were included as additional explanatory variables: body mass (Bordes et al. 2008, Gay et al. 2014, Han et al. 2015, Kamiya et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009), range size (Kamiya et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009). These other factors were included to avoid spurious positive results occurring simply due to correlations between pathogen richness and a different, causal factor. Despite commonly being associated with pathogen richness (Arneberg 2002, Kamiya et al. 2014, Nunn et al. 2003), population density is not included in the analysis as there is very little data for bat densities. Measures of body mass were taken from Pantheria (Jones et al. 2009) and primary literature (Aldridge 1987, Arita 1993, Canals et al. 2005, Heaney et al. 2012, Henderson & Broders 2008, Lim & Engstrom 2001, López-Baucells et al. 2014, Ma et al. 2003, Oleksy et al. 2015, Orr & Zuk 2013, Owen et al. 2003, Zhang et al. 2009). *Pipistrellus pygmaeus* was assigned the same mass as *P. pipistrellus* as they are indistinguishable by mass. Body mass measurements were log transformed as they were strongly right skewed. Distribution size was estimated by downloading range maps for all species from IUCN (IUCN 2010) and were also log transformed due to right skew.

2.3.2 Statistical analysis

Statistical analysis for both response variables — number of subspecies and effective level of gene flow — was conducted using an information theoretical approach (Burnham & Anderson 2002), specifically following Whittingham et al. (2006) and Whittingham et al. (2005). All analysis were performed in R (R Development Core Team 2010) and all code is available on GitHub. I chose a credible set of models including all combinations of explanatory variables and a model with just an intercept. In the analysis using the number of subspecies response variable I also modelled the interaction study effort and number of subspecies by including their product. This interaction was included as I believed *a priori* that this interaction may be present as subspecies in well studied species are more likely to be identified. The interaction was only included in models with both study effort and number of subspecies as individual terms. Following Whittingham et al. (2005) I included a uniformly distributed random variable. This variable can be used to benchmark how important other explanatory variables are. The whole analysis was run 50 times, resampling the random variable each time.

To control for phylogenetic non-independence of datapoints I used the best-supported phylogeny from Fritz et al. (2009) (Figure 2.1) which is the supertree from Bininda-Emonds et al. (2007) with names updated to match the taxonomy by Wilson & Reeder (2005). Phylogenetic manipulation was performed using the *ape* package (Paradis et al. 2004). I also performed the analysis using the phylogeny from Jones et al. (2005) as this has some broad topological differences including the Rhinolophoidea being sister to the Pteropodidae rather than being related to the other insectivorous bats (Figure A.2).

The importance of the phylogeny on each variable separately was examined by estimating the λ parameter when regressing the variable against an intercept using the *pgls* function in *caper* (Orme et al. 2012). The parameter λ usually takes values between zero and one and *pgls* constrains λ within these bounds. $\lambda = 0$ implies no autocorrelation while a trait evolving by Brownian motion along the tree would have $\lambda = 1$. I tested fitted λ values against the null hypothesis of $\lambda = 0$ (no correlation between species) with log-likelihood ratio tests using *caper* (Orme et al. 2012).

I fitted phylogenetic regressions for all models in the credible set using the function *gls* in the package *nlme* (Pinheiro et al. 2015). The explanatory variables were centred and scaled to allow direct comparison of the coefficients (Schielzeth 2010). For each regression model I simultaneously fitted the λ parameter as this

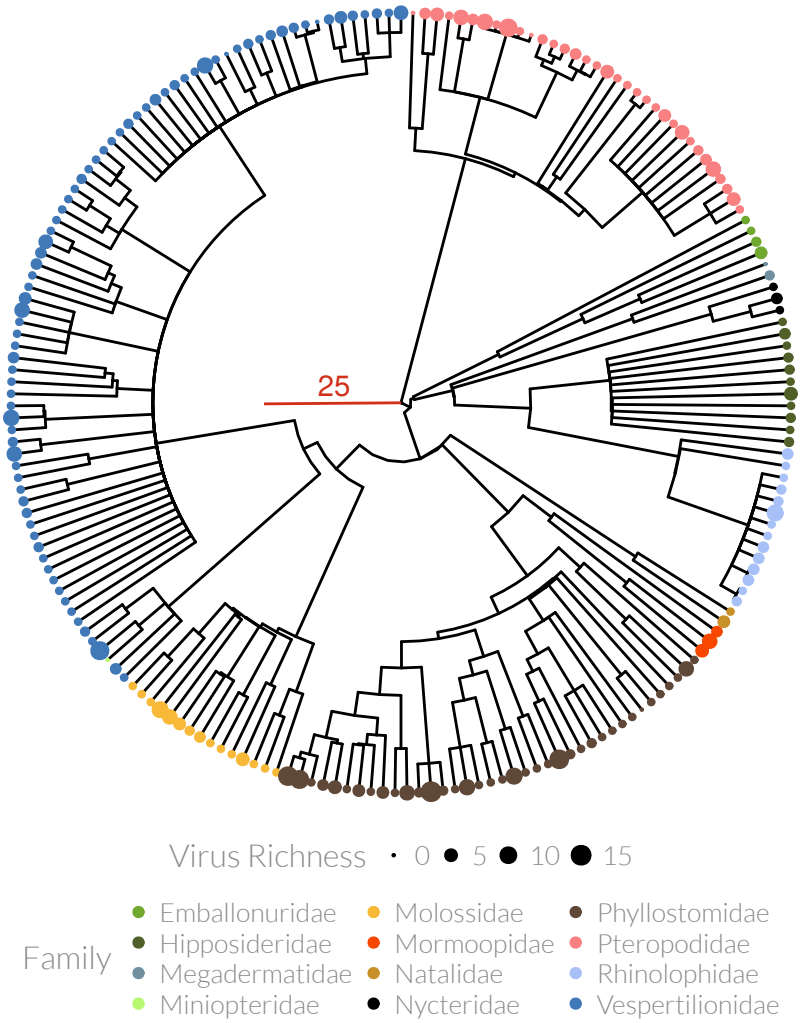


Figure 2.1 The phylogenetic distribution of viral richness. There is no clear association between phylogeny and virus richness (pgls: $\lambda = 0.04$, $p = 0.21$). The phylogeny is from (Bininda-Emonds et al. 2007) pruned to include all species used in either the number of subspecies or gene flow analysis. Dot size shows the number of known viruses for that species and colour shows family. The red scale bar shows 25 million years.

avoids misspecifying the model (Revell 2010). Unlike the *pgls* function, *gls* does not constrain λ to be in the range $[0,1]$. $\lambda < 0$ indicates that residuals from the fitted model are distributed on the phylogeny more uniformly than expected by chance. κ and δ parameters were constrained to one as they are more concerned with when along a branch evolution occurs than the importance of the phylogeny. Further, fitting multiple parameters makes interpretation difficult.

To establish the importance of variables I calculated the probability, *Pr*, that each variable would be in the best model amongst those examined (under the assumption that all models are *a priori* equally likely). This value can more generally, and with fewer assumptions, be considered as simply the relative weight of evidence for each variable being in the best model amongst those examined. I calculated AICc for each model. I calculated the average AICc, $\bar{\text{AICc}}$, by averaging AICc scores within models. ΔAICc was calculated as $\min(\bar{\text{AICc}}) - \bar{\text{AICc}}$, not the mean of the individual ΔAICc scores, to guarantee that the best model has $\Delta\text{AICc} = 0$. From these ΔAICc values I calculated Akaike weights, w. This value can be interpreted as the probability that a model is the best model, given the data, amongst those examined. For each variable, the sum of the Akaike weights of models containing that variable are summed to give *Pr*. This value can be interpreted as the probability that the given variable is in the best model.

To determine the direction and strength of the effect of each variable the mean of its regression coefficient, b, in all models that contained that variable, weighted by the model's Akaike weight, was also calculated. In the subspecies analysis the inclusion of an interaction term between number of subspecies and study effort makes interpretation of this mean coefficient more difficult, particularly because the interaction term greatly affects the estimated value of b. To aid interpretation, the mean coefficient for the number of subspecies was calculated for: *i*) all models containing the number of species, *ii*) only models with the interaction term and *iii*) only models with the number of subspecies but not the interaction term.

2.4 Results

2.4.1 Number of Subspecies

The number of described virus species for a bat host ranged up to 15 viruses in *Carollia perspicillata*. There appears to be a positive relationship between the number of subspecies and viral richness (Figure 2.2) though few species have more than five subspecies. Out of 39 fitted models, the top seven models all had $\Delta\text{AICc} < 4$ meaning there was no clear best model (Table 2.1 and Table A.2). However these top seven models all contained study effort, number of subspecies and the interaction between these two variables. The explanatory variables log(Mass), log(Range Size) and the uniformly random variable are each in three

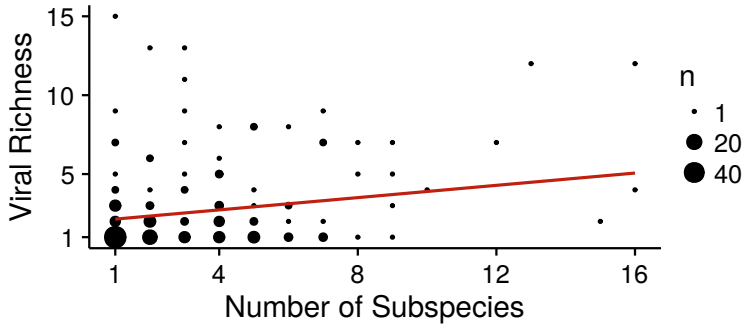


Figure 2.2 The relationship between number of subspecies and viral richness for 196 bat species. The area of the circle shows the number of bat species at each discrete value. 48 bat species have one subspecies and one known virus species. The red line represents a phylogenetic multiple regression including all the explanatory variables but no interaction term. The line shows the slope from the multiple regression with the intercept being calculated by setting other explanatory variables to their median values.

of the top seven models. These top seven models had a combined weight of 0.96 meaning that there is a 96% chance that one of these models is the best model amongst those examined.

Summing the Akaike weights of all models that contain a given variable gives a probability, Pr, that the variable would be in the best model amongst those in the plausible set (Whittingham et al. 2006). The number of subspecies is very likely in the best model ($Pr > 0.99$) as is the interaction term between the number of subspecies and study effort ($Pr = 0.96$) compared to the benchmark random variable which has $Pr = 0.25$ (Figure 2.4A and Table 2.2). When models with the interaction term are removed there is, on average (mean weighted by Akaike weights), a positive relationship between the number of subspecies and viral richness ($b = 0.63$, variance = 0.02). Models with an interaction term between the number of subspecies and study effort have a positive interaction term ($b = 0.5$, variance = 5.11×10^{-5}) and linear term ($b = 0.31$, variance = 2.13×10^{-4}).

Study effort is very likely in the best model ($b = 0.99$, $Pr > 0.99$). Body mass and range size are also probably in the best model ($b = 0.48$, $Pr = 0.73$ and $b = 0.35$, $Pr = 0.54$ respectively) with positive relationships of slightly lower strength than the number of subspecies in models without an interaction term ($b = 0.63$, variance = 0.02).

When using the phylogeny from Jones et al. (2005) the results are broadly similar (Figure A.3 and Tables A.4 and A.6). Study effort, the number of subspe-

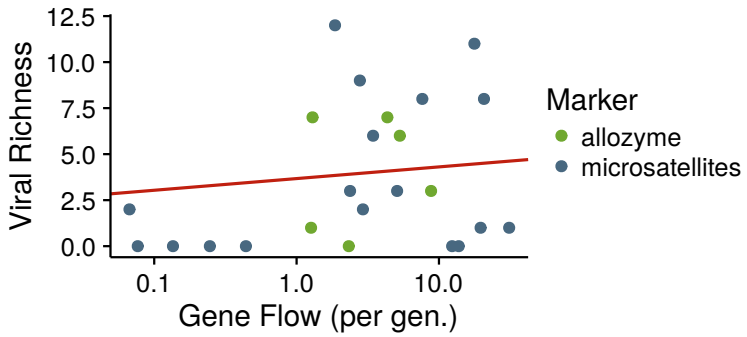


Figure 2.3 Relationship between viral richness and log effective gene flow per generation for 24 bat species. Green points are studies that estimated effective gene flow using allozymes and blue points are studies using microstatellites.

cies and the interaction between the number of subspecies and study effort have strong support while range size and mass have intermediate support. However, mass, range size and the interaction between number of subspecies and study effort have slightly weaker support than in the analysis using the phylogeny from Bininda-Emonds et al. (2007).

2.4.2 Gene Flow

The number of described virus species for a bat host ranged up to 12 viruses in *Miniopterus schreibersii* (Figure 2.3). Only the model with study effort, gene flow and body mass was well supported with the second model having an ΔAICc of 34 (Table 2.1 and Table A.2). The effective level of gene flow was likely in the best model ($Pr > 0.999$, see Figure 2.4B and Table 2.2). On average (mean weighted by Akaike weights) there was a negative relationship between gene flow and viral richness ($b = -0.67$, variance = 5.48×10^{-3}) despite the insignificant positive relationship (Figure 2.3) estimated by the single-predictor model (pgls: $b = 0.63$, $t = 1.16$, df = 13, $p = 0.27$). Possibly due to the smaller sample size, or a weaker relationship, this coefficient was much more varied than the number of subspecies coefficient with 22% of multiple-regression models estimating a positive relationship.

Study effort was very likely in the best model ($Pr > 0.999$) as was body mass ($Pr > 0.999$). However, body mass had a negative average coefficient ($b = -0.35$, variance = 0.04). In contrast to the number of subspecies analysis, range size was almost certainly not in the best model with $Pr = 3.96 \times 10^{-8}$. Of the three

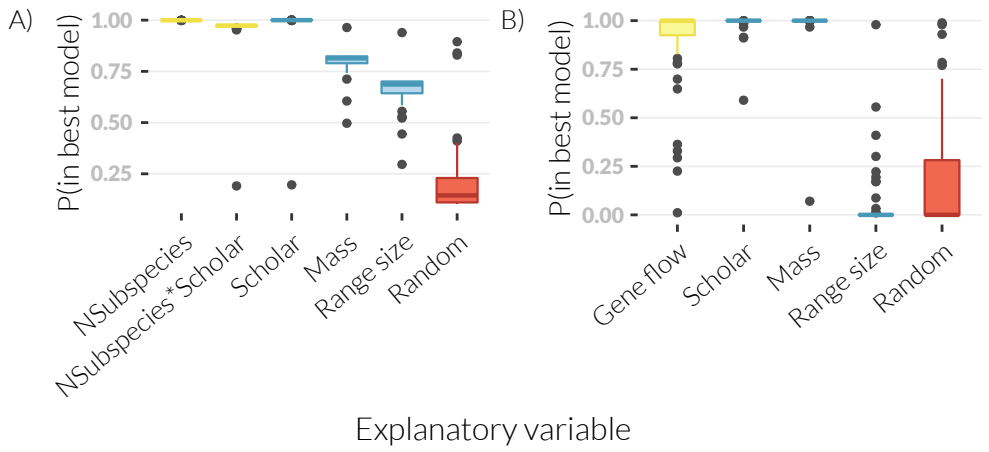


Figure 2.4 The relative weight of evidence that each explanatory variable is in the best model for explaining viral richness. The probability that each variable is in the best model (amongst the models tested) is shown for A) the number of subspecies analysis and B) the effective gene flow analysis. The boxplots show the variation of the results over 50 resamplings of the uniformly random "null" variable. The thick bar of the boxplot shows the median value, the interquartile range is represented by a box, vertical lines represent range, and outliers are shown as filled circles. The red "Random" box is the uniformly random variable. Population structure (number of subspecies and effective gene flow), shown in yellow, is likely to be in the best model in both analyses.

explanatory variables in the best model, study effort had the largest effect ($b = 2.49$, variance $= 0.08$). The effect size of gene flow ($b = -0.67$, variance $= 5.48\times10^{-3}$) was approximately twice the size of that of body mass ($b = -0.35$, variance $= 0.04$)

When using the phylogeny from Jones et al. (2005) the analysis became very unstable (Figure A.3). The support for each variable changed dramatically with each resampling of the random variable. On average however, only the model containing mass and range size is supported (Tables A.3 and A.6).

2.4.3 Phylogenetic Analysis

2.4.3.1 Number of subspecies

Figure 2.1 shows the phylogeny used and the number of viruses for each species. The mean number of viruses across families is fairly constant with Nycteridae having the smallest mean, (1.67). The highest mean is Mormoopidae with 5 virus

Table 2.1 Model selection results for number of subspecies and effective level of gene flow analysis. Models are ranked according to $\bar{AIC}c$ and only the best nine and three models are shown respectively. Models were fitted to all combinations of variables (in total 39 number of subspecies models and 32 effective gene flow models). $\bar{AIC}c$ is the mean AICc score across 50 resamplings of the null random variable. ΔAICc is the model's $\bar{AIC}c$ score minus $\min(\bar{AIC}c)$. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights. log(Scholar)*NSubspecies indicates the interaction term between study effort and number of subspecies.

Model	$\bar{AIC}c$	ΔAICc	w	$\sum w$
Number of Subspecies				
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + log(Mass) + log(RangeSize)	882	0.00	0.38	0.38
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + log(Mass)	884	1.39	0.19	0.57
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + Random + log(Mass)	885	2.24	0.12	0.70
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies	885	3.14	0.08	0.78
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + log(RangeSize)	886	3.18	0.08	0.86
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + Random + log(RangeSize)	886	3.94	0.05	0.91
log(Scholar) + NSubspecies + log(Scholar)*NSubspecies + Random	886	3.95	0.05	0.96
log(Scholar) + NSubspecies + log(Mass) + Random	889	6.93	0.01	0.97
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize) + rand	890	7.80	0.01	0.98
Gene flow				
log(Scholar) + log(Gene flow) + log(Mass)	71	0.00	1.00	1.00
log(Range size)	105	34.09	0.00	1.00
log(Mass)	106	35.06	0.00	1.00

Table 2.2 Estimated variable weights (probability that a variable is in the best model) and their estimated coefficients for both number of subspecies and gene flow analyses. The coefficients for the number of subspecies variable are given for models with and without the interaction term because this term strongly changes the coefficient and because the coefficient can only be usefully interpreted when estimated without the interaction. However, there are no weights for these separated terms as they are not directly compared in the model selection framework.

	Number of Subspecies		*Gene flow*	
Variable	*Pr*	Coefficient	*Pr*	Coefficient
Number of subspecies				
Total	1.00	0.32		
Models without interaction term		0.63		
Models with interaction term		0.31		
Number of subspecies*log(Scholar)	0.96	0.50		
Gene flow			1	−0.67
log(Scholar)	1.00	0.99	1	2.49
log(Mass)	0.73	0.48	1	−0.35
log(Range size)	0.54	0.35	3.96×10^{-8}	1.57
Random	0.25	0.05	2.21×10^{-9}	0.23

species per bat species, but this is based on only 3 species. The Phyllostomidae have the second highest mean of 3.49 ($n = 37$).

The small change in mean pathogen richness across families and the lack of clear pattern in Figure 2.1 implies that viral richness is not strongly phylogenetic. This is corroborated by the small estimated size of λ ($\lambda = 0.04$, $p = 0.21$).

Of the explanatory variables, the number of subspecies had no phylogenetic autocorrelation ($\lambda = 10^{-6}$, $p > 0.999$), study effort and distribution size had weak but significant autocorrelation (Study Effort: $\lambda = 0.1$, $p = 9.12 \times 10^{-3}$, Distribution size: $\lambda = 0.46$, $p < 10^{-5}$) and body mass was strongly phylogenetic ($\lambda = 0.93$, $p < 10^{-5}$). Across all multiple regression models the mean value of λ was 0.08 which implied that the residuals from the models were very weakly phylogenetic. A small number of models (0.4%) had negatively phylogenetically distributed residuals.

2.4.3.2 Effective gene flow

There was no phylogenetic signal in the number of virus species ($\lambda = 10^{-6}$, $p > 0.999$). Gene flow also had no phylogenetic autocorrelation ($\lambda = 10^{-6}$, $p > 0.999$). Due to the limited sample size, significance tests are unlikely to have much power. There is little evidence of phylogenetic autocorrelation in study effort ($\lambda = 0.15$, $p = 0.56$). However, there is some weak evidence of phylogenetic signal in range size as the estimated size of λ is large while p is also large, potentially due to a lack of statistical power ($\lambda = 0.67$, $p = 0.53$). Body mass showed significant phylogenetic autocorrelation ($\lambda = 0.79$, $p = 2.69 \times 10^{-3}$).

Across all multiple regression models the mean value of λ is -1.64 and a large number of individual models (58%) had negatively phylogenetically distributed residuals implying the residuals from the model are spread more uniformly on the phylogeny than expected by chance. Due to the small sample size this was probably due to a small number of data points with large residuals being distant on the tree.

2.5 Discussion

In this study I have used known viral richness in bats as a case study for the more general hypothesis that increased population structure promotes pathogen richness. In both analyses I found that a positive effect of increasing population structure (a positive effect of the number of subspecies and a negative effect of gene flow) is likely to be in the best model for explaining viral richness. Only the effective gene flow analysis, when performed using the phylogeny from Jones et al. (2005), does not support this hypothesis. Therefore my study supports the broader hypothesis that increased population structure promotes pathogen richness. Furthermore it contradicts the assumption that factors that promote high R_0 will automatically promote high pathogen richness (Morand 2000, Nunn et al. 2003). The positive relationship between increased population structure and pathogen richness implies that direct or indirect competitive mechanisms are acting such that increased population structure allows escape from competition which promotes pathogen richness.

This analysis is in agreement with two studies that have specifically tested this same hypothesis (Maganga et al. 2014, Turmelle & Olival 2009). These two studies used F_{ST} (Turmelle & Olival 2009) and fragmentation of species distributions (Maganga et al. 2014). Combined with the analysis here using the number

of subspecies, three different measures of population structure have been shown to correlate with pathogen richness in bats. By analysing data on two measures of population structure, and using larger data sets than previous studies, it is hoped that the results here may be more robust than in previous analyses (Gay et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009).

In contrast, one study Gay et al. (2014) found the opposite relationship using fragmentation of species distribution. Furthermore, Bordes et al. (2008) found no relationship between increased colony size and pathogen richness while Gay et al. (2014) found relationships in opposite directions for virus and ectoparasite richness. However, the study by Gay et al. (2014) uses relatively few species while the study by Bordes et al. (2008) uses group size which is a measure of local rather than global population structure. The overall weight of evidence suggests that poplulation structure and pathogen richness are associated. fur

There was strong support for a positive interaction between the number of subspecies and study effort. The support for this interaction implies that increased population structure has a stronger relationship with known pathogen richness in the presence of study effort. One interpretation of this is that increased population structure alone does not predict high known viral richness; reasonable study effort is also needed to turn the expected high viral richness into known and recorded viral richness. Biases in identification of subspecies have been noted before (Gippoliti & Amori 2007). The number of subspecies is more commonly used as a variable in comparative analyses of birds than mammals but the fact that it is associated with study effort is often not taken into account (Belliure et al. 2000, Phillimore et al. 2007).

Of the other explanatory variables considered, study effort and body mass were selected as being in the best model while there was marginal evidence for range size being associated with viral richness. Study effort positively predicted pathogen richness, confirming the expectation that additional study of a bat species yields more known viruses infecting that host species. Therefore, this bias cannot be ignored in studies using known pathogen richness as a proxy for total pathogen richness (Gregory 1990, Nunn et al. 2003). While body mass is selected as being in the best model in both the number of subspecies analysis and the effective gene flow analyis the estimated coefficients have opposite signs in the two analyses. In the number of subspecies analysis, body mass has a positive relationship with pathogen richness which is in agreement with previous studies (Bordes et al. 2008, Gay et al. 2014, Kamiya et al. 2014, Maganga et al. 2014,

Turmelle & Olival 2009). However, in the effective gene flow analysis, body mass has a negative estimated coefficient. This is in contrast to the number of subspecies analysis, previous studies in the literature and the single-predictor model. This result is probably due to correlations with other variables in the analysis and exacerbated by the small sample size in this analysis.

2.5.1 Broader implications

The relationship between increased population structure and pathogen richness suggests that population structure has at least some potential as being predictive of high pathogen richness and therefore of a species' likelihood of being a reservoir of a potentially zoonotic pathogen. However, given that it is difficult to measure population structure and given that the relationship appears to be weak at best, this trait on its own is unlikely to be useful in predicting zoonotic risk. However, as number of other factors are also associated with pathogen richness such as body mass and to a lesser extent range size as shown here as well as other traits studied elsewhere (Luis et al. 2013, Turmelle & Olival 2009). Therefore, using a combination of traits in a predictive (i.e. machine learning) framework has potential to be used in prioritising zoonotic disease surveillance. The main hurdle in this approach is finding a way to validate models; due to the study effort bias in current data, predictive models will also be biased. As unbiased pathogen surveys (e.g., Anthony et al. (2013)) become more common this may become possible. Alternatively, predictive models could be trained on all available — and therefore biased — data and validated by predicting smaller, unbiased data sets such as the data collected in Maganga et al. (2014).

The relationship between increased population structure and pathogen richness also has implications for habitat fragmentation and range shifts due to global change. In short, habitat fragmentation and range shifts that reduce movement between populations would be predicted to increase pathogen richness. However, depending on the mechanisms by which increased population structure increases pathogen richness this may not be a cause for concern. If the main mechanism is one that reduces pathogen extinction rates, a newly fragmented population is unlikely to increase its pathogen richness over any appreciable timescale. If, however, increased population structure actively promotes the evolution of new pathogen strains or allows the persistence of more virulent strains (Blackwood et al. 2013, Plowright et al. 2011, Pons-Salort et al. 2014) this could have important public health implications. Therefore further studies on the exact

mechanisms by which increased population structure affects pathogen richness is needed.

2.5.2 Study limitations

Although I have used measures of study effort to try to control for biases in the viral richness data, this bias could still make the results here unreliable — this is especially true as study effort is by far the strongest predictor of viral richness in both data sets. It is hoped that as untargeted sequencing of viral genetic material (e.g., Anthony et al. (2013)) becomes cheaper and more common this bias can be reduced. The strength of the relationship between study effort and known viral richness also highlights the number of bat-virus host-pathogen relationships yet to be discovered and the number of virus species that are yet to be described.

I have a number of explanatory variables to avoid spurious correlations. However, there is little data on bat density. Given that studies in other mammalian groups have found relationships between density and pathogen richness this would be a useful variable to include in further analyses (Arneberg 2002, Kamiya et al. 2014, Nunn et al. 2003). Acoustic monitoring is becoming cheaper and less labour intensive and may provide suitable data for estimating population densities for more bat species.

I have used two measures of population structure and the number of subspecies data set is larger than those used in previous studies. However it is clear that the gene flow data set is small ($n = 24$). This may explain some unexpected results. While the model averaging approach has given a negative model averaged coefficient for gene flow, the single-predictor model of gene flow against viral richness gave a positive coefficient. Furthermore body mass has a negative average coefficient. This is in contrast to the number of subspecies analysis, many studies in the literature (Gay et al. 2014, Kamiya et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009) and the single-predictor model. It is not easy to interpret these contradictions but it is clear that the results from the gene flow analysis alone should not be considered strong evidence for a relationship between increased population structure and pathogen richness. These contradictions also reiterate the need to use large data sets where possible and the need to use multiple measures of population structure to promote robust conclusions.

Finally, while comparative studies are a useful tool for examining broad trends of pathogen richness across large taxonomic groups, they cannot examine the specific mechanisms that may be underpinning the correlations found. There-

fore, further work is needed to test which mechanisms are actually causing the relationship between increased population structure and pathogen richness that I have identified here. A number of mechanisms might be involved. A reduced rate of pathogen extinction might be caused by a reduction in competition due to the slow dispersal of competing pathogens. Alternatively, increased population structure may promote the invasion of new pathogens, by creating localised areas of low competition or host immunity. One method for testing these mechanisms would be through mechanistic epidemiological models.

2.5.3 Conclusions

I have used phylogenetic linear models to identify positive relationships between two measures of population structure (the number of subspecies and effective levels of gene flow) and viral richness in bats. This study adds to the evidence that increased population structure may promote pathogen richness. It does not support the view that factors that increase R_0 will increase pathogen richness. Using larger data sets and multiple measurements makes the weight of the evidence here stronger than in previous studies. However, caution must still be taken in interpreting these results as the data is biased and particularly sparse in one of the analyses.

3 Understanding how population structure affects pathogen richness in a mechanistic model of bat populations

This work was conducted in collaboration with Kate Jones and Hilde Wilkinson-Herbots.

3.1 Abstract

An increasingly large fraction of emerging human diseases come from animals and these diseases have a huge impact on human health, healthcare systems and economic development. The chance that a new zoonosis will come from any particular wild host species increases with the number of pathogen species occurring in that host species. However, the factors that control pathogen richness of wild animal species are still unclear. Comparative, phylogenetic studies have shown that host-species traits such as population density, longevity and body size correlate with pathogen richness. Further comparative studies have found correlations between population structure and pathogen richness. Typically it is assumed that well-connected, unstructured populations (that therefore have a high basic reproductive number, R_0) promote the invasion of new pathogens and therefore increase pathogen richness. However, this assumption is largely untested. In the presence of inter-pathogen competition, the opposite effect might occur; increased population structure may increase pathogen richness by reducing the effects of competition. A more mechanistic understanding of how population structure affects pathogen richness could discriminate between these two broad hypotheses. I hypothesised that low dispersal rates and a low number of connections in a metapopulation network would allow invading pathogens to establish more easily, thus increasing pathogen richness. I tested these hypotheses using metapopulation networks parameterised to mimic wild bat populations as bats have highly varied social structures and have recently been implicated in a number of high profile diseases such as Ebola, SARS, Hendra and Nipah. I simulated the process of a new pathogen invading into a metapopulation already occupied by an identical pathogen. I varied the dispersal rate, topology of the metapopulation and transmission rate. I found significant evidence that increased dispersal rate increased the probability that a new pathogen would invade into a population. I found marginal evidence that network topology affected the probability that a new pathogen would invade. The assumption that factors causing high R_0 allow new pathogens to invade and therefore increase pathogen richness was supported. However, my results contradict many theoretical studies which predict that increased population structure should promote coexistence of pathogens. My results also contradict empirical patterns of pathogen richness with respect to population structure. Therefore, it is likely that population structure affects pathogen richness via a different mechanism to the one modelled here.

3.2 Introduction

Over 50% of emerging infectious diseases have an animal source (Jones et al. 2008, Smith et al. 2014). Zoonotic pathogens can be highly virulent (Lefebvre et al. 2014, Luby et al. 2009) and potentially have huge public health impacts (Granich et al. 2015), economic costs (Knobler et al. 2004) and slow down international development (World Bank 2014). Therefore understanding and predicting changes in the process of zoonotic spillover is a global health priority (Taylor et al. 2001). The number of pathogen species hosted by a wild animal species affects the chance that a disease from that species will infect humans (Wolfe et al. 2000). However, the factors that control the number of pathogen species in a wild animal population are still unclear (Metcalf et al. 2015); in particular our mechanistic understanding of how population processes inhibit or promote pathogen richness is poor.

In comparative studies, a number of host traits have been shown to correlate with pathogen richness including body size (Arneberg 2002, Kamiya et al. 2014), population density (Arneberg 2002, Nunn et al. 2003) and range size (Bordes & Morand 2011, Kamiya et al. 2014). A further factor that may affect pathogen richness is population structure. In comparative studies it is often assumed that factors that promote fast disease spread should promote high pathogen richness; the faster a new pathogen spreads through a population, the more likely it is to persist (Altizer et al. 2003, Morand 2000, Nunn et al. 2003, Poulin 2014, Poulin & Morand 2000). However, this assumption ignores competitive mechanisms such as cross-immunity and depletion of susceptible hosts. If competitive mechanisms are strong, pathogens in populations structured such that R_0 will be high will be able to easily out-compete invading pathogens. Only if competitive mechanisms are weak will high R_0 enable the invasion of new pathogens and allow higher pathogen richness.

Overall, the evidence from comparative studies indicates that increased population structure correlates with higher pathogen richness. This conclusion is based on studies using a number of measures of population structure: genetic measures, the number of subspecies, the shape of species' distributions and social group size (Chapter 2, Maganga et al. 2014, Turmelle & Olival 2009, Vitone et al. 2004). However, there are number of studies that contradict this conclusion (Bordes et al. 2007, Ezenwa et al. 2006, Gay et al. 2014). Comparative studies are often contradictory due to small sample sizes, noisy data and because empirical relationships often do not extrapolate well to other taxa. Furthermore, multi-

collinearity between many traits also makes it hard to clearly distinguish which factors are important (Nunn et al. 2015). However, meta-analyses can be used to combine studies to help generalise conclusions (Kamiya et al. 2014).

Furthermore, knowing which factors correlate with pathogen richness does not tell us if, or how, they causally control pathogen richness. This lack of a solid mechanistic understanding of these processes precludes predictions of how wild populations will respond to peterbations such as increased human pressure and global change. As habitats fragment we expect wild populations to change in a number of ways including becoming smaller and less well connected (Andren 1994, Cushman et al. 2012). As multiple population-level factors are likely to change simultaneously due to global change, the correlative relationships examined in comparative studies are unlikely to effectively predict future changes in pathogen richness. Mechanistic models are needed to project how these highly non-linear disease systems will respond to the multiple, simultaneous stressors affecting them.

There are a number of mechanisms by which population structure could increase pathogen richness. Firstly, population structure may reduce competition between pathogens. In analytical models of well-mixed populations competitive exclusion has been predicted (Ackleh & Allen 2003, Allen et al. 2004, Bremermann & Thieme 1989, Martcheva & Li 2013, Qiu et al. 2013). When competitive exclusion occurs, population structure has sometimes been shown to allow coexistence (Allen et al. 2004, Garmer et al. 2016, Nunes et al. 2006, Qiu et al. 2013). Alternatively, population structure may promote the evolution of new strains within a species (Buckee et al. 2004), reduce the rate of pathogen extinction (Rand et al. 1995) or increase the probability of pathogen invasion from other host species (Nunes et al. 2006). These separate mechanisms have not been examined and it is difficult to see how they could be distinguished through comparative methods.

Currently the literature contains very abstract, simplified models (Allen et al. 2004, Garmer et al. 2016, Nunes et al. 2006, Qiu et al. 2013). These cannot be easily applied to real data. They also do not easily give quantitative predictions of pathogen richness; typically they predict either no pathogen coexistence (Bremermann & Thieme 1989, Martcheva & Li 2013) or infinite pathogen richness (May & Nowak 1994). Models that can give quantitive predictions of pathogen richness in wild populations are more applicable to real-world issues such as zoonotic disease surveillance. While predicting an absolute value of pathogen richness

for a wild species is likely to be impossible, models that attempt to rank species from highest to lowest pathogen richness are still useful for prioritising species for surveillance. This requires a middle ground of model complexity.

In order to capture this middle ground, I have used metapopulation models. Unlike two-patch models, that are used to add population structure while keeping model complexity to a minumum (Allen et al. 2004, Garmer et al. 2016, Qiu et al. 2013), the metapopulations used here split the population into multiple subpopulations. I have used two independant variables that alter population structure, dispersal rate and metapopulation network topology. I have studied the invasion of new pathogens as a mechanism for increasing pathogen richness. In particular I have focused on studying the invasion of a newly evolved pathogen that is therefore identical in epidemiological parameters to the endemic pathogen. Furthermore, this close evolutionary relationship means that competition via cross-immunity is strong.

The metapopulations were parameterised to broadly mimic wild bat populations. Population structure has already been found to correlate with pathogen richness in bats (Chapter 2, Gay et al. 2014, Maganga et al. 2014, Turmelle & Olival 2009). Furthermore, bats have an unusually large variety of social structures. Colony sizes range from 10 to 1 million (Jones et al. 2009) individuals and colonies can be very stable (Kerth et al. 2011, McCracken & Bradbury 1981). This strong colony fidelity means they fit the assumptions of metapopulations well. Bats have also, over the last decade, become a focus for disease research (Calisher et al. 2006, Hughes et al. 2007). The reason for this focus is that they have been implicated in a number of high profile diseases including Ebola, SARS, Hendra and Nipah (Calisher et al. 2006, Li et al. 2005).

Here I show that, given the assumptions of a metapopulation, increased dispersal significantly increased the probability of invasion of new pathogens. Furthermore, structured populations nearly always had a lower probability of pathogen invasion than a fully-mixed population of equal size. The topology of the network did not strongly affect the probability of pathogen invasion as long as the population was not completely unconnected. Overall, I found significant evidence that reduced population structure increases the probability of invasion of a new pathogen, implying a role for the generation of pathogen richness more generally.

3.3 Methods

3.3.1 Two pathogen SIR model

I developed a multipathogen, SIR compartment model with individuals being classed as susceptible, infected or recovered with immunity (Figure 3.1). Susceptible individuals are counted in class S (see Table 3.1 for a list of symbols and values used). There are three infected classes, I_1, I_2 and I_{12}, being individuals infected with pathogen 1, pathogen 2 or both respectively . Recovered individuals, R, are immune to both pathogens, even if they have only been infected with one. Furthermore, recovery from one pathogen moves an individual straight into the recovered class, even if the individual is infected with both pathogens (Figure 3.1). This modelling choice allows the model to be easily expanded to include more than two pathogens though here I only simulate two pathogens. The assumption of immediate recovery from all other diseases is likely to be reasonable for very closely related pathogens as is being studied here. Once an acquired immune response is activated, all infections are likely to be cleared quickly and upregulation of the innate immune system will speed up recovery. The coinfection rate (the rate at which an infected individual is infected with a second pathogen) is adjusted compared to the infection rate by a factor α (here I used $\alpha = 0.1$ which means coinfection happens at a tenth the rate of first infections). Both negative and positive associations have been found between closely related bat viruses (Anthony et al. 2013). I modelled a strong negative association between viruses based on the aim of modelling a newly evolved pathogen.

In the application of long term existence of pathogens it is necessary to include vital dynamics (births and deaths) as the SIR model without vital dynamics has no endemic state. Birth and death rates (μ and Λ) are set as being equal meaning the population does not systematically increase or decrease. The population size does however change as a random walk. New born individuals enter the susceptible class. The time scale of the simulations are determined by setting μ = 0.05 per year, yielding an average host generation time of 20 years. Infection amd coinfection were assumed to cause no extra mortality as for a number of viruses, bats show no clinical signs of infection (de Thoisy et al. 2016, Halpin et al. 2011). In humans, coinfection generally worsens health (Griffiths et al. 2011) but as there are

The population is modelled as a metapopulation, being divided into a number of subpopulations (colonies). This model is an intermediate level of com-

plexity between fully-mixed populations and contact networks. The existence of subspecies, measurements of genetic dissimilarity and ecological studies provide ample evidence that bat populations are structured to some extent (Burns & Broders 2014, Kerth et al. 2011, McCracken & Bradbury 1981, Wilson & Reeder 2005). Therefore a fully mixed population is a large oversimplification. However, trying to study the contact network relies on knowledge of detailed individual behaviour which is not available.

The metapopulation is modelled as a network with colonies being nodes and dispersal between colonies being indicated by edges (Figure 3.2). Individuals within a colony interact randomly so that the colony is fully mixed. Dispersal between colonies occurs at a rate ξ. Individuals can only disperse to colonies connected to theirs in the network. The rate of dispersal is not affected by the number of edges a colonies has (known as the degree of the colony and denoted k). Therefore, the dispersal rate from a colony x with degree k_x to colony y is ξ/k_x. Note this rate is independent of the degree and size of colony y.

I examined this model using stochastic, continuous-time simulations implemented in *R* (R Development Core Team 2010). The implementation is available as an R package on GitHub (Lucas 2015a). The model can be written as a continuous-time Markov chain. The Markov chain contains the random variables $((S_x)_{x=1\ldots m}, (I_{x,q})_{x=1\ldots m,\, q\in\{1,2,12\}}, (R_x)_{x=1\ldots m})$. Here, $(S_x)_{x=1\ldots m}$ is a length m vector of the number of susceptibles in each colony. $(I_{x,q})_{x=1\ldots m, q\in\{1,2,12\}}$ is a length $m\times 3$ vector describing the number of individuals of each disease class ($q \in \{1, 2, 12\}$) in each colony. Finally, $(R_x)_{x=1\ldots m}$ is a length m vector of the number of individuals in the recovered class. The model is a Markov chain where extinction of both pathogens species and extinction of the host species are absorbing states. However, the expected time to reach this state is much larger than the duration of the simulations.

At any time, suppose the system is in state $((s_x), (i_{x,q}), (r_x))$. At each step in the simulation we calculate the rate at which each possible event might occur. One event is then randomly chosen, weighted by its rate

$$p(\text{event } i) = \frac{e_i}{\sum_j e_j}, \tag{3.1}$$

where e_i is the rate at which event i occurs and $\sum_j e_j$ is the sum of the rates of all possible events. Finally, the length of the time step, δ, is drawn from an exponential distribution

$$\delta \sim \text{Exp}\left(\sum_i e_i\right). \tag{3.2}$$

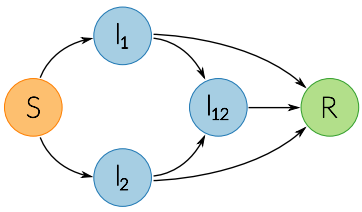


Figure 3.1 Schematic of the SIR model used. Individuals are in one of five classes, susceptible (orange, S), infected with pathogen 1, pathogen 2 or both (blue, I_1, I_2, I_{12}) or recovered and immune from further infection (green, R). Transitions between classes occur only as indicated by arrows. Note that individuals in I_{12} move into R, not back to I_1 or I_2. That is, recovery from one pathogen causes immediate recovery from the other pathogen.

We can now write down the rates of all events. Assuming asexual reproduction, that all classes reproduce at the same rate and that individuals are born into the susceptible class we get

$$s_x \to s_x + 1 \text{ at a rate of } \Lambda\left(s_x + \sum_q i_{qx} + r_x\right) \tag{3.3}$$

where $s_x \to s_x + 1$ is the event that the number of susceptibles in colony x will increase by 1 (a single birth) and $\sum_q i_{qx}$ is the sum of all infection classes $q \in \{1, 2, 12\}$. The rates of death, given a death rate μ, and no increased mortality due to infection, are given by

$$s_x \to s_x - 1 \text{ at a rate of } \mu s_x, \tag{3.4}$$

$$i_{qx} \to i_{qx} - 1 \text{ at a rate of } \mu i_{qx}, \tag{3.5}$$

$$r_x \to r_x - 1 \text{ at a rate of } \mu r_x. \tag{3.6}$$

I modelled transmission as being density-dependent. This assumption was more suitable than frequency-dependent transmission as I was modelling a disease transmitted by saliva or urine in highly dense populations confined to caves, buildings or potentially a small number of tree roosts. I was notably not modelling an STD as spillover of STDs from bats to humans is is likely to be rare. Infection of a susceptible with either pathogen 1 or 2 is therefore given by

$$i_{1x} \to i_{1x} + 1, s_x \to s_x - 1 \text{ at a rate of } \beta s_x \left(i_{1x} + i_{12x}\right), \tag{3.7}$$

$$i_{2x} \to i_{2x} + 1, s_x \to s_x - 1 \text{ at a rate of } \beta s_x \left(i_{2x} + i_{12x}\right), \tag{3.8}$$

while coinfection, given the coinfection adjustment factor α, is given by

$$i_{12,x} \to i_{12,x} + 1,\ i_{1x} \to i_{1x} - 1 \text{ at a rate of } \alpha\beta i_{1x}\left(i_{2x} + i_{12x}\right), \tag{3.9}$$

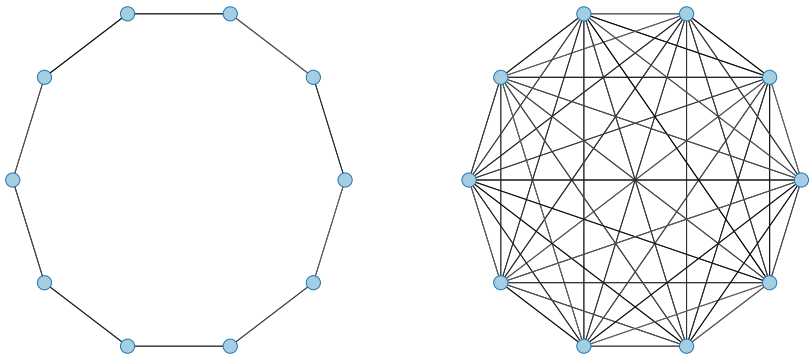
$$i_{12,x} \to i_{12,x} + 1,\ I_{2x} \to i_{2x} - 1 \text{ at a rate of } \alpha\beta i_{2x}\left(i_{1x} + i_{12x}\right). \tag{3.10}$$

Note that lower values of α give lower rates of infection as in Castillo-Chavez et al. (1989). $\alpha = 0$ would be the case of no coinfection while $\alpha = 1$ would be the case where coinfection occurs at the same rate as the initial infection.

The probability of migration from colony y (with degree k_y) to colony x, given

Table 3.1 A summary of all symbols used in Chapters 3 and 4 along with their units and default values. The justification for parameter values is given in Section 3.3.2.

Symbol	Explanation	Units	Value
ρ	Number of pathogens		2
x, y	Colony index		
p	Pathogen index i.e. $p \in \{1, 2\}$ for pathogens 1 and 2		
q	Disease class i.e. $q \in \{1, 2, 12\}$		
S_x	Number of susceptible individuals in colony x		
$I_q x$	Number of individuals infected with disease(s) $q \in 1, 2, 12$ in colony x		
R_x	Number of individuals in colony x in the recovered with immunity class		
N	Total Population size		30,000
m	Number of colonies		10
n	Colony size		3,000
a	Area	km^2	10,000
β	Transmission rate		0.1 – 0.4
α	Coinfection adjustment factor.	Proportion	0.1
γ	Recovery rate	$year^{-1}.individual^{-1}$	1
ξ	Dispersal	$year^{-1}.individual^{-1}$	0.001–0.1
Λ	Birth rate	$year^{-1}.individual^{-1}$	0.05
μ	Death rate	$year^{-1}.individual^{-1}$	0.05
k_x	Degree of node x (number of colonies that individuals from colony x can disperse to).		
δ	Waiting time until next event	years	
e_i	The rate at which event i occurs	$year^{-1}$	



A Fully connected **B** Minimally connected

Figure 3.2 The two network topologies used to test whether network connectedness influences a pathogens ability to invade. A) Animals can only disperse to neighbouring colonies. B) Dispersal can occur between any colony. Blue circles are colonies of 3,000 individuals. Dispersal only occurs between colonies connected by an edge (black line). The dispersal rate is held constant between the two topologies.

a dispersal rate ξ is given by

$$s_x \to s_x + 1,\ s_y \to s_y - 1 \ \text{ at a rate of } \ \frac{\xi s_y}{k_y}, \tag{3.11}$$

$$i_{qx} \to i_{qx} + 1,\ i_{qy} \to i_{qy} - 1 \ \text{ at a rate of } \ \frac{\xi i_{qy}}{k_y}, \tag{3.12}$$

$$r_x \to r_x + 1,\ r_y \to r_y - 1 \ \text{ at a rate of } \ \frac{\xi r_y}{k_y}. \tag{3.13}$$

Not that the dispersal rate does not change with infection. As above, this is due to the low virulence of bat viruses. Finally, recovery from any infectious class occurs at a rate γ

$$i_{qx} \to i_{qx} - 1,\ r_x \to r_x + 1 \ \text{ at a rate of } \ \gamma i_{qx}. \tag{3.14}$$

3.3.2 Parameter selection

The fixed parameters were chosen to roughly reflect realistic wild bat populations. The death rate μ was set as 0.05 per year giving a generation time of 20 years. The birth rate Λ was set to be equal to μ. This yields a population that does not

systematically increase or decrease. However, the size of each colony changes as a random walk. Given the length of the simulations, colonies were very unlikely to go extinct (Figure B.3) The recovery rate γ was set to one, giving an average infection duration of one year. This is therefore a long lasting infection but not a chronic infection. It is very difficult to directly estimate infection durations in wild populations but it seems that these infections might sometimes be long lasting (Peel et al. 2012, Plowright et al. 2015). However, other studies have found much shorter infectious periods (Amengual et al. 2007). These shorter infections are not studied further here.

The coinfection adjustment parameter, α, was set to 0.1 so that an individual infected with one pathogen is 90% less likely to be infected with another. This is a rather arbitrary value. However, the rationale of the model was that the invading species might be a newly speciated strain of the endemic species. Furthermore, the model assumes complete cross-immunity after recovery from infection. Therefore cross-immunity to coinfection is likely to be very strong as well.

The starting size of each colony was set to 3,000. This is appropriate for many bat species (Jones et al. 2009), especially the large, frugivorous *Pteropodidae* that have been particularly associated with recent zoonotic diseases. Four values of the transmission rate β were used, 0.1, 0.2, 0.3 and 0.4. These values were chosen to cover the range of behaviours, from very high probabilities of invasion of the second pathogen, to very low probabilities. All simulations were run under all four transmission rates as this is such a fundamental parameter.

3.3.3 Experimental setup

The metapopulation contains 10 colonies. 10 colonies was selected as a trade off between computation time and a network complex enough that structure might have an effect. This value is artificially small compared to wildlife populations. In each simulation the population was seeded with 10 sets of 200 infected individuals of pathogen 1. These groups were seeded into randomly selected colonies with replacement. For each 200 infected individuals added, 200 susceptible individuals were removed to keep starting colony sizes constant. Pathogen 1 was then allowed to spread until the initial, large epidemic had ended. Visual inspection of preliminary simulations was used to decide on 3×10^5 as being long enough for the epidemic to end and the pathogen to be in an endemic state (Figures B.1 and B.2). After 3×10^5 events, five individuals infected with pathogen 2 were added to one randomly selected colony. After another 5×10^5 events the invasion of

pathogen 2 was considered successful if any individuals were still infected with pathogen 2. Therefore, if at least one individual was in class I_2 or I_{12} at the end of the simulation, this was considered an invasion. Again visual inspection of preliminary simulations was used to determine that after 5×10^5 events, if an invading pathogen was still present, it was well established (Figures B.1 and B.2). The choice to use a fixed number of events, rather than fixed amount number of years, was for computaiton convenience. However, this choice creates a risk of bias as simulations with more events per unit time will last for a shorter time overall. However, visual inspection of the dynamics of disease extinction (B.2), and examination of the typical time to extinction implies that this bias is very weak. For example, of the simulations where extinction occurred, the extinction occurred more than 50 years before the end of the simulation in 90% of cases. On a preliminary run of 106 simulations across all combinations of dispersal and transmission rates, examining the population after 7×10^5 events instead of 8×10^5 events gave exactly the same result with respect to the binary state of invasion or no invasion.

3.3.4 Population structure

As a baseline for comparison, I ran simulations of a fully unstructured population. These simulations were run with a population of 30,000 so that the total population size was equal to that of the total metapopulation size in the structured simulations. I ran 100 simulations at each transmission rate.

Two parameters control population structure in the model: dispersal rate and the topology of the metapopulation network. The values used for these parameters were chosen to highlight the effects of population structure. I selected the dispersal values $\xi = 0, 0.1, 0.01$ and 0.001 dispersals per individual per year. The probability that an individual disperses at least once in its lifetime is given by $\xi / (\xi + \mu)$. $\xi = 0.1$ relates to 67% of individuals dispersing between colonies at least once in their lifetime. Therefore exclusively juvenile dispersal would have dispersal rates similar to this value. $\xi = 0.01$ relates to 17% of individuals dispersing at least once in their lifetime. This value is relatively close to male-biased dispersal, with female philopatry. $\xi = 0.001$ relates to 2% of individuals dispersing in their lifetime. This therefore relates to a species that does not habitually disperse. Finally, I ran simulations with no dispersal. Given zero dispersal, only the colony seeded with pathogen 2 can ever recieve infections of the invading pathogen. Therefore, only one colony was simulated for $\xi = 0$. I ran 100 simu-

lations for most parameter sets. However, examing these simulations had some marginal and difficult to interpret results. Therefore, I ran 150 simulations for $\beta = 0.2$ and 0.3 (not including the zero dispersal case).

I also altered the topology of the metapopulation network. The network topology was created to be either fully or minimally connected (Figure 3.2). To model a completely unconnected population the $\xi = 0$ simulations from above were used. I again ran 100 simulations for each parameter set.

3.3.5 Statistical analysis

I used generalised linear models (GLMs) with a binary response variable, invasion or not, to test the hypothesis that probability of invasion increased with dispersal. Seperate GLMs were fitted for each transmission rate. These tests were performed both with and without the $\xi = 0$ results as the complete lack of dispersal makes these simulations qualitatively different to the other simulations. To test whether the different topologies had different probabilities of invasion I used Fisher's exact tests as topology is best described as a categorical variable. As with the $\xi = 0$ results, these tests were performed both with and without the completely unconnected topology results. Finally, I also used binomial GLMs to test the hypothesis that the probability of invasion increased with transmission rate. Seperate GLMs were fitted for each dispersal rate and network topology. All statistical analyses were performed using the *stats* package in *R*. The code used for running the simulations and analysing the results is available at GitHub.

3.4 Results

3.4.1 Dispersal

In the unstructured population, the second pathogen invaded in 100 out of 100 simulations. This was true at all four transmission rates.

When the $\xi = 0$ simulations were included, there was a positive relationship between dispersal rate and invasion probability for $\beta = 0.2, 0.3$ and 0.4 (Figure 3.3A). These positive relationships were all significant (GLM. $\beta = 0.2$: $b = 12.59$, $p < 10^{-5}$. $\beta = 0.3$: $b = 12.07$, $p = 7.62 \times 10^{-4}$. $\beta = 0.4$: $b = 13.44$, $p = 0.03$.) At $\beta = 0.1$ there was no significant relationship as invasion probaiblity was very close to zero at all dispersal rates (GLM. $b = -220.19$, $p = 0.62$).

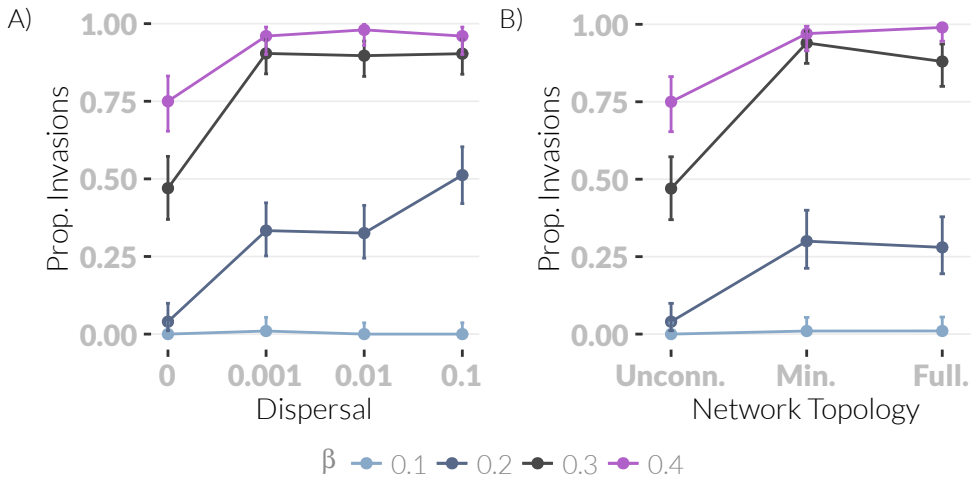


Figure 3.3 The probability of successful invasion for different A) dispersal rates or B) network topologies (with network topologies "unconnected", "minimally connected" and "fully connected" as in Figure 3.2). Error bars are 95% confidence intervals of probability of invasion. 100 simulations were run for each treatment except $\beta = 0.2$ in A) which has 150 per treatment. Other parameters were kept constant at: $m = 10$, $\mu = \Lambda = 0.05$, $\gamma = 1$, $\alpha = 0.1$. When dispersal is varied, the population structure is fully connected. When network topology is varied, $\xi = 0.01$.

However, when the $\xi = 0$ simulations were removed, this significant, positive relationship largely disappeared. At $\beta = 0.2$, the significant positive relationship remained (GLM: $b = 7.97$, $p = 7.82 \times 10^{-4}$). At all other transmission rates, the probability of invasion did not significantly change with dispersal rate (GLM. $\beta = 0.1$: $b = -1.93 \times 10^{3}$, $p = 1$. $\beta = 0.3$: $b = 0.27$, $p = 0.94$. $\beta = 0.4$: $b = -2.7$, $p = 0.70$.)

3.4.2 Network structure

When the completely unconnected topology simulations were included, the probability of invasion was different across topologies for $\beta = 0.1, 0.2$ and 0.3 (Fisher's exact test. $\beta = 0.2$: $p < 10^{-5}$. $\beta = 0.3$: $p < 10^{-5}$. $\beta = 0.4$: $p < 10^{-5}$). In each case, the fully unconnected population had a lower probability of invasion than the minimally and completely connected topologies (Figure 3.3B). At $\beta = 0.1$ there was no significant difference ($p = 0.77$) and the probability of invasion was close to zero for all topologies (Figure 3.3B).

When the completely unconnected topology simulations were removed, there

were no significant differences between topologies i.e. between the minimally and fully connected topologies (Figure 3.3B). This was true at all transmission rates (Fisher's exact test. $\beta = 0.1$, $p = 1.00$. $\beta = 0.2$, $p = 0.88$. $\beta = 0.3$, $p = 0.22$. $\beta = 0.4$, $p = 0.62$).

3.4.3 Transmission

Increasing the transmission rate increased the probability of invasion (Figure 3.3). This is true for all four dispersal values (GLM. $\xi = 0$: $b = 19.73$, $p < 10^{-5}$. $\xi = 0.001$: $b = 26.75$, $p < 10^{-5}$. $\xi = 0.01$: $b = 29.56$, $p < 10^{-5}$. $\xi = 0.1$: $b = 24.74$, $p < 10^{-5}$.) and both network structures (GLM. Minimally connected: $b = 30.4$, $p < 10^{-5}$. Fully connected: $b = 30.06$, $p < 10^{-5}$).

3.5 Discussion

I have used mechanistic, metapopulation models to test whether increased population structure can promote pathogen richness by facilitating invasion of new pathogens. I found that dispersal does affect the ability of a new pathogen to invade and persist in a population. I also found evidence that the pathogen invasion is less likely in completely isolated colonies. However, as long as the metapopulation is slightly connected, the topology of the metapopulation network did not affect invasion probability. Increasing transmission rate quickly reaches a state where new pathogens always invade aslong there is some dispersal while decreasing the transmission rate quickly reaches a state where invasion is impossible.

That increased population structure increases pathogen richness supports many predictions that increasing R_0 should increase pathogen richness (Altizer et al. 2003, Morand 2000, Nunn et al. 2003, Poulin 2014, Poulin & Morand 2000). However, many comparative studies have found the opposite relationship, with increased population structure increasing pathogen richness (Chapter 2, Maganga et al. 2014, Turmelle & Olival 2009, Vitone et al. 2004). Furthermore, simple analytical models suggest that population structure should increase pathogen richness (Allen et al. 2004, Nunes et al. 2006, Qiu et al. 2013) and I find no evidence of this.

These results suggest that if population structure does in fact affect pathogen richness, as observed in comparative studies (Chapter 2, Maganga et al. 2014, Turmelle & Olival 2009, Vitone et al. 2004), it must occur by a mechanism other

than the one studied here. Here, the hypothesised mechanism for the relationship between population structure and pathogen richness was that the spread and persistence of a newly evolved pathogen is facilitated by decreased competition due to population structure. Instead, reduced population structure allows the new pathogen to quickly spread outside of the colony in which it evolved. As the mechanism studied here cannot explain the relationship between population structure and pathogen richness seen in wild species (Chapter 2, Maganga et al. 2014, Turmelle & Olival 2009, Vitone et al. 2004), other mechanisms should be studied. Other mechanisms that should be examined include reduced competitive exclusion of already established pathogens or increased invasion of less closely related and less strongly competing pathogens, perhaps mediated by ecological competition of pathogens (i.e. reduction of the susceptible pool by disease induced mortality). Furthermore, single pathogen dynamics could have an important role such as population structure causing a much slower, asynchronous epidemic preventing acquired herd immunity (Plowright et al. 2011).

I ran simulations of a completely unstructured population as a baseline comparison of pathogen invasion probability. However, this unstructured population could also be considered one, very large, subpopulation or colony. The fact that invasion occured 100% of the time in these simulations suggest that colony size has an important role in pathogen richness. Therefore the interplay between population structure and colony size should be studied further especially as the range of colony size in bats is large, ranging from 10 to 1 million (Jones et al. 2009) individuals.

My simulations also highlight the importance of competition for the spread of a new pathogen. All parameters used corresponded to pathogens with $R_0 > 1$ (as seen in the unstructured simulations). However, the competition with the endemic pathogen means that for some transmission rates the chance of epidemic spread and persistence is close to zero. This has implications for human epidemics as well — if there is strong competition between a newly evolved strain and an endemic strain, we are unlikely to see the new strain spread, regardless of population structure.

3.5.1 Model assumptions

Complete cross-immunity I have assumed that once recovered, individuals are immune to both pathogens. Furthermore, when a coinfected individual recovers from one pathogen, it immediately recovers from the other as well. This is prob-

ably a reasonable assumption given that I am modelling a newly evolved strain. The recovery rate of coinfected individuals has not been well studied however. In humans, recent recovery from a co-circulating virus, increased the rate of recovery from respiratory syncytial virus (though co-infected individuals recovered more slowly) (Munywoki et al. 2015).

However, further work could relax this assumption using a model similar to (Poletto et al. 2015) which contains additional classes for 'infected with pathogen one, immune to pathogen two' and 'infected with pathogen two, immune to pathogen one'. The model here was formulated such that the study of systems with greater than two pathogens is still computationally feasible while a model such as used in (Poletto et al. 2015) contains 3^ρ classes for a system with ρ pathogen species. This quickly becomes computationally restrictive. The interaction of multiple pathogens is one avenue for further study. It might be expected that there is an upper limit to the total number of pathogen species that can coexist in a population. In particular, it is possible that once a certain number of species are endemic in a population, no more pathogens can invade into the population. This has not been studied in the context of metapopulations.

Identical strains Many papers on pathogen richness have focused on the evolution of pathogen traits and have considered a trade off between transmission rate and virulence (Nowak & May 1994) or infectious period (Poletto et al. 2013). However, here I am interested in host traits. Therefore we have assumed that pathogen strains are identical. It is clear however that there are a number of factors that affect pathogen richness and our focus on host population structure does not imply that pathogen traits are not important.

Complex social structure and behaviour With the models here I have aimed to tread a middle ground between the overly simple models employed in analytical studies (Allen et al. 2004) and the full complexity and variety of true bat social systems (Kerth 2008). Omissions include seasonal migration, maternity roosts, hibernation roosts and swarming sites (Cryan et al. 2014, Fleming et al. 2003, Kerth 2008, Richter & Cumming 2008). While future models might aim to model this complexity more fully, the number of parameters that are required to be estimated and varied becomes very large. Furthermore, not all of these social complexities exist in all bat species, so in limiting my analysis to the simpler end of bat social systems it is hoped that the results are more broadly representative of the order.

Furthermore, I have considered a single host species in isolation. It seems likely that sympatry in bats and other mammals is epidemiologically important (Brierley et al. 2016, Luis et al. 2013, Pilosof et al. 2015) but was beyond the scope of this study. There is potential for this to be effectively modelled as a multi-layered network (Funk & Jansen 2010, Wang et al. 2016) and this would be expected to act to reduce population structure. Conversely, the case of interspecies roost sharing could be modelled as an additional layer of within-colony, population structure which would tend to increase population structure.

Finally, many species of bat exhibit strong seasonal birth pulses which are known to affect disease dynamics (Amman et al. 2012, Hayman 2015, Peel et al. 2014). This would be expected to facilitate the invasion of new pathogen species; if a new strain evolved or entered the population by migration during a period of low population immunity, it would have a higher chance of invading and establishing in the population.

3.5.2 Conclusions

In conclusion I have found evidence that reduced population structure facilitates the invasion and establishment of newly evolved pathogen species. However, the direction of the relationship contradicts those found in wild species. This suggests that if population structure does have a role in shaping pathogen communities, it is unlikely to be by this specific mechanism.

4 A mechanistic model to compare the importance of interrelated population measures: population size, population density and colony size

This work was conducted in collaboration with Kate Jones and Hilde Wilkinson-Herbots.

4.1 Abstract

An increasingly large fraction of emerging diseases come from wild animals and these diseases have a huge impact on human health. The chance that a new disease will come from any particularly host species increases with the diversity of pathogens in that species. However, the factors that control pathogen diversity in wild populations are still poorly understood.

Population density is thought to increase pathogen richness while theory suggests that population structure and population size may also play a role. However, these factors are intrinsically linked — reducing density reduces contacts between individuals and directly reduces population size. In group living species group size and the total number of groups both contribute to total population size. As these factors are all completely interdependent, it is very difficult to study them empirically e.g., in a comparative frame work.

It is unknown whether it is specifically density that controls pathogen diversity or whether density merely correlates with other causal factors such as population structure, group size or population size. Here I use metapopulation SIR models to test whether it is density *per se* that increases the ability of a newly evolved pathogen to invade and persist in a population as apposed to colony size, population size or population structure.

I found that increased group size increases the chance that a new pathogen will invade into a population to the largest extent. Both group size and the number of groups (i.e. the components of population size) promote pathogen richness more than population density. This implies that, in comparative studies, population density is merely a correlate of group size or population size. As these factors are not equally important it is expected that the pathogen communities of different host species will respond to climate change in different ways. Species which experience changes in group size are expected to have larger changes in pathogen richness than other species.

This study helps clarify both the inter-relationships between, and relative importance of, a number of population level factors affecting pathogen richness. It also highlights the necessity for studying the mechanisms underlying pathogen community construction as comparative approaches do not have the specificity to do so.

4.2 Introduction

Zoonotic diseases are an increasingly important source of human infectious diseases (Jones et al. 2008, Taylor et al. 2001, Woolhouse & Gowtage-Sequeria 2006). The diversity of pathogens in wild animal populations is huge and largely unknown (Poulin 2014). Furthermore, the factors that allow large numbers of pathogen species to coexist in a host (e.g., Anthony et al. (2013)) are still unclear. It is well known that population level factors such as population density, range size and population structure have an important role in controlling pathogen community dynamics (Anderson & May 1979, Colizza & Vespignani 2007, May & Anderson 1979, May & Lloyd 2001). Global change is strongly perturbing wild animal populations (Craigie et al. 2010, Thomas et al. 2004), but without clear mechanistic models of how these populations maintain pathogen species richness, we can not predict how pathogen communities, and the risks of zoonotic outbreaks, will change in the coming decades.

Variables that describe populations, such as population density and structure, are well established as having a central role in pathogen dynamics (Anderson & May 1979, Barthélemy et al. 2010, Colizza & Vespignani 2007, May & Anderson 1979, Wu et al. 2013). More recently, the role of the population has been examined with respect to pathogen richness and the coexistence of competing pathogens (Allen et al. 2004, Nunes et al. 2006, Qiu et al. 2013). Yet even in theoretical studies there is confusion as to how exactly we should measure populations. There is disagreement on whether population density (individuals per unit area) should be preferred over population size (number of individuals) and how exactly area should be incorporated (Begon et al. 2002).

With the increase of novel zoonotic pathogens (Jones et al. 2008) attention has turned to comparatively assessing the factors that are associated with high or low pathogen richness in wild animal species (Poulin & Morand 2000). Here again there is little clarity on the relationship between a number of species measurements. Population density is commonly studied (Arneberg 2002, Kamiya et al. 2014, Lindenfors et al. 2007, Morand & Poulin 1998, Nunn et al. 2003) as is range size (Huang et al. 2015, Kamiya et al. 2014, Lindenfors et al. 2007, Nunn et al. 2003, Turmelle & Olival 2009). However it is rarely if ever acknowledge that these two values are intrinsically linked by $d = N/a$ where d is density, N is the population size and a is area (See Table 3.1 for all parameters used) or that the relationship $N \propto a$ has broad empirical support (Blackburn et al. 2006, Borregaard & Rahbek 2010). In contrast, population size has never been directly studied as

a predictor of pathogen richness — although confusingly, population range size is sometimes used as a measure of size e.g., (Vögeli et al. 2011). Furthermore, population size is considered the more relevant measure in terms of pathogen dynamics, especially when area cannot be assumed to be constant (Begon et al. 2002) as is commonly the case in wild populations, especially in the face of global warming and habitat degradation.

It is clear that animals are neither randomly distributed in space nor epidemiologically 'will-mixed': social groups are common (Kerth 2008) and distance and geographic boundaries reduce contacts between isolated populations (Jenkins et al. 2010, Peel et al. 2012). In social species, measures such as global population density are largely meaningless with respect to the number of infectious contacts individuals may have. Rather, contacts are based on group size and rates of movements between groups. Two aspects of non-random transmission have been studied in particular: group size (Ezenwa et al. 2006, Gay et al. 2014, Nunn et al. 2003, Rifkin et al. 2012, Vitone et al. 2004) and global measures of population structure including genetic measures and measures derived from geographic distribution shapes (Gay et al. (2014), Maganga et al. (2014), Turmelle & Olival (2009) and see Chapter 2). Again however, the relationships between these terms and range size, population size and density are rarely examined. Population size can be decomposed into two components, the number of groups and the average size of a group with $N = nm$ where n is group size and m is the number of groups. The amount of movement between groups is at least partially dependent on the distance between them (Jenkins et al. 2010, Le Galliard et al. 2012, Schooley & Branch 2009). The distance between neighbouring groups decreases with the number of groups per area m/a or N/na.

Importantly, these factors, although interrelated, will respond differently to global change and the response will be species dependent. Some species may suffer large range contractions, and therefore large falls in population size, while their density remains fairly constant. Other species might retain their distribution but have a depressed population density. Similarly with population structure, species particularly affected by habitat fragmentation can expect increased reduced movement of individuals between groups, while other species may be most affected by a reduction in group size. Furthermore, different mechanisms of maintenance and creation of pathogen richness will respond to changes in these factors differently as well. If pathogen richness ultimately depends on the "island size" of the host population, then falls in population size will reduce pathogen

richness the most. If local group size affects the ability of new pathogens to invade (Nunn et al. (2003), Chapter 3) then changes in group size are likely to be more important. Finally, if increased population structure allows pathogens to coexist (Allen et al. (2004), Nunes et al. (2006) and Qiu et al. (2013) and Chapter 2) increased habitat fragmentation could be expected to increase pathogen richness.

As these population factors — population size and density, range size and group size — are likely to be intercorrelated, correlative comparative studies will struggle to distinguish between them. Furthermore, even if some factors are statistically supported or rejected, the specific mechanisms by which they promote pathogen richness will remain unknown, and these may suggest different responses to global change. Finally, mechanistic models are expected to be more predictive into the future and into hitherto unseen population regimes.

Therefore there is great need for mechanistic models that try to disentangle the interplay between these many factors: density, population size, range size, population structure, group size and the number of groups. Here, I have used multipathogen, metapopulation models to individually vary these population parameters. I examined how these factors affect the ability of a newly evolved pathogen to invade and persist in a population in the presence of strong competition from an endemic pathogen strain. I used these simulations to test two specific hypotheses. First I tested whether population size or population density more strongly promotes the invasion of a new pathogen. Secondly I tested whether the invasion of a new pathogen is more strongly promoted by colony size or the number of colonies.

4.3 Methods

4.3.1 Metapopulation model

I used a two-pathogen, metapopulation SIR model to compare the roles of demographic parameters on pathogen species richness. Specifically I let two identical pathogens — an endemic pathogen and an invading pathogen — compete and used persistence or not of the second pathogen as my response variable. I tested whether population size is more important than population density. I then tested whether colony size or the number of colonies is the more important component of population size. The multipathogen SIR model is identical to that in Chapter 3 and is implemented in R (R Development Core Team 2010).

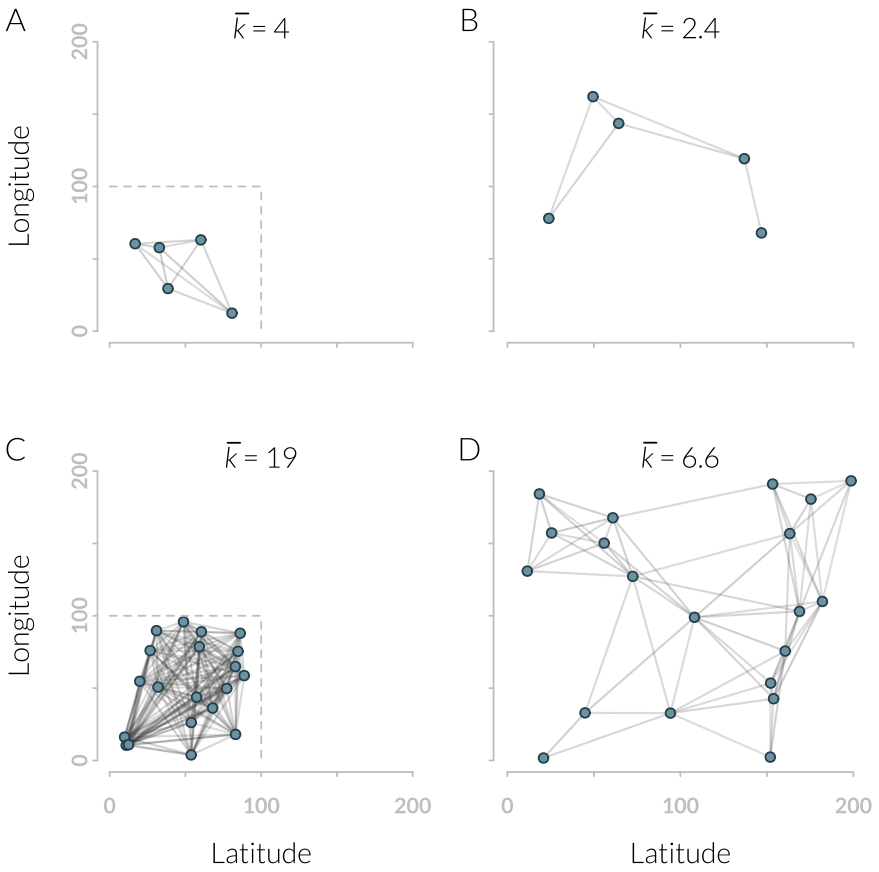


Figure 4.1 Examples of the metapopulation networks used. They include the smallest number of colonies (five, A and B) and the default (20, C and D). They also include the default area ($10^4\,km^2$, grey dashed lines, A and C) and the largest area ($4 \times 10^4\,km^2$, full plot, B and C), though all networks are plotted on the same spatial scale. Colonies are connected if they are within 100km. As area increases, the number of connections each subpopulation has decreases as seen by the changes in mean degree, $\bar{k}$.

In each simulation the population is seeded with 20 individuals infected with pathogen 1 in each colony. Pathogen 1 is then allowed to spread and reach equilibrium. After 7×10^5 events, 5 individuals infected with pathogen 2 are added to one randomly selected colony. After another 3×10^5 events the invasion of pathogen 2 is considered successful if any individuals with pathogen 2 still remain.

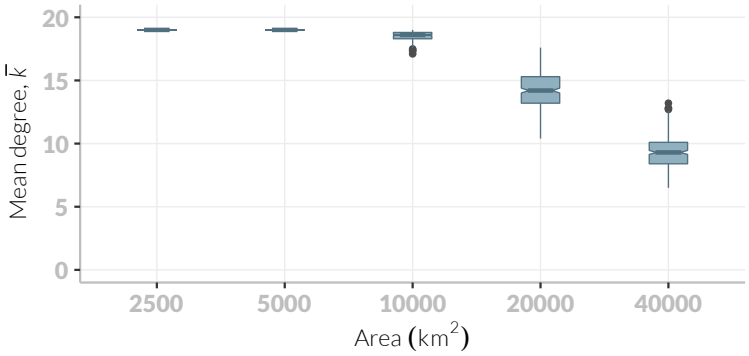


Figure 4.2 Change in average metapopulation network degree ($\bar{k}$) with increasing area. Bars show the median, boxes show the interquartile range, vertical lines show the range and grey dots indicate outlier values. Notches indicate the 95% confidence interval of the mean. All simulations had 20 colonies, meaning 19 is the maximum value of $\bar{k}$.

4.3.2 Independent variables

Three independent variables were varied: colony size, number of colonies and area. From these parameters, population size and population density can be calculated. The default values of these parameters was a population size of 8000 individuals split into 20 colonies of 400 individuals. The default area of the simulations was 10^4km^2 (space is given in square kilometres for simplicity even though they are in fact arbitrary units).

Three sets of simulations were run. First, colony size was varied using values 100, 200, 400, 800 and 1600. The number of colonies was kept constant and so population size varied directly proportionally with colony size. Area was scaled to keep population density constant. Secondly, number of colonies (and therefore population size) was varied and again area was varied to keep density constant. 5, 10, 20, 40 and 80 colonies were used. Finally, colony size and number of colonies were kept constant (therefore keeping population size constant) and area was varied alone to alter population density. The values of area used were 4×10^4, 2×10^4, 10^4, 5×10^3 and 2.5×10^3km^2 which gave density values of 0.2, 0.4, 0.8, 1.6 and 3.2 animals$\cdot$km^{-2}.

The affects of area occur through changing the metapopulation network. The metapopulation structure was created for each simulation by randomly placing colonies in space (Figure 4.1). The spatial scale of the simulations vary between 2.5×10^3 and 4×10^4 km^2. This corresponds to square areas with sides

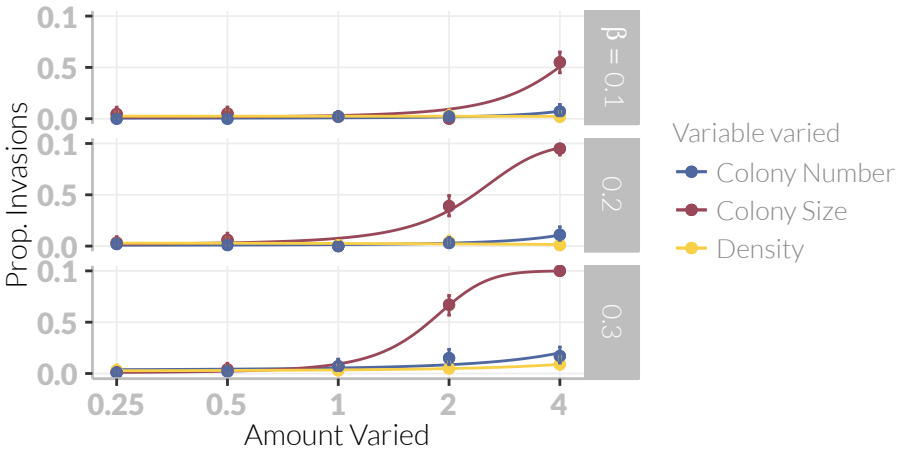


Figure 4.3 Comparison of the effect of colony size, colony number and area on probability of invasion. Default values are: colony number = 20, colony size = 400 and density = 0.8 animals·km^{-2}. The x-axis shows the relative change in each of these values ($\times 0.25, 0.5, 1, 2$ and 4). For colony size and number, area is altered so that density remains constant. For density, population size is constant at 8,000 and area is altered. Each point is the mean of 100 simulations and bars are 95% confidence intervals. Curves are simple logistic regression fits for each independent variable. Relationships are shown separately for each transmission value.

of 50 to 200km. Dispersal can only occur between two colonies if they are within 100 kilometres of each other i.e. they are then counted as connected nodes in the metapopulation network. The number of connections each colony has is called its degree, k. How well connected the metapopulation network is overall is measured by the mean degree, $\bar{k}$. This random placement does not guarantee that the population is connected (i.e. made up of a single connected component) but as the endemic pathogen is seeded in all colonies, the invading pathogen cannot be seeded into a fully susceptible colony. This was considered more realistic than repeatedly resampling the population until a fully connected population occurred. The threshold of 100 kilometres was arbitrary but I aimed to maximise the range of $\bar{k}$ (Figure 4.2) while not having many simulations with networks that were not fully connected. Given this setup, simulations with low densities had relatively unconnected metapopulation networks while high density populations had fully connected networks.

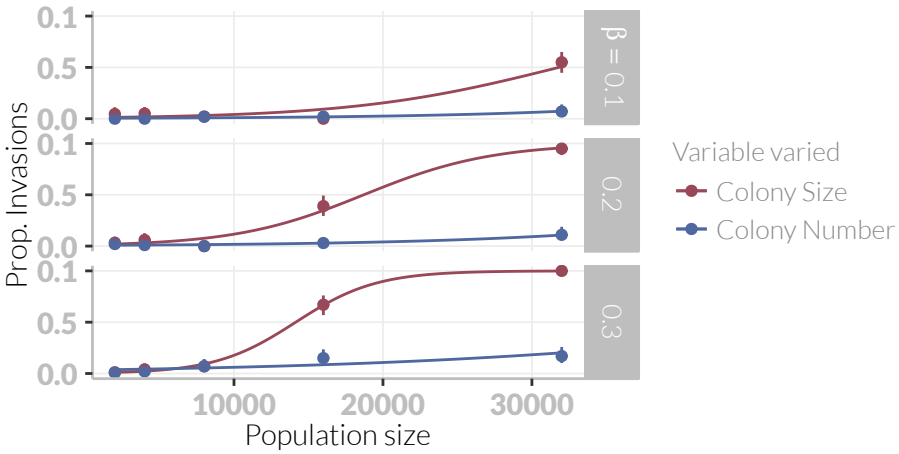


Figure 4.4 Comparison of the effect of population size on probability of invasion when population size is altered by changing colony size or colony number. Relationship is shown separately for each transmission value. It can be seen that changes in colony size give a much greater increase in invasion probability than changes in colony number. Note that this is the same data as Figure 4.3 but with the x-axis scaled by population size, not parameter change.

4.3.3 Other Parameters

The fixed parameters used are chosen to roughly reflect realistic wild bat populations. The death rate Λ is set as 0.05 per year giving a generation time of 20 years. The birth rate is set to be equal to μ so that the population size is stable. The recovery rate γ is set to 1 giving a average infection duration of 1 years. This is therefore a long lasting infection but not a chronic infection. It is difficult to estimate the duration of infections in wild bats but it seems that these infections may often be long lasting (Peel et al. 2012, Plowright et al. 2015). However, much shorter infectious periods have also been identified (Amengual et al. 2007). These shorter lived infections are not studied further here.

Cross-immunity is set to 0.1 so that coinfection is 90% less likely than an initial infection. This is an arbitrary value that is based on the fact that the rationale of the model is that the invading species is a newly speciated strain of the endemic species. Furthermore, the model assumes complete cross-immunity after recovery from infection. Therefore cross-immunity to coinfection is also likely to be strong. Three values, 0.1, 0.2 and 0.3, of the transmission rate β are used. All simulations are run under all three transmission rates.

4.3.4 Statistical comparisons

I tested two hypotheses. Firstly I tested the hypothesis that an increase in population size creates a stronger increase in invasion probability (of the second pathogen) than an equal increase in population density. Secondly, I tested the hypothesis that an increase in colony size creates a stronger increase in invasion probability than a proportionally equal increase in number of colonies. To statistically test these hypotheses I combined the results from different simulations and fitted multiple logistic regressions, centering and scaling the independent variables. Specifically, I fitted the model

$$\text{Invasion} = b_1 d + b_2 n + b_3 m + c + \epsilon \tag{4.1}$$

where d, n and m are density, colony size and number of colonies respectively and b_i are the regression coefficients. c is a fitted intercept and ϵ is a binomially distributed error term. To test the first hypothesis I compared the size (and 95% confidence intervals) of b_1 to b_2 and b_3. To test the second hypothesis I compared b_2 to b_3.

4.4 Results

At the default parameter settings, the probability of invasion and establishment of the second pathogen, $P(I)$, was rare ($\beta = 0.1$, $P(I) = 0.02$; $\beta = 0.2$, $P(I) = 3.33 \times 10^{-3}$; $\beta = 0.3$, $P(I) = 0.06$). Although there is no clear, directional relationship, these proportions are significantly different (χ^2 test: $\chi^2 = 17.21$, df $= 2$, $p = 1.83 \times 10^{-4}$).

In 37 simulations, both of the pathogens went extinct. This did not depend on transmission rate (χ^2 test: $\chi^2 = 1.51$, df $= 2$, $p = 0.47$). However they were all either in simulations with the smallest colony size (colony size = 100, 29 simulations) or with the fewest colonies (5 colonies, 8 simulations). Results from these simulations were removed before further analyses.

4.4.1 Population density or size

To test whether population density or size has a stronger affect on invasion probability I compared the regression coefficients of the multiple regressions fitted to simulation results (Figure 4.3). Increasing population size, either by increasing colony size or number of colonies, increased the probability of invasion (Table 4.1).

Table 4.1 Regression results comparing effects of colony size, colony number and density. Coefficients are from multiple logistic regressions with invasion as the dependent variable and all independent variables being scaled and centred. Colony size and colony number were varied while keeping density equal while density was varied by changing area while keeping population size equal. p is for the test against the null hypothesis that $b = 0$.

β	*Variable*	*Estimate* (b)	(95% *CI*)	p
0.1	Intercept	−3.52	(−3.87, −3.2)	$< 10^{-5}$
	Colony Size	1.07	(0.75, 1.49)	$< 10^{-5}$
	Colony Number	0.35	(−0.02, 0.79)	0.08
	Density	0.01	(−0.66, 0.52)	0.97
0.2	Intercept	−2.84	(−3.12, −2.58)	$< 10^{-5}$
	Colony Size	2.11	(1.71, 2.6)	$< 10^{-5}$
	Colony Number	0.51	(0.16, 0.95)	0.009
	Density	−0.31	(−0.96, 0.19)	0.29
0.3	Intercept	−2.11	(−2.34, −1.9)	$< 10^{-5}$
	Colony Size	2.74	(2.35, 3.16)	$< 10^{-5}$
	Colony Number	0.25	(0.04, 0.48)	0.02
	Density	0.27	(−0.06, 0.57)	0.09

The relationship between colony size and invasion is strong and significant at all transmission rates, while the relationship between colony number and invasion is weaker and more marginally significant. In contrast, varying population density does not alter invasion probability. Therefore the simulations support the hypothesis that population size affects invasion more strongly than population density.

4.4.2 Colony size or number of groups

To test whether colony size or the number of colonies is the more important component of population size, I compared the regression coefficients, b_2 and b_3, of the multiple regressions fitted to simulation results (Figure 4.4). Increasing colony size or the number of colonies increases the probability of invasion but this affect is much stronger and more statistically significant, for colony size (Table 4.1). Therefore the simulations support the hypothesis that colony size is the more important component of population size.

4.5 Discussion

Overall, my results suggest that population size promotes pathogen richness significantly more than population density in the context of metapopulations or group living. Furthermore, the component of population size that is important is group size.

These results lead to a number of other conclusions. All else being equal, increasing range size (with density remaining constant) will not increase pathogen richness significantly unless the increased range size promotes larger groups. Furthermore, social species that live in large groups are likely to harbour more pathogen species, even if sociality promotes reduced interactions between groups due to territory defence or simply because of larger distances between groups due to groups needing larger home ranges than solitary individuals.

For related, strongly competing strains, the factor that most strongly allows new pathogens to invade is the number of susceptible individuals in the local group. As long as there are enough susceptible individuals that the new pathogen species persists through the stochastic, early stages of the epidemic, the new pathogen will persist. As dispersal is a very slow process compared to infection, the global pool of susceptibles is not important. This is why increasing the number of colonies does not increase pathogen invasion as quickly as the size of a colony does. Similarly, the density — at the global scale — of the species has little affect. In these simulations, increasing density without increasing population size implies a reduction in range size, which simply increases the number of colonies which are connected to the colony experiencing the invading pathogen. This increases the pool of susceptibles that are within one dispersal of the invading pathogen. However, again, this affect is very weak compared to the strong changes in local disease dynamics caused by increasing colony size.

4.5.1 Global change

It is clear that many species are suffering strong population changes due to climate change (Thomas et al. 2004). However these changes might affect range size (Thomas et al. 2004), population size (Craigie et al. 2010), population connectivity (Fontúrbel et al. 2014, Rivera-Ortíz et al. 2015, Wasserman et al. 2013) or group size (Atwood 2006, Lehmann et al. 2010, Manor & Saltz 2003, Zunino et al. 2007) to different extents. My results suggest that pathogen communities will respond differently depending on which factors are affected, although it should be

noted that the mechanism here — invasion of a new pathogen — is possibly more relevant over longer, multi-generation time scales than decade-long time scales. In short, species suffering reductions in groups size (Atwood 2006, Lehmann et al. 2010, Manor & Saltz 2003, Zunino et al. 2007) are predicted to experience decreases in pathogen richness in the long term. Species that are experiencing increases in group size (Lehmann et al. 2010) would be expected gain new pathogen species more quickly. In contrast, species suffering range contractions (Thomas et al. 2004) and decreases in population size (Craigie et al. 2010) are expected to experience smaller changes in pathogen richness.

4.5.2 Comparative studies

Many comparative studies measure some aspect of a species population size or structure, yet it is rarely discussed how these relate. Instead most studies use the data that are available, without considering *a priori* how it may depend on other factors (though statistical correlations between independent variables is usually considered and dealt with using PCA or by removing collinear variables). Population density is often measured (Arneberg 2002, Lindenfors et al. 2007, Morand & Poulin 1998, Nunn et al. 2003) yet density is directly associated with population size. This study suggests that it is in fact population size that is important (in the context of social species as studied here). Therefore, the density measures in these comparative studies are more likely to be proxies for population size than the true causal factor. Similarly, this study suggests that host range size does not promote pathogen richness by the mechanism studied here, yet a number of studies have found evidence of this relationship (Kamiya et al. 2014, Nunn et al. 2003). This suggests that either the relationship found in comparative studies is in fact due to a correlation with another factor, or that mechanisms other than rate of invasion of new pathogens are important. Range size has been suggested to affect pathogen richness by a number of mechanisms such as increasing the diversity of sympatric species and these other mechanisms should be specifically tested.

The studies that have tested specifically the affect of group size have in fact found both positive (Vitone et al. 2004) and negative associations (Gay et al. 2014) or no relationship (Ezenwa et al. 2006). Meta-analyses suggest that the relationship between social group size and pathogen richness is weak (Rifkin et al. 2012). This suggests that the mechanism studied here — invasion of recently evolved pathogens — is not the major cause of pathogen richness in wild populations.

4.5.3 Assumptions and limitation

Being based on the same model as used earlier, the work presented here relies on many of the same assumptions (see Section 3.5). Furthermore, as a comparison is being made between the effects of area and population size, the exact specifications of how the metapopulation is affected by area is important. I have conducted this study at one rate of dispersal, 0.01 dispersals per individual per year. In practice this relates to only 20% of individuals dispersing in their lifetime. This low rate of dispersal is expected to exaggerate the effect of area; at high rates of dispersal the population is essentially well-mixed, despite the metapopulation.

Also, I have assumed that dispersal only occurs between colonies a certain distance apart. Based on *a priori* considerations such as the time and energy required to disperse long distances this is a reasonable assumption. The exact threshold was chosen to attempt to maximise the range of $\bar{k}$ studied (Figure 4.2). However, a similar assumption could be made in other ways. Instead of a threshold distance, individuals could be expected to disperse in a random direction and stop at the first colony they encounter; this could create some long distance links in the network and increase network connectivity, potentially reducing the effects of area. Alternatively, the metapopulation could have been modelled as a weighted network with dispersal occurring at a higher rate to nearby colonies. Depending on the parameterisation of this distance-dispersal relationship this could serve to increase the affect of area — by exaggerating dispersal to very nearby colonies — or decrease the affect of area by allowing rare, but significant, global dispersal creating a small-world network structure. Ultimately, the modelling choices could increase or decrease the affects of area relative to colony size and the number of colonies but I have aimed to make the effect of area as strong as possible.

I have used the simple relationships between demographic factors — density = population size / area for example — to illustrate that these are tightly linked. In order to isolate the effects of these factors I have assumed these simple relationships hold; to examine density without altering population size I have fixed population size and manipulated area. However, in reality, these are likely to covary both within species across time and also between species. Therefore, while these quantities are certainly linked, they cannot be assumed to have simple linear relationships and should not be used as proxies of each other without further examination. For example, rates and distances of dispersal — which affect the influence of space — may be related to local density (Marjamäki et al. 2013). Sim-

ilarly it is unlikely that a species whose range size decreases will not experience a decrease in total population size as well; the range contraction is likely to occur over generations rather than a simple squeezing of the existing individuals into a smaller area.

4.5.4 Conclusions

Overall I have shown that while a number of demographic factors are intrinsically linked, they have different effects on the rate at which new pathogens will invade. I found that population size, not density, has the stronger impact on the ability of a pathogen to invade. Furthermore, species with large groups are likely to harbour more pathogens than species with many, smaller groups. Due to the correlations between these factors, they are particularly hard to study within a comparative framework; this highlights the utility of mechanistic models.

5 A generalised random encounter model for estimating animal density with remote sensor data

This work was conducted in collaboration with Elizabeth Moorcroft, Robin Freeman, Marcus Rowcliffe and Kate Jones and is now published in Methods in Ecology and Evolution (Lucas et al. 2015). The text here is almost completely reproduced from Lucas et al. (2015). I formulated and analysed the analytical model. Elizabeth Moorcroft wrote to code for and carried out the simulations. I led the writing of the manuscript with contributions fro the other coauthors.

5.1 Abstract

Wildlife monitoring technology is advancing rapidly and the use of remote sensors such as camera traps and acoustic detectors is becoming common in both the terrestrial and marine environments. Current methods to estimate population size or density require individual recognition of animals or knowing the distance of the animal from the sensor, which is often difficult. A method without these requirements, the random encounter model (REM), has been successfully applied to estimate animal densities from count data generated from camera traps. However, count data from acoustic detectors do not fit the assumptions of the REM due to the directionality of animal signals.

We developed a generalised REM (gREM), to estimate absolute animal density from count data from both camera traps and acoustic detectors. We derived the gREM for different combinations of sensor detection widths and animal signal widths (a measure of directionality). We tested the accuracy and precision of this model using simulations of different combinations of sensor detection widths and animal signal widths, number of captures, and models of animal movement.

We find that the gREM produces accurate estimates of absolute animal density for all combinations of sensor detection widths and animal signal widths. However, larger sensor detection and animal signal widths were found to be more precise. While the model is accurate for all capture efforts tested, the precision of the estimate increases with the number of captures. We found no effect of different animal movement models on the accuracy and precision of the gREM.

We conclude that the gREM provides an effective method to estimate absolute animal densities from remote sensor count data over a range of sensor and animal signal widths. The gREM is applicable for count data obtained in both marine and terrestrial environments, visually or acoustically (e.g., big cats, sharks, birds, echolocating bats and cetaceans). As sensors such as camera traps and acoustic detectors become more ubiquitous, the gREM will be increasingly useful for monitoring unmarked animal populations across broad spatial, temporal and taxonomic scales.

5.2 Introduction

The density of animal populations is one of the fundamental measures in ecology and conservation and has important implications for a range of issues, such as sensitivity to stochastic fluctuations (Wright & Hubbell 1983) and extinction risk (Purvis et al. 2000). Monitoring animal population changes in response to anthropogenic pressure is becoming increasingly important as humans rapidly modify habitats and change climates (Everatt et al. 2014). Sensor technology, such as camera traps (Karanth 1995, Rowcliffe & Carbone 2008) and acoustic detectors (Acevedo & Villanueva-Rivera 2006, Walters et al. 2012) are widely used to monitor changes in animal populations as they are efficient, relativity cheap and non-invasive, allowing for surveys over large areas and long periods (Kessel et al. 2014, Rowcliffe & Carbone 2008, Walters et al. 2013). However, converting sampled count data into estimates of density is problematic as detectability of animals needs to be accounted for (Anderson 2001).

Existing methods for estimating animal density often require additional information that is often unavailable. For example, capture-mark-recapture methods (Borchers et al. 2014, Karanth 1995) require recognition of individuals, and distance methods (Harris et al. 2013) require estimates of how far away individuals are from the sensor (Barlow & Taylor 2005, Marques et al. 2011). When individuals cannot be told apart, an extension of occupancy modelling can be used to estimate absolute population size (Royle & Nichols 2003). However, as the model is originally formulated to estimate occupancy, count information is simplified to presence–absence data. Assumptions about the distribution of individuals (e.g., a Poisson distribution) must also be made (Royle & Nichols 2003) which may be a poor assumption for nonrandomly distributed species. Furthermore repeat, independent surveys must be performed and the definition of a site can be difficult, especially for wide-ranging species (MacKenzie & Royle 2005).

The REM method has been successfully applied to estimate animal densities from camera trap surveys (Zero et al. 2013). However, extending the REM method to other types of sensors (e.g., acoustic detectors) is more problematic, because the original derivation assumes a relatively narrow sensor width (up to $\pi/2$ radians) and that the animal is equally detectable irrespective of its heading (Rowcliffe et al. 2008).

Whilst these restrictions are not problematic for most camera trap makes (e.g., Reconyx, Cuddeback), the REM cannot be used to estimate densities from camera traps with a wider sensor width (e.g., canopy monitoring with fish eye

lenses, (Brusa & Bunker 2014)). Additionally, the REM method is not useful in estimating densities from acoustic survey data as acoustic detector angles are often wider than $\pi/2$ radians. Acoustic detectors are designed for a range of diverse tasks and environments (Kessel et al. 2014), which naturally leads to a wide range of sensor detection widths and detection distances. In addition to this, calls emitted by many animals are directional (Blumstein et al. 2011), breaking the assumption of the REM method.

There has been a sharp rise in interest around passive acoustic detectors in recent years, with a 10 fold increase in publications in the decade between 2000 and 2010 (Kessel et al. 2014). Acoustic monitoring is being developed to study many aspects of ecology, including the interactions of animals and their environments (Blumstein et al. 2011, Rogers et al. 2013), the presence and relative abundances of species (Marcoux et al. 2011), biodiversity of an area (Depraetere et al. 2012), and monitoring population trends (Walters et al. 2013).

Acoustic data suffers from many of the problems associated with data from camera trap surveys in that individuals are often unmarked, making capture-mark-recapture methods more difficult to use (Marques et al. 2013). In some cases the distance between the animal and the sensor is known, for example, when an array of sensors is deployed and the position of the animal is estimated by triangulation (Lewis et al. 2007). In these situations distance-sampling methods can be applied (Buckland et al. 2008). However, in many cases distance estimation is not possible, for example, when single sensors are deployed, a situation typical in the majority of terrestrial acoustic surveys (Buckland et al. 2008). In these cases, only relative measures of local abundance can be calculated, and not absolute densities. This means that comparison of populations between species and sites is problematic without assuming equal detectability (Schmidt 2003, Walters et al. 2013). Equal detectability is unlikely because of differences in environmental conditions, sensor type, habitat, and species biology.

In this study, we create a generalised REM (gREM) as an extension to the camera trap model of (Rowcliffe et al. 2008), to estimate absolute density from count data from acoustic detectors, or camera traps, where the sensor width can vary from 0 to 2π radians, and the signal given from the animal can be directional. We assessed the accuracy and precision of the gREM within a simulated environment, by varying the sensor detection widths, animal signal widths, number of captures and models of animal movement. We use the simulation results to recommend best survey practice for estimating animal densities from remote

sensors.

5.3 Methods

5.3.1 Analytical Model

The REM presented by (Rowcliffe et al. 2008) adapts the gas model to count data collected from camera trap surveys. The REM is derived assuming a stationary sensor with a detection width less than $\pi/2$ radians. However, in order to apply this approach more generally, and in particular to stationary acoustic detectors, we need both to relax the constraint on sensor detection width, and allow for animals with directional signals. Consequently, we derive the gREM for any detection width, θ, between 0 and 2π with a detection distance r giving a circular sector within which animals can be captured (the detection zone) (Figure 5.1). Additionally, we model the animal as having an associated signal width α between 0 and 2π (Figure 5.1, see Appendix S1 for a list of symbols). We start deriving the gREM with the simplest situation, the gas model where $\theta = 2\pi$ and $\alpha = 2\pi$.

Gas Model Following (Yapp 1956), we derive the gas model where sensors can capture animals in any direction and animal signals are detectable from any direction ($\theta = 2\pi$ and $\alpha = 2\pi$). We assume that animals are in a homogeneous environment, and move in straight lines of random direction with velocity v. We allow that our stationary sensor can capture animals at a detection distance r and that if an animal moves within this detection zone they are captured with a probability of one; while outside this zone, animals are never captured.

In order to derive animal density, we need to consider relative velocity from the reference frame of the animals. Conceptually, this requires us to imagine that all animals are stationary and randomly distributed in space, while the sensor moves with velocity v. If we calculate the area covered by the sensor during the survey period, we can estimate the number of animals the sensor should capture. As a circle moving across a plane, the area covered by the sensor per unit time is $2rv$. The expected number of captures, z, for a survey period of t, with an animal density of D is $z = 2rvtD$. To estimate the density we rearrange to get $D = z/2rvt$. Note that as z is the number of encounters, not individuals, the possibility of repeated detections of the same individual is accounted for (Hutchinson & Waser 2007).

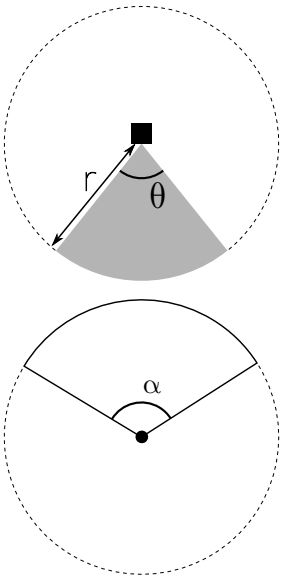


Figure 5.1 Representation of sensor detection width and animal signal width. The filled square and circle represent a sensor and an animal, respectively; θ, sensor detection width (radians); r, sensor detection distance; dark grey shaded area, sensor detection zone; α, animal signal width (radians). Dashed lines around the filled square and circle represents the maximum extent of θ and α, respectively.

gREM derivations for different detection and signal widths Different combinations of θ and α would be expected to occur (e.g., sensors have different detection widths and animals have different signal widths). For different combinations θ and α, the area covered per unit time is no longer given by $2rv$. Instead of the size of the sensor detection zone having a diameter of $2r$, the size changes with the approach angle between the sensor and the animal. The width of the area within which an animal can be detected is called the profile, p. The size of p depends on the signal width, detector width and the angle that the animal approaches the sensor. The size of the profile (averaged across all approach angles) is defined as the average profile $\bar{p}$. However, different combinations of θ and α need different equations to calculate $\bar{p}$.

We have identified the parameter space for the combinations of θ and α for which the derivation of the equations are the same (defined as sub-models in the gREM) (Figure 5.2). For example, the gas model becomes the simplest gREM sub-model (upper right in Figure 5.2) and the REM from (Rowcliffe et al. 2008) is another gREM sub-model where $\theta < \pi/2$ and $\alpha = 2\pi$. We derive one gREM

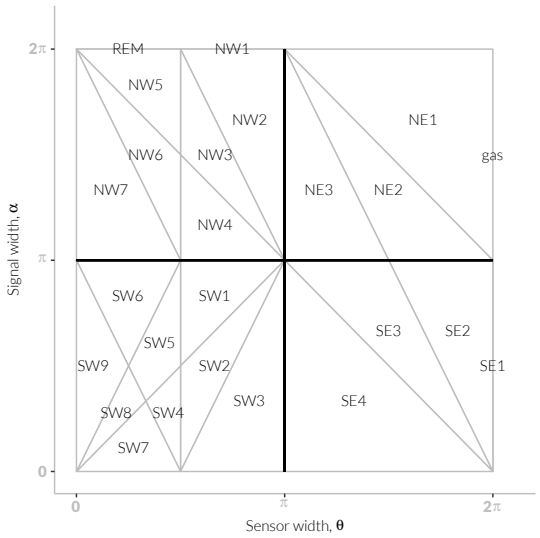


Figure 5.2 Locations where derivation of the average profile $\bar{p}$ is the same for different combinations of sensor detection and animal signal widths. Symbols within each polygon refer to each gREM submodel named after their compass point, except for Gas and REM which highlight the position of these previously derived models within the gREM. Symbols on the edge of the plot are for submodels where $\alpha, \theta = 2\pi$

sub-model SE2 as an example below, where $2\pi - \alpha/2 < \theta < 2\pi,\ 0 < \alpha < \pi$ (see Appendix S2 for derivations of all gREM sub-models). Any estimate of density would require prior knowledge of animal velocity, v and animal signal width, α taken from other sources, for example, existing literature (Brinkløv et al. 2011, Carbone et al. 2005). Sensor width, θ, and detection distance, r would also need to be measured or obtained from manufacturer specifications (Adams et al. 2012, Holderied & Von Helversen 2003).

Example derivation of SE2 In order to calculate $\bar{p}$, we have to integrate over the focal angle, x_1 (Figure 5.3a). This is the angle taken from the centre line of the sensor. Other focal angles are possible (x_2, x_3, x_4) and are used in other gREM submodels (see Appendix S2). As the size of the profile depends on the approach angle, we present the derivation across all approach angles. When the sensor is directly approaching the animal $x_1 = \pi/2$.

Starting from $x_1 = \pi/2$ until $\theta/2 + \pi/2 - \alpha/2$, the size of the profile is $2r \sin \alpha/2$ (Figure 5.3b). During this first interval, the size of α limits the width of the profile.

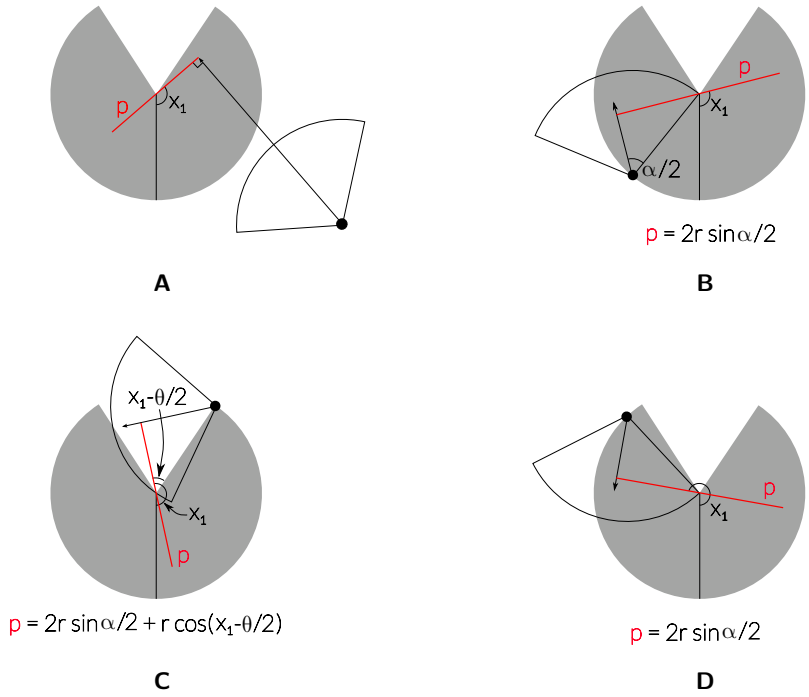


Figure 5.3 An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2, where (a) shows the location of the profile p (the line an animal must pass through in order to be captured) in red and the focal angle, x_1, for an animal (filled circle), its signal (unfilled sector), and direction of movement (shown as an arrow). The detection zone of the sensor is shown as a filled grey sector with a detection distance of r. The vertical black line within the circle shows the direction the sensor is facing. The derivation of p changes as the animal approaches the sensor from different directions (shown in b-d), where (b) is the derivation of p when x_1 is in the interval $[\frac{\pi}{2}, \frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}]$, (c) p when x_1 is in the interval $[\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}, \frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}]$ and (d) p when x_1 is in the interval $[\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}, \frac{3\pi}{2}]$, where θ, sensor detection width; α, animal signal width. The resultant equation for p is shown beneath b-d. The average profile $\bar{p}$ is the size of the profile averaged across all approach angles.

When the animal reaches $x_1 = \theta/2 + \pi/2 - \alpha/2$ (Figure 5.3c), the size of the profile is $r\sin(\alpha/2) + r\cos(x_1 - \theta/2)$ and the size of θ and α both limit the width of the profile (Figure 5.3c). Finally, at $x_1 = 5\pi/2 - \theta/2 - \alpha/2$ until $x_1 = 3\pi/2$, the width of the profile is again $2r\sin\alpha/2$ (Figure 5.3d) and the size of α again limits the width of the profile.

The profile width p for π radians of rotation (from directly towards the sensor to directly behind the sensor) is completely characterised by the three intervals (Figure 5.3b–d). Average profile width $\bar{p}$ is calculated by integrating these profiles over their appropriate intervals of x_1 and dividing by π which gives

$$\bar{p} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} 2r\sin\frac{\alpha}{2}\,\mathrm{d}x_1 + \int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\frac{\alpha}{2} + r\cos\left(x_1-\frac{\theta}{2}\right)\mathrm{d}x_1 \right.$$
$$\left. + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{3\pi}{2}} 2r\sin\frac{\alpha}{2}\,\mathrm{d}x_1 \right) \tag{5.1}$$
$$= \frac{r}{\pi}\left(\theta\sin\frac{\alpha}{2} - \cos\frac{\alpha}{2} + \cos\left(\frac{\alpha}{2}+\theta\right)\right) \tag{5.2}$$

We then use this expression to calculate density

$$D = z/vt\bar{p}. \tag{5.3}$$

Rather than having one equation that describes $\bar{p}$ globally, the gREM must be split into submodels due to discontinuous changes in p as α and β change. These discontinuities can occur for a number of reasons such as a profile switching between being limited by α and θ, the difference between very small profiles and profiles of size zero, and the fact that the width of a sector stops increasing once the central angle reaches π radians (i.e., a semi-circle is just as wide as a full circle). As an example, if α is small, there is an interval between Figure 5.3c and 5.3d where the 'blind spot' would prevent animals being detected giving $p = 0$. This would require an extra integral in our equation, as simply putting our small value of α into 5.1 would not give us this integral of $p = 0$.

gREM submodel specifications were done by hand, and the integration was done using SymPy (SymPy Development Team 2014) in Python (Appendix S3). The gREM submodels were checked by confirming that: (1) submodels adjacent in parameter space were equal at the boundary between them; (2) submodels that border $\alpha = 0$ had $p = 0$ when $\alpha = 0$; (3) average profile widths $\bar{p}$ were between 0 and $2r$ and; (4) each integral, divided by the range of angles that it was integrated over, was between 0 and $2r$. The scripts for these tests are included in Appendix S3 and the R (R Development Core Team 2010) implementation of the gREM is given in Appendix S4.

5.3.2 Simulation Model

We tested the accuracy and precision of the gREM by developing a spatially explicit simulation of the interaction of sensors and animals using different combinations of sensor detection widths, animal signal widths, number of captures, and models of animal movement. One hundred simulations were run where each consisted of a 7.5 km by 7.5 km square with periodic boundaries. A stationary sensor of radius r, 10 m, was set up in the exact centre of each simulated study area, covering seven sensor detection widths θ, between 0 and 2π ($2/9\pi$, $4/9\pi$, $6/9\pi$, $8/9\pi$, $10/9\pi$, $14/9\pi$, and 2π). Each sensor was set to record continuously and to capture animal signals instantaneously from emission. Each simulation was populated with a density of 70 animals km^{-2}, calculated from the equation in (Damuth 1981) as the expected density of mammals weighing 1 g. This density therefore represents a reasonable estimate of density of individuals, given that the smallest mammal is around 2 g (Jones et al. 2009). A total of 3937 individuals per simulation were created which were placed randomly at the start of the simulation. 11 signal widths α between 0 and π were used ($1/11\pi$, $2/11\pi$, $3/11\pi$, $4/11\pi$, $5/11\pi$, $6/11\pi$, $7/11\pi$, $8/11\pi$, $9/11\pi$, $10/11\pi$, π).

Each simulation lasted for N steps (14400) of duration T (15 minutes) giving a total duration of 150 days. The individuals moved within each step with a distance d, with an average speed, v. The distance, d, was sampled from a normal distribution with mean distance, $\mu_d = vT$, and standard deviation, $\sigma_d = vT/10$, where the standard deviation was chosen to scale with the average distance travelled. An average speed, $v = 40\,\mathrm{km\,day}^{-1}$, was chosen based on the largest day range of terrestrial animals (Carbone et al. 2005), and represents the upper limit of realistic speeds. At the end of each step, individuals were allowed to either remain stationary for a time step (with a given probability, S), or change direction where the change in direction has a uniform distribution in the interval $[-A, A]$. This resulted in seven different movement models where: (1) simple movement, where S and $A = 0$; (2) stop-start movement, where (i) $S = 0.25$, $A = 0$, (ii) $S = 0.5$, $A = 0$, (iii) $S = 0.75$, $A = 0$; (3) correlated random walk movement, where (i) $S = 0$, $A = \pi/3$, (ii) $S = 0$, $A = 2\pi/3$, iii) $S = 0$, $A = \pi$. Individuals were counted as they moved into the detection zone of the sensor per simulation.

We calculated the estimated animal density from the gREM by summing the number of captures per simulation and inputting these values into the correct gREM submodel. The accuracy of the gREM was determined by comparing the true simulation density with the estimated density. Precision of the gREM

was determined by the standard deviation of estimated densities. We used this method to compare the accuracy and precision of all the gREM submodels. As these submodels are derived for different combinations of α and θ, the accuracy and precision of the submodels was used to determine the impact of different values of α and θ.

The influence of the number of captures and animal movement models on accuracy and precision was investigated using four different gREM submodels representative of the range α and θ values (submodels NW1, SW1, NE1, and SE3, Figure 5.2). From a random starting point we ran the simulation until a range of different capture numbers were recorded (from 10 to 100 captures), recorded the length of time this took, and estimated the animal density for each of the four sub-models. These estimated densities were compared to the true density to assess the impact on the accuracy and precision of the gREM. We calculated the coefficient of variation in order to compare the precision of the density estimates from simulations with different expected numbers of captures. The gREM also assumes that individuals move continuously with straight-line movement (simple movement model) and we therefore assessed the impact of breaking the gREM assumptions. We used the four submodels to compare the accuracy and precision of a simple movement model, stop-start movement models (using different average amounts of time spent stationary), and random walk movement models. Finally, as the parameters (α, β, r and v) are likely to be measured with error, we compared true simulation densities to densities estimated with parameters with errors of 0%, $\pm$5% and $\pm$10%, for all gREM submodels.

5.4 Results

5.4.1 Analytical model

The equation for $\bar{p}$ has been newly derived for each submodel in the gREM, except for the gas model and REM which have been calculated previously. However, many models, although derived separately, have the same expression for $\bar{p}$. Figure 5.4 shows the expression for $\bar{p}$ in each case. The general equation for density, 5.3, is used with the correct value of $\bar{p}$ substituted. Although more thorough checks are performed in Appendix S3, it can be seen that all adjacent expressions in Figure 5.4 are equal when expressions for the boundaries between them are substituted in.

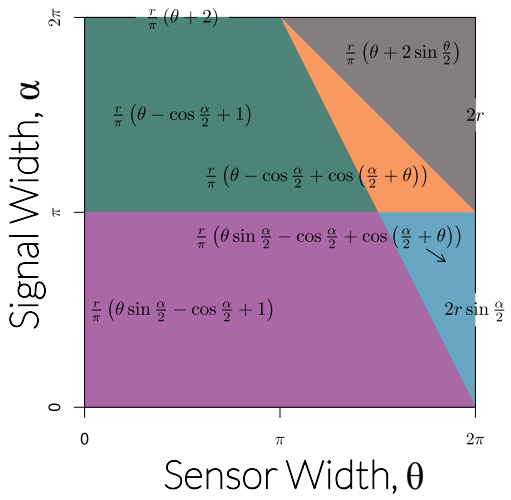


Figure 5.4 Expressions for the average profile width, $\bar{p}$, given a range of sensor and signal widths. Despite independent derivation within each block, many models result in the same expression. These are collected together and presented as one block of colour. Expressions on the edge of the plot are for submodels with $\alpha, \theta = 2\pi$.

5.4.2 Simulation model

gREM submodels All gREM submodels showed a high accuracy, i.e., the median difference between the estimated and true values was less than 2% across all models (Figure 5.5). However, the precision of the submodels do vary, where the gas model is the most precise and the SW7 sub model the least precise, having the smallest and the largest interquartile range, respectively (Figure 5.5). The standard deviation of the error between the estimated and true densities is strongly related to both the sensor and signal widths (Appendix S5), such that larger widths have lower standard deviations (greater precision) due to the increased capture rate of these models.

Number of captures Within the four gREM submodels tested (NW1, SW1, SE3, NE1), the accuracy was not strongly affected by the number of captures. The median difference between the estimated and true values was less than 15% across all capture rates (Figure 5.6). However, the precision was dependent on the number of captures across all four of the gREM submodels, where precision increases as number of captures increases, as would be expected for any statistical estimate (Figure 5.6). For all gREM submodels, the the coefficient of variation falls to 10%

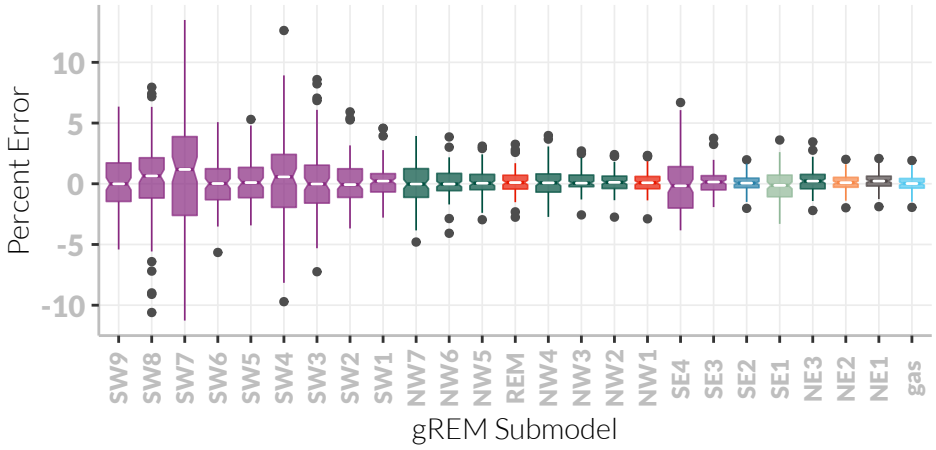


Figure 5.5 Simulation model results of the accuracy and precision for gREM submodels. The percentage error between estimated and true density for each gREM sub model is shown within each box plot, where the white line represents the median percentage error across all simulations, boxes represent the middle 50% of the data, whiskers represent variability outside the upper and lower quartiles with outliers plotted as individual points. Notches indicate 95% confidence intervals. Box colours correspond to the expressions for average profile width $\bar{p}$ given in Figure 5.4.

at 100 captures.

Movement models Within the four gREM submodels tested (NW1, SW1, SE3, NE1), neither the accuracy or precision was affected by the average amount of time spent stationary. The median difference between the estimated and true values was less than 2% for each category of stationary time (0, 0.25, 0.5 and 0.75) (Figure 5.7). Altering the maximum change in direction in each step (0, $\pi/3$, $2\pi/3$, and π) did not affect the accuracy or precision of the four gREM submodels (Figure 5.7).

Impact of parameter error The percentage error in the density estimates across all parameters and gREM submodels shows a similar response for under and over estimated parameters, suggesting the accuracy is reasonable with respect to parameter error (Appendix S6). The impact of parameter error on the precision of the density estimate varies across gREM submodels and parameters, where α shows the largest variation including the largest values. However, in all cases

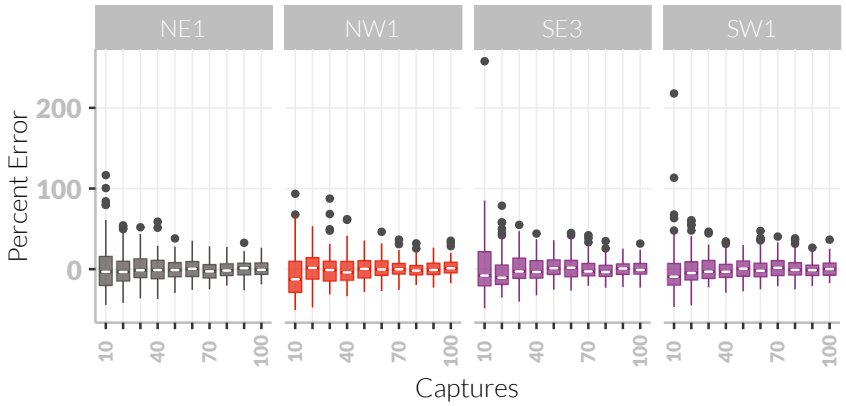


Figure 5.6 Simulation model results of the accuracy and precision of four gREM submodels (NW1, SW1, SE3 and NE1) given different numbers of captures. The percentage error between estimated and true density within each gREM sub model for capture rate is shown within each box plot, where the white line represents the median percentage error across all simulations, boxes represent the middle 50% of the data, whiskers represent variability outside the upper and lower quartiles with outliers plotted as individual points. Notches show the 95% confidence interval. Sensor and signal widths vary between submodels. The numbers beneath each plot represent the coefficient of variation. The colour of each box plot corresponds to the expressions for average profile width $\bar{p}$ given in Figure 5.4.

the percentage error in the density estimate is not more than 5% greater than the error in the parameter estimate (Appendix S6).

5.5 Discussion

5.5.1 Analytical model

We have developed the gREM such that it can be used to estimate density from acoustic sensors and camera traps. This has entailed a generalisation of the gas model and the REM in (Rowcliffe et al. 2008) to be applicable to any combination of sensor width θ and signal directionality α. We emphasise that the approach is robust to multiple detections of the same individual. We have used simulations to show, as a proof of principle, that these models are accurate and precise.

There are a number of possible extensions to the gREM which could be developed in the future. The original gas model was formulated for the case where

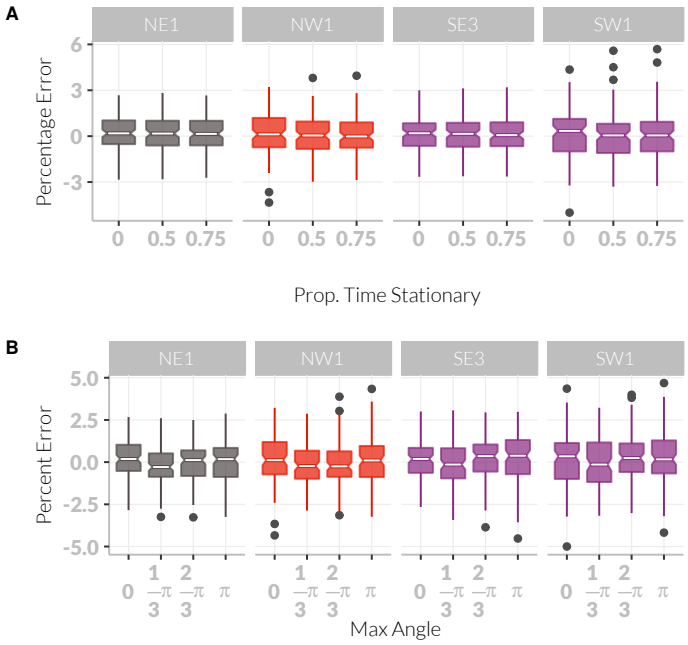


Figure 5.7 Simulation model results of the accuracy and precision of four gREM submodels (NW1, SW1, SE3 and NE1) given different movement models where (a) average amount of time spent stationary (stop-start movement) and (b) maximum change in direction at each step (correlated random walk model). The percentage error between estimated and true density within each gREM sub model for the different movement models is shown within each box plot, where the white line represents the median percentage error across all simulations, boxes represent the middle 50% of the data, whiskers represent variability outside the upper and lower quartiles with outliers plotted as individual points. Notches in boxplots show the 95% confidence for the median. The simple model is represented where time and maximum change in direction equals 0. The colour of each box plot corresponds to the expressions for average profile width $\bar{p}$ given in Figure 4.

both animals and sensor are moving (Hutchinson & Waser 2007). Indeed any of the models which have animals that are equally detectable in all directions ($\alpha = 2\pi$) can be trivially expanded by replacing animal speed v with $v + v_s$ where v_s is the speed of the sensor. However, when the animal has a directional call the extension becomes less simple. The approach would be to calculate again the mean profile width. However, for each angle of approach, one would have to average the profile width for an animal facing in any direction (i.e., not necessarily moving towards the sensor) weighted by the relative velocity of that direction. There are a number of situations where a moving detector and an-

imal could occur, e.g., an acoustic detector towed from a boat when studying porpoises (Kimura et al. 2014) or surveying echolocating bats from a moving car (Jones et al. 2011).

Interesting but unstudied problems impacting the gREM are firstly, edge effects caused by sensor trigger delays (the delay between sensing an animal and attempting to record the encounter) (Rovero et al. 2013), and secondly, sensors which repeatedly turn on an off during sampling (Jones et al. 2011). The second problem is particularly relevant to acoustic detectors which record ultrasound by time expansion. Here ultrasound is recorded for a set time period and then slowed down and played back, rendering the sensor 'deaf' periodically during sampling. Both of these problems may cause biases in the gREM, as animals can move through the detection zone without being detected. As the gREM assumes constant surveillance, the error created by switching the sensor on and off quickly will become more important if the sensor is only on for short periods of time. We recommend that the gREM is applied to constantly sampled data, and the impacts of breaking these assumptions on the gREM should be further explored.

5.5.2 Accuracy, Precision and Recommendations for Best Practice

Based on our simulations, we believe that the gREM has the potential to produce accurate estimates for many different species, using either camera traps or acoustic detectors. However, the precision of the gREM differed between submodels. For example, when the sensor and signal width were small, the precision of the model was reduced. Therefore when choosing a sensor for use in a gREM study, the sensor detection width should be maximised. If the study species has a narrow signal directionality, other aspects of the study protocol, such as length of the survey, should be used to compensate.

The precision of the gREM is greatly affected by the number of captures. The coefficient of variation falls dramatically between 10 and 60 captures and then after this continues to slowly reduce. At 100 captures the submodels reach 10% coefficient of variation, considered to be a very good level of precision and better than many previous studies (Foster & Harmsen 2012, O'Brien et al. 2003, Thomas & Marques 2012). The length of surveys in the field will need to be adjusted so that enough data can be collected to reach this precision level. Populations of fast moving animals or populations with high densities will require less survey effort than those species that are slow moving or have populations with low densities.

We found that the sensitivity of the gREM to inaccurate parameter estimates was both predictable and reasonable (Appendix S6), although this varies between different parameters and gREM submodels. Whilst care should be taken in parameter estimation when analysing both acoustic and camera trap data, acoustic data poses particular problems. For acoustic surveys, estimates of r (detection distance) can be measured directly or calculated using sound attenuation models (Holderied & Von Helversen 2003), while the sensor angle is often easily measured (Adams et al. 2012) or found in the manufacturer's specifications. When estimating animal movement speed v, only the speed of movement during the survey period should be used. The signal width is the most sensitive parameter to inaccurate estimates (Appendix S6) and is also the most difficult to measure. While this parameter will typically be assumed to be 2π for camera trap surveys, fewer estimates exist for acoustic signal widths. Although signal width has been measured for echolocating bats using arrays of microphones (Brinkløv et al. 2011), more work should be done on obtaining estimates for a range of acoustically surveyed species.

5.5.3 Limitations

Although the REM has been found to be effective in field tests (Rowcliffe et al. 2008, Zero et al. 2013), the gREM requires further validation by both field tests and simulations. For example, capture-mark-recapture methods could be used alongside the gREM to test the accuracy under field conditions (Rowcliffe et al. 2008). While we found no effect of the movement model on the accuracy or precision of the gREM, the models we have used in our simulations to validate the gREM are still simple representations of true animal movement. Animal movement may be highly nonlinear and often dependent on multiple factors such as behavioural state and existence of home ranges (Smouse et al. 2010). Therefore testing the gREM against real animal data, or further simulations with more complex movement models, would be beneficial.

The assumptions of our simulations may require further consideration, for example, we have assumed an equal density across the study area. However, in a field environment the situation may be more complex, with additional variation coming from local changes in density between sensor sites. Athough unequal densities should theoretically not affect accuracy (Hutchinson & Waser 2007), it will affect precision and further simulations should be used to quantify this effect. Additionally, we allowed the sensor to be stationary and continuously de-

tecting, negating the triggering, and non-continuous recording issues that could exist with some sensors and reduce precision or accuracy. Finally, in the simulation animals moved at the equivalent of the largest day range of terrestrial animals (Carbone et al. 2005). Slower speed values should not alter the accuracy of the gREM, but precision would be affected since slower speeds produce fewer records. The gREM was both accurate and precise for all the movement models we tested (stop-start movement and correlated random walks).

A feature of the gREM is that it does not fit a statistical model to estimate detection probability as occupancy models and distance sampling do (Barlow & Taylor 2005, Marques et al. 2011, Royle & Nichols 2003). Instead it explicitly models the process, with animals only being detected if they approach the sensor from a suitable direction. Other processes that affect detection probability could be included in the model to improve realism.

5.5.4 Implications for ecology and conservation

The gREM is applicable for count data obtained either visually or acoustically in both marine and terrestrial environments, and is suitable for taxa including echolocating bats (Walters et al. 2012), songbirds (Buckland & Handel 2006), whales (Marques et al. 2011) and forest primates (Hassel-Finnegan et al. 2008). Many of these taxa contain critically endangered species and monitoring their populations is of conservation interest. For example, current methods of density estimation for the threatened Franciscana dolphin (*Pontoporia blainvillei*) may result in underestimation of their numbers (Crespo et al. 2010). In addition, using gREM may be easier than other methods for measuring the density of animals which may be useful in quantifying ecosystem services, such as songbirds with a known positive influence on pest control (Jirinec et al. 2011).

The gREM will aid researchers to study species with non-invasive methods such as remote sensors, which allows for large, continuous monitoring projects with limited human resources (Kelly et al. 2012). The gREM is also suitable for species that are sensitive to human contact or are difficult or dangerous to catch (Thomas & Marques 2012). As sensors such as camera traps and acoustic detectors become more ubiquitous, the gREM will be increasingly useful for monitoring unmarked animal populations across broad spatial, temporal and taxonomic scales.

6 Discussion

6.1 Overview

In this thesis I have aimed to examine the importance of population size and structure on the accumulation of pathogen richness. I used bats as a case study throughout due to their interesting and varied social structure (Kerth 2008) and their association with a number of important, recent zoonoses (Field et al. 2001, Halpin et al. 2011, Leroy et al. 2005, Li et al. 2005). I have studied the role of these population factors using both simulation studies and empirical, comparative approaches in order to both examine the specific, epidemiological mechanisms involved in a controlled and interpretable *in silico* environment, while also being able to link these results back to real-world data. I have found the most robust evidence so far that population structure does relate to higher pathogen richness in bats. However, my simulation study testing whether newly evolved pathogens would invade more easily in a structured population did not recover the same relationship. This implies that it is not this specific mechanism that is important in wild populations. Subsequently, I examined a number of intrinsically linked factors — population size, density and range size as well as colony size and the number of colonies — and found that contrary to beliefs commonly held in the literature, only colony size strongly promotes the invasion of newly evolved pathogens. Finally, I derived and validated a method for estimating bat population sizes from acoustic data; as bat population sizes are very difficult to estimate, this method fills a great need in bat ecology and zoonotic surveillance.

In Chapter 2 I tested the hypothesis that bat species with more structured populations harbour more virus species. I tested this hypothesis with two measurements of population structure: the number of subspecies (a novel measure and the largest dataset yet used to test this hypothesis) and gene flow. With both measures I found that, after controlling for phylogeny and study bias, a positive relationship between population structure and pathogen richness was very likely in the best model. This relationship was of similar strength, and at least as likely to be in the best model, as other measures (body mass and range size) which have been thought to promote pathogen richness in bats and other mammals (Arneberg 2002, Gay et al. 2014, Kamiya et al. 2014, Nunn et al. 2003, Turmelle & Olival 2009).

While the results from Chapter 2 suggest that there is a relationship between population structure and pathogen richness, comparative studies like these cannot identify which specific mechanisms maintain high pathogen richness. To examine this I developed a model of two recently diverged — and therefore

identical — pathogen lineages competing in a metapopulation i.e. a population of large bat colonies with limited movement between them (Chapter 3). I tested whether population structure (specifically network topology and dispersal rate) allowed a second pathogen to invade and persist in the presence of strong competition from the first, endemic pathogen. However, I found no relationship between probability of invasion and population structure. Instead it appeared that if transmission rate was high enough for the invading pathogen to survive the initial, highly stochastic part of its spread, it would then survive and spread throughout the metapopulation irregardless of how structured it was. This implies that local dynamics, defined in part by colony size, are controlling disease invasion and that a different mechanism must be causing the relationship seen in Chapter 2.

Group (or colony) size is one of many demographic parameters measured in comparative studies of pathogen richness. Other commonly measured parameters include population density and range size (Ezenwa et al. 2006, Gay et al. 2014, Kamiya et al. 2014, Lindenfors et al. 2007, Morand & Poulin 1998, Nunn et al. 2003) yet the intrinsic relationships between these variables are rarely acknowledged or discussed. Therefore in Chapter 4 I used the same model as Chapter 3 to test whether population density or population size more strongly promoted pathogen richness and whether a pathogen invaded more easily into a population comprising many small colonies or few big colonies. I found that population size has a much stronger affect than density and that the component of population size that has the strongest affect is colony size.

Theory (Anderson & May 1979, May & Anderson 1979), previous literature (Kamiya et al. 2014, Morand & Poulin 1998, Nunn et al. 2003) and Chapters 3 and 4 suggested that population size (either local group size or global population size) strongly influences the dynamics of disease and pathogen richness. However, there are very few estimates of population size for bats and colony counts are time consuming and costly (Kloepper et al. 2016). I therefore aimed to obtain estimates of population size from acoustic data such as the iBats program (Jones et al. 2011). I developed a general method for estimating population size and density from acoustic detectors (Chapter 5). I used spatial simulations of animal movement to validate the method and found it to be precise and unbiased.

6.2 Applications and implications for research

I have found evidence, both empirical and theoretical, that demographic parameters can influence pathogen richness. However it seems likely that this effect alone is not strong enough to be a useful predictor of viral richness with respect to surveillance for zoonotic diseases. While there is potential for population structure and colony size to be useful variables when combined with other variables in a predictive framework, the biases in all pathogen richness datasets makes these approaches difficult. However, as more unbiased data is collected — as in Anthony et al. (2013) and Anthony et al. (2015) — or using much larger pathogen data sets — such as Wardeh et al. (2015) — predictive models may become a more viable tool. Furthermore, the method provided in Chapter 5 makes the collection of population size data more feasible over broad taxonomic, spatial or temporal scales, further increasing the potential of predictive models. Field trials should test the gREM's ability to estimate population size and density and to ensure it is not strongly biased by species specific factors; only if it is unbiased can it be effectively used in predictive models and other applications.

While predictive models are difficult to build due to a lack of data and strong biases in pathogen richness data, the mechanistic understanding obtained by the theoretical chapters here can suggest how pathogen richness may respond to global change. Firstly, when global change acts to reduce group size (Atwood 2006, Lehmann et al. 2010, Manor & Saltz 2003, Zunino et al. 2007) pathogen richness is expected to decrease while in species where group size is increasing (Lehmann et al. 2010) pathogen richness is expected to increase. In contrast, species suffering range contractions (Thomas et al. 2004) and decreases in population size (Craigie et al. 2010) are expected to experience smaller changes in pathogen richness despite these being the more commonly studied effects of global change. This suggests that further research should study in more detail the effects of climate change on social group size.

Furthermore, I have shown that while population factors such as population size, density and range size are directly linked, they have very different effects on pathogen richness. Therefore future studies should be careful to acknowledge these relationships and, where possible, compare multiple demographic measurements to further test which factors are in fact causally affecting pathogen richness.

There is a common assumption that factors that increase R_0 should increase pathogen diversity (Morand 2000, Nunn et al. 2003). However, my results im-

ply a more nuanced relationship. In Chapters 3 and 4 I found that populations with large groups sizes, and therefore many localised contacts (i.e. high R_0), promote the invasion of new pathogen species, but that at the global level there is little or no affect of population structure. However, in Chapter 2 I found that in wild bat populations, global population structure does promote pathogen species richness. This implies that there are two distinct phases or scales to pathogen competition. When a new pathogen first enters a population, the local scale is important, and many contacts (i.e. a highly connected population) allows the pathogen to spread and avoid stochastic extinction. However, after this initial spread, the global scale may be more important as shown by the stronger support for mechanisms such as population structure (Chapter 2, Maganga et al. (2014) and Turmelle & Olival (2009)) and range size (Kamiya et al. 2014, Nunn et al. 2003) than group size (Ezenwa et al. 2006, Rifkin et al. 2012). This highlights the distinction between factors that promote the addition of new pathogens to the community and those factors that instead allow a larger number of pathogens to coexist or reduce the rate of extinction of pathogens due to competition or other processes. Little research has so far been conducted contrasting these different processes and examining which mechanisms could promote high pathogen richness at each scale.

Much research in multipathogen systems has been conducted over the short time scales of a single epidemic (Funk & Jansen 2010, Poletto et al. 2013, Poletto et al. 2015, van de Bovenkamp et al. 2014). While this time scale has important human health consequences, when examining the slow process of the accumulation of pathogen species, a longer term view needs to be examined. Interestingly, my results, along with previously published studies show quite strong differences between these timescales. Competing epidemics are strongly affected by population structure with structure promoting coexistence of pathogens and allowing less competitive pathogens to persist (Poletto et al. 2013, Poletto et al. 2015). In contrast, in the longer time scales studied here, I have found that population structure does not seem to allow an invading pathogen to escape competition (Chapters 3 and 4). This can be understood by considering that at very long time scales, any population is well mixed unless there is complete separation of subpopulations.

6.2.1 Further work

Other mechanisms controlling pathogen richness Colony size has been found to have a negative relationship (Gay et al. 2014) and no relationship (Turmelle & Olival 2009) with parasite richness in previous comparative studies using relatively small datasets. However, in Chapter 4 I found that colony size is particularly important for promoting pathogen richness. I did not include colony size in my comparative analysis (Chapter 2) for three reasons: the focus of the chapter was broad-scale population structure, the lack of previous evidence of a positive relationship (Gay et al. 2014, Turmelle & Olival 2009) and the lack of data. However, given the results of Chapter 4, filling these data gaps would be a useful avenue for further research. In particularly, testing the relative effects of colony size, population structure and range size would be a useful test of the model used in Chapter 4.

In this thesis I have only examined one mechanism by which demographic attributes may affect pathogen richness. I have only examined the ability of a newly evolved pathogen (i.e. a pathogen, identical to an endemic pathogen and in the presence of strong competition) to invade and persist.

However, there are a number of other mechanisms that could equally strongly affect pathogen richness in the wild. Closely related to the mechanism studied here is the case of pathogens invading from other host species. These pathogens are likely to have different epidemiological parameters (transmission rate, virulence, recovery rate) to the endemic pathogen. Furthermore, the competition between pathogens is expected to be less strong. Alternatively, host population traits could affect the rate of pathogen extinction. Once a number of pathogens are established in a population, there is still likely to be occasional extinctions, especially in the presence of inter-pathogen competition. A number of population factors could affect this rate. It is expected that large populations will experience slower rates of pathogen loss as stochastic extinction will be more rare. Furthermore, populations with strongly varying disease prevalences are likely to have higher rates of pathogen extinction. This includes populations where epidemic cycles are common (Altizer et al. 2006); subsequent to a large epidemic, and after a time lag, the number of susceptible individuals in the populations will be low due to immunity, host death or low birth rates induced by infection (Hethcote 1994, Scott & Lewis 1987). While the number of susceptibles is low, incidence could stochastically drop to zero for one disease generation, thus causing extinction. This effect will be exacerbated in the case where an epidemic cycle is syn-

chronous across the whole population (Duke-Sylvester et al. 2011). Structured populations with asynchronous epidemic cycles may experience local pathogen extinction but rarely global extinction; this pattern of local extinction and recolonisation has been well studied in the ecological literature (Grenfell et al. 1995, Hanski 1998, Levin 1974), but less so in the epidemiological literature.

Bat social structure Finally it is important to note that I have ignored much of the social complexity found in bats. In Chapters 3 and 4 I have modelled bat populations as a metapopulation where the only social structure is the grouping of individuals into subpopulations. There is dispersal between these subpopulations but otherwise they are static. Similarly, information on these other social behaviours was not explicitly included in the regression in Chapter 2. Firstly, I have not modelled the creation of new colonies, or the disbanding of colonies (Metheny et al. 2008). Especially in the face of habitat destruction, it is likely that the number of colonies of a species will be decreasing. Furthermore, in some species, colonies are likely to be more fluid, with groups joining and splitting (August et al. 2014, Kerth & Van Schaik 2012). Secondly, there are a number of behaviours common in bats, particularly in temperate regions, that has been excluded from these models. For example, many species have different types of colonies — maternity colonies, mating colonies and hibernation colonies (Kerth 2008). Epidemiological dynamics are likely to be altered by the physiological differences in bats while in these different colony types but also due to their role in population structure (Blehert 2012, George et al. 2011, Langwig et al. 2015, Webber et al. 2016). The extent to which the individuals move together when switching between these colony types is largely unknown (Baerwald & Barclay 2016, Kurta & Murray 2002) but if there is a large degree of mixing during the transition between colony types, then there will be considerably less population structure overall. Similarly, swarming behaviour — the coming together of many bats from different colonies — is likely to decrease epidemiological population structure (Kerth & Van Schaik 2012).

Furthermore, many bat species, both temperate and tropical, are migratory (Fleming et al. 2003, Hutterer et al. 2005, Krauel & McCracken 2013, Popa-Lisseanu & Voigt 2009). Again, it is largely unknown whether colonies travel together during migration. It is therefore also unknown whether colony structure is similar before and after migration (Carter & Wilkinson 2013) though Kurta & Murray (2002) find that individuals do not migrate together. There is also little data on whether parameters such as dispersal rate are constant before and after

migration, though it is likely that most dispersal between colonies is juvenile dispersal and so dispersal rate is likely to be much higher in one location than the other. Even if colonies remain fairly constant during migration, the spatial relationships may be different; colonies that were far apart in one area could subsequently be near neighbours after migration. Migratory status has been included in previous comparative analyses and not been found to be a strong predictor of pathogen richness (Maganga et al. 2014, Turmelle & Olival 2009).

Another potentially important factor that has been ignored here is roost sharing by different bat species (de Thoisy et al. 2016, López-Roig et al. 2014, Maganga et al. 2014, Pons-Salort et al. 2014, Serra-Cobo et al. 2002). If the species are very similar in most epidemiological factors, this could potentially be sensibly modelled by ignoring species identity and treating the whole population as one. However, it is more likely that there will be fewer close contact events between individuals of different species even if they roost share. It is also likely that species will have different dispersal patterns between colonies. Therefore, more complex models such as overlay network models might be needed in order to effectively model these populations (Funk & Jansen 2010, Marceau et al. 2011). Roost sharing and the amount of sympatry has been included in comparative studies of bat pathogen richness (Maganga et al. 2014) but was not found to correlate with pathogen richness.

Finally, birth and deaths have been modelled here as occurring randomly through time but many bat species have very tightly controlled birth pulses (Dietrich et al. 2015, George et al. 2011, Greiner et al. 2011, Porter & Wilkinson 2001). This has important epidemiological consequences; there will be a pulse of susceptible individuals each year with very few new susceptibles during the rest of the year (Dietrich et al. 2015). Models of these population dynamics have found that birth pulses can drive pathogen extinction (Peel et al. 2014). Hayman (2015) found that certain Filoviruses were less likely to persist in bat species with an annual birth pulse than a biannual birth pulse. In other mammals, birth pulses have also been shown to reduce synchrony of dynamics (Duke-Sylvester et al. 2011).

Overall, there is much complexity that could be added to epidemiological models of bats. However, there is little data for many species which makes parameterisation difficult. Furthermore, as these factors differ between species, trying to make general models that apply across the order is difficult. Further work should include specific, detailed models of well studied species and further examination of how important these various factors might be.

6.2.2 Conclusions

Overall my studies suggest that population size and structure do have consequences for pathogen richness. However, the exact mechanisms by which these effects occur are not clear. I have found that colony size is particularly important in the case of closely related, strongly competing pathogens. I have also provided a tool to facilitate the estimation of population sizes in echolocating bats; data which is currently sparse despite its importance to epidemiology and bat ecology more generally.

Bibliography

Acevedo MA & Villanueva-Rivera LJ (2006). 'Using automated digital recording systems as effective tools for the monitoring of birds and amphibians'. *Wildl. Soc. Bull.* **34**(1), 211–214. DOI: 10.1016/s1769-4493(08)70101-8

Ackleh AS & Allen LJ (2003). 'Competitive exclusion and coexistence for pathogens in an epidemic model with variable population size'. *J. Math. Biol.* **47**(2), 153–168. DOI: 10.1007/s00285-003-0207-9

Ackleh AS & Salceanu PL (2014). 'Robust uniform persistence and competitive exclusion in a nonautonomous multi-strain SIR epidemic model with disease-induced mortality'. *J. Math. Biol.* (1-2), 453–475. DOI: 10.1007/s00285-012-0636-4

Adams A, Jantzen M, Hamilton R & Fenton M (2012). 'Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats'. *Methods Ecol. Evol.* DOI: 10.1126/science.291.5510.1855k

Aldridge H (1987). 'Turning flight of bats'. *J. Exp. Biol.* **128**(1), 419–425

Allen LJ, Kirupaharan N & Wilson SM (2004). 'SIS epidemic models with multiple pathogen strains'. *J. Differ. Equations Appl.* **10**(1), 53–75. DOI: 10.1080/10236190310001603680

Allen LJ, Langlais M & Phillips CJ (2003). 'The dynamics of two viral infections in a single host population with applications to hantavirus'. *Math. Biosci.* **186**(2), 191–217. DOI: 10.1016/j.mbs.2003.08.002

Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M & Rohani P (2006). 'Seasonality and dynamics of infectious diseases.' *Ecol. Lett.* **9**, 467–484. DOI: 10.1111/j.1461-0248.2005.00879.x

Altizer S, Nunn CL, Thrall PH, Gittleman JL, Antonovics J, Cunnningham AA, Dobson AP, Ezenwa V, Jones KE et al. (2003). 'Social organization and parasite risk in mammals: integrating theory and empirical studies'. *Annu. Rev. Ecol. Evol. Syst.* 517–547. DOI: 10.1146/annurev.ecolsys.34.030102.151725

Amengual B, Bourhy H, López-Roig M & Serra-Cobo J (2007). 'Temporal dynamics of European bat Lyssavirus type 1 and survival of *Myotis myotis* bats in natural colonies'. *PLoS One* **2**(6), e566. DOI: 10.1371/journal.pone.0000566

Amman B, Carroll S, Reed Z, Sealy T, Balinandi S, Swanepoel R, Kemp A, Erickson B, Comer J, Campbell S et al. (2012). 'Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection'. *PLoS Pathog.* **8**(10), e1002877

Anderson DR (2001). 'The need to get the basics right in wildlife field studies'. *Wildl. Soc. Bull.* **29**(4), 1294–1297. DOI: 10.1211/pj.2013.11124181

Anderson RM & May RM (1979). 'Population biology of infectious diseases: Part I.' *Nature* (280), 361–7. DOI: 10.1038/280361a0

Andren H (1994). 'Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review'. *Oikos*, 355–366. DOI: 10.2307/3545823

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, Solovyov A, Ojeda-Flores R, Arrigo NC, Islam A, Khan SA, Hosseini P, Bogich TL, Olival KJ, Sanchez-Leon MD, Karesh WB, Goldstein T, Luby SP, Morse SS, Mazet JAK, Daszak P & Lipkin WI (2013). 'A strategy to estimate unknown viral diversity in mammals'. *MBio* **4**(5), e00598–13. DOI: 10.1128/mBio.00598-13

Anthony SJ, Islam A, Johnson C, Navarrete-Macias I, Liang E, Jain K, Hitchens PL, Che X, Soloyvov A, Hicks AL, Ojeda-Flores R, Zambrana-Torrelio C, Ulrich W, Rostal MK, Petrosov A, Garcia J, Haider N, Wolfe N, Goldstein T, Morse SS, Rahman M, Epstein JH, Mazet JK, Daszak P & Lipkin WI (2015). 'Non-random patterns in viral diversity'. *Nat. Commun.* **6**. DOI: 10.1038/ncomms9147

Arita HT (1993). 'Rarity in Neotropical bats: correlations with phylogeny, diet, and body mass'. *Ecol. Appl.* 506–517. DOI: 10.2307/1941919

Arneberg P (2002). 'Host population density and body mass as determinants of species richness in parasite communities: comparative analyses of directly transmitted nematodes of mammals'. *Ecography* (1), 88–94. DOI: 10.1034/j.1600-0587.2002.250110.x

Atwood TC (2006). 'The influence of habitat patch attributes on coyote group size and interaction in a fragmented landscape'. *Can. J. Zool.* **84**(1), 80–87. DOI: 10.1139/z05-180

August TA, Nunn MA, Fensome AG, Linton DM & Mathews F (2014). 'Sympatric woodland *Myotis* bats form tight-knit social groups with exclusive roost home ranges'. *PloS one* **9**(10), e112225. DOI: 10.1371/journal.pone.0112225

Baerwald EF & Barclay RMR (2016). 'Are migratory behaviours of bats socially transmitted?' *R. Soc. Open Sci.* **3**(4). DOI: 10.1098/rsos.150658. eprint: http://rsos.royalsocietypublishing.org/content/3/4/150658.full.pdf. URL: http://rsos.royalsocietypublishing.org/content/3/4/150658

Barlow J & Taylor B (2005). 'Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey'. *Mar. Mamm. Sci.* **21**(3), 429–445. DOI: 10.1111/j.1748-7692.2005.tb01242.x

Barthélemy M, Godreche C & Luck JM (2010). 'Fluctuation effects in metapopulation models: percolation and pandemic threshold'. *J. Theor. Biol.* **267**(4), 554–564

Begon M, Bennett M, Bowers RG, French NP, Hazel S, Turner J et al. (2002). 'A clarification of transmission terms in host-microparasite models: numbers, densities and areas'. *Epidemiol. Infect.* **129**(1), 147–153

Belliure J, Sorci G, Møller A & Clobert J (2000). 'Dispersal distances predict subspecies richness in birds'. *Journal of Evolutionary Biology* **13**(3), 480–487

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL & Purvis A (2007). 'The delayed rise of present-day mammals'. *Nature* **446**(7135), 507–512. DOI: 10.3410/f.1076757.529672

Birt P, McCoy M & Palmer C (2008). 'Little Red Flying-fox, *Pteropus scapulatus*'. In: *The Mammals of Australia*. Ed. by S Van Dyck & R Strahan. 3rd ed. Sydney, Australia: Reed New Holland, 446–447

Blackburn TM, Cassey P & Gaston KJ (2006). 'Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution'. *J. Anim. Ecol.* **75**(6), 1426–1439. DOI: 10.1111/j.1365-2656.2006.01167.x

Blackwood JC, Streicker DG, Altizer S & Rohani P (2013). 'Resolving the roles of immunity, pathogenesis, and immigration for rabies persistence in vampire bats'. *PNAS* **110**(51), 20837–20842. DOI: 10.1073/pnas.1308817110

Blehert DS (2012). 'Fungal disease and the developing story of bat white-nose syndrome'. *PLoS Pathog.* **8**(7), e1002779. DOI: 10.1371/journal.ppat.1002779

Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, Deppe JL, Krakauer AH, Clark C, Cortopassi KA, Hanser SF, McCowan B, Ali AM & Kirschel ANG (2011). 'Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus'. *J. Appl. Ecol.* **48**(3), 758–767. DOI: 10.1111/j.1365-2664.2011.01993.x

Borchers D, Distiller G, Foster R, Harmsen B & Milazzo L (2014). 'Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey'. *Methods Ecol. Evol.* **5**(7), 656–665

Bordes F, Blumstein DT & Morand S (2007). 'Rodent sociality and parasite diversity'. *Biol. Lett.* **3**(6), 692–694. DOI: 10.1098/rsbl.2007.0393

Bordes F & Morand S (2011). 'The impact of multiple infections on wild animal hosts: a review'. *Infect. Ecol. Epidemiol.* **1**. DOI: 10.3402/iee.v1i0.7346

Bordes F, Morand S & Ricardo G (2008). 'Bat fly species richness in Neotropical bats: correlations with host ecology and host brain'. *Oecologia* **158**(1), 109–116. DOI: 10.1007/s00442-008-1115-x

Borregaard MK & Rahbek C (2010). 'Causality of the relationship between geographic distribution and species abundance'. *Q. Rev. Biol.* **85**(1), 3–25. DOI: 10.1086/650265

Bremermann HJ & Thieme H (1989). 'A competitive exclusion principle for pathogen virulence'. *J. Math. Biol.* **27**, 179–190. DOI: 10.1007/BF00276102

Brierley L, Vonhof M, Olival K, Daszak P & Jones K (2016). 'Quantifying Global Drivers of Zoonotic Bat Viruses: A Process-based Perspective'. *Am. Nat.* **187**

Brinkløv S, Jakobsen L, Ratcliffe J, Kalko E & Surlykke A (2011). 'Echolocation call intensity and directionality in flying short-tailed fruit bats, *Carollia perspicillata* (Phyllostomidae)'. *J. Acoust. Soc. Am.* **129**(1), 427–435. DOI: 10.3389/fphys.2013.00089

Brown JH, Gillooly JF, Allen AP, Savage VM & West GB (2004). 'Toward a metabolic theory of ecology'. *Ecology* **85**(7), 1771–1789. DOI: 10.1890/03-9000

Brusa A & Bunker DE (2014). 'Increasing the precision of canopy closure estimates from hemispherical photography: Blue channel analysis and under-exposure'. *Agric. For. Meteorol.* **195**, 102–107. DOI: 10.1016/j.agrformet.2014.05.001

Buckee CO, Koelle K, Mustard MJ & Gupta S (2004). 'The effects of host contact network structure on pathogen diversity and strain structure'. *PNAS* **101**(29), 10839–10844. DOI: 10.1073/pnas.0402000101

Buckland ST & Handel C (2006). 'Point-transect surveys for songbirds: robust methodologies'. *The Auk* **123**(2), 345–357. DOI: 10.1642/0004-8038(2006)123[345:psfsrm]2.0.co;2

Buckland ST, Marsden SJ & Green RE (2008). 'Estimating bird abundance: making methods work'. *Bird Conserv. Int.* **18**(S1), S91–S108. DOI: 10.1017/s0959270908000294

Burland T, Barratt E, Beaumont M & Racey P (1999). 'Population genetic structure and gene flow in a gleaning bat, *Plecotus auritus*'. *Proc. R. Soc. B* **266**(1422), 975–980. DOI: 10.1007/978-3-642-70837-4_20

Burnham KP & Anderson DR (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag. DOI: 10.1007/978-0-387-22456-5_1

Burns LE & Broders HG (2014). 'Correlates of dispersal extent predict the degree of population genetic structuring in bats'. *Conserv. Genet.* **15**(6), 1371–1379

Calisher C, Childs J, Field H, Holmes K & Schountz T (2006). 'Bats: important reservoir hosts of emerging viruses'. *Clin. Microbiol. Rev.* **19**(3), 531–545. DOI: 10.1128/CMR.00017-06

Campos PR & Gordo I (2006). 'Pathogen genetic variation in small-world host contact structures'. *J. Stat. Mech. Theor. Exp.* **2006**(12), L12003. DOI: 10.1088/1742-5468/2006/12/L12003

Canals M, Atala C, Grossi B & Iriarte-Díaz J (2005). 'Relative size of hearts and lungs of small bats'. *Acta Chiropt.* **7**(1), 65–72. DOI: 10.3161/1733-5329(2005)7[65:rsohal]2.0.co;2

Carbone C, Cowlishaw G, Isaac NJ & Rowcliffe JM (2005). 'How far do animals go? Determinants of day range in mammals'. *Am. Nat.* **165**(2), 290–297. DOI: 10.1086/426790

Carter GG & Wilkinson GS (2013). 'Cooperation and conflict in the social lives of bats'. In: *Bat evolution, ecology, and conservation*. Ed. by RA Adams & SC Pedersen. Springer, 225–242. DOI: 10.1007/978-1-4614-7397-8_12

Castillo-Chavez C, Hethcote H, Andreasen V, Levin S & Liu WM (1989). 'Epidemiological models with age structure, proportionate mixing, and cross-immunity'. *Journal of mathematical biology* **27**(3), 233–258. DOI: 10.1007/BF00275810

Castillo-Chavez C, Huang W & Li J (1995). 'Dynamics of multiple pathogen strains in heterosexual epidemiological models'. Ed. by M Martelli, K Cooke, E Cumberbatch, B Tang & H Thieme, 289–298

Chamberlain SA & Szöcs E (2013). 'taxize: taxonomic search and retrieval in R'. *F1000Research* **2**. DOI: 10.12688/f1000research.2-191.v1

Chinnasamy K, Pitchamuthu M, Doss PS, Marimuthu G & Rajan KE (2013). 'Genetic diversity and population structure of leaf-nosed bat Hipposideros speoris (Chiroptera: Hipposideridae) in Indian subcontinent'. *African Journal of Biotechnology* **10**(8), 1320–1328

Clement MJ & Castleberry SB (2013). 'Estimating density of a forest-dwelling bat: a predictive model for Rafinesque's big-eared bat'. *Popul. Ecol.* **55**(1), 205–215. DOI: 10.1007/s10144-012-0356-z

Colizza V & Vespignani A (2007). 'Invasion threshold in heterogeneous metapopulation networks'. *Phys. Rev. Lett.* **99**(14), 148701. DOI: 10.1103/physrevlett.99.148701

Craigie ID, Baillie JE, Balmford A, Carbone C, Collen B, Green RE & Hutton JM (2010). 'Large mammal population declines in Africa's protected areas'. *Biol. Cons.* **143**(9), 2221–2228. DOI: 10.1016/j.biocon.2010.06.007

Crespo EA, Pedraza SN, Grandi MF, Dans SL & Garaffo GV (2010). 'Abundance and distribution of endangered Franciscana dolphins in Argentine waters and conservation implications'. *Mar. Mamm. Sci.* **26**(1), 17–35. DOI: 10.3354/esr00391

Cryan PM, Stricker CA & Wunder MB (2014). 'Continental-scale, seasonal movements of a heterothermic migratory tree bat'. *Ecol. Appl.* **24**(4), 602–616. DOI: 10.1890/13-0752.1

Cushman SA, Shirk A & Landguth EL (2012). 'Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes'. *Landscape ecology* **27**(3), 369–380. DOI: 10.1007/s10980-011-9693-0

Damuth J (1981). 'Population density and body size in mammals'. *Nature* **290**(5808), 699–700. DOI: 10.1038/290699a0

De Thoisy B, Bourhy H, Delaval M, Pontier D, Dacheux L, Darcissac E, Donato D, Guidez A, Larrous F, Lavenir R, Salmier A, Lacoste V & Lavergne A (2016). 'Bioecological Drivers of Rabies Virus Circulation in a Neotropical Bat Community'. *PLoS Negl. Trop. Dis.* **10**(1). Ed. by M Kasper,

e0004378. DOI: 10.1371/journal.pntd.0004378. URL: http://dx.doi.org/10.1371/journal.pntd.0004378

Depraetere M, Pavoine S, Jiguet F, Gasc A, Duvail S & Sueur J (2012). 'Monitoring animal diversity using acoustic indices: implementation in a temperate woodland'. *Ecol. Indic.* **13**(1), 46–54. DOI: 10.1016/j.ecolind.2011.05.006

Dietrich M, Wilkinson DA, Benlali A, Lagadec E, Ramasindrazana B, Dellagi K & Tortosa P (2015). 'Leptospira and paramyxovirus infection dynamics in a bat maternity enlightens pathogen maintenance in wildlife'. *Environ. Microbiol.* **17**(11), 4280–4289

Dobson AP (2005). 'What links bats to emerging infectious diseases?' *Science* **310**(5748), 628–629. DOI: 10.1126/science.1120872

Drosten C, Günther S, Preiser W, Van Der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA, Berger A, Burguière AM, Cinatl J, Eickmann M, Escriou N, Grywna K, Kramme S, Manuguerra JC, Müller S, Rickerts V, Stürmer M, Vieth S, Klenk HD, Osterhaus AD, Schmitz H, Doerr HW et al. (2003). 'Identification of a novel coronavirus in patients with severe acute respiratory syndrome'. *N. Engl. J. Med.* **348**(20), 1967–1976. DOI: 10.1056/NEJMoa030747

Duke-Sylvester SM, Bolzoni L & Real LA (2011). 'Strong seasonality produces spatial asynchrony in the outbreak of infectious diseases'. *J. R. Soc. Interface* **8**(59), 817–825. DOI: 10.1098/rsif.2010.0475

Dunn RR, Davies TJ, Harris NC & Gavin MC (2010). 'Global drivers of human pathogen richness and prevalence'. *Proc. R. Soc. B*, rspb20100340. DOI: 10.1098/rspb.2010.0340

Everatt KT, Andresen L & Somers MJ (2014). 'Trophic scaling and occupancy analysis reveals a lion population limited by top-down anthropogenic pressure in the Limpopo National Park, Mozambique'. *PLoS One* **9**(6), e99389. DOI: 0120099389

Ezenwa VO, Price SA, Altizer S, Vitone ND & Cook KC (2006). 'Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla'. *Oikos* **115**(3), 526–536. DOI: 10.1111/j.2006.0030-1299.15186.x

Fenton A & Perkins SE (2010). 'Applying predator-prey theory to modelling immune-mediated, within-host interspecific parasite interactions'. *Parasitology* **137**(06), 1027–1038. DOI: 10.1017/s0031182009991788

Field H, Young P, Yob J, Mills J, Hall L & Mackenzie J (2001). 'The natural history of Hendra and Nipah viruses'. *Microbes Infect.* **3**(4), 307–314. DOI: 10.1016/S1286-4579(01)01384-3

Fleming TH, Eby P, Kunz T & Fenton M (2003). 'Ecology of bat migration'. In: *Bat ecology*. Ed. by T Kunz & M Brock Fenton. Chicago: The University of Chicago Press, 156–208

Fontúrbel FE, Candia AB, Salazar DA, Malebrán J, González-Browne C & Botto-Mahan C (2014). 'How forest marsupials are affected by habitat degradation and fragmentation? A meta-analysis'. *Naturwissenschaften* **101**(7), 599–602. DOI: 10.1007/s00114-014-1193-z

Foster RJ & Harmsen BJ (2012). 'A critique of density estimation from camera-trap data'. *J. Wildl. Manage.* **76**(2), 224–236. DOI: 10.1002/jwmg.275

Fox S (2006). 'Population structure in the spectacled flying fox, *Pteropus conspicillatus*: a study of genetic and demographic factors'. PhD thesis. James Cook University

Fritz SA, Bininda-Emonds OR & Purvis A (2009). 'Geographical variation in predictors of mammalian extinction risk: big is bad, but only in the tropics'. *Ecol. Lett.* **12**(6), 538–549. DOI: 10.3410/f.1166120.627034

Funk S & Jansen VA (2010). 'Interacting epidemics on overlay networks'. *Phys. Rev. E* **81**(3), 036118

Garmer S, Lynn R, Rossi D & Capaldi A (2016). 'Multistrain Infections in Metapopulations'. *Spora: A Journal of Biomathematics* **1**(1), 17–27. URL: http://journals.sfu.ca/spora/index.php/spora/article/view/7

Gay N, Olival KJ, Bumrungsri S, Siriaroonrat B, Bourgarel M & Morand S (2014). 'Parasite and viral species richness of Southeast Asian bats: Fragmentation of area distribution matters'. *Int. J. Parasitol. Parasites Wildl.* **3**(2), 161–170. DOI: 10.1016/j.ijppaw.2014.06.003

George DB, Webb CT, Farnsworth ML, O'Shea TJ, Bowen RA, Smith DL, Stanley TR, Ellison LE & Rupprecht CE (2011). 'Host and viral ecology determine bat rabies seasonality and maintenance'. *PNAS* **108**(25), 10208–10213. DOI: 10.1073/pnas.1010875108

Gippoliti S & Amori G (2007). 'The problem of subspecies and biased taxonomy in conservation lists: the case of mammals'. *Folia Zool.* **56**(2), 113

Goffman W & Newill V (1964). 'Generalization of epidemic theory'. *Nature* **204**(4955), 225–228

Gómez-Rodríguez RA, Gutiérrez-Granados G, Montiel-Parra G, Rodríguez-Moreno Á & Sánchez-Cordero V (2015). 'Diversity and Coexistence of Ectoparasites in Small Rodents in a Tropical Dry Forest'. *Biotropica* **47**(4), 484–490. DOI: 10.1111/btp.12229

Goodman SM, Chan LM, Nowak MD & Yoder AD (2010). 'Phylogeny and biogeography of western Indian Ocean Rousettus (Chiroptera: Pteropodidae)'. *J. Mammal.* **91**(3), 593–606

Granich R, Gupta S, Hersh B, Williams B, Montaner J, Young B & Zuniga JM (2015). 'Trends in AIDS deaths, new infections and ART coverage in the top 30 countries with the highest AIDS mortality burden; 1990–2013'. *PLoS One* **10**(7), e0131353. DOI: 10.1371/journal.pone.0131353

Gregory R (1990). 'Parasites and host geographic range as illustrated by waterfowl'. *Funct. Ecol.* 645–654. DOI: 10.2307/2389732

Greiner S, Schwarzenberger F & Voigt CC (2011). 'Predictable timing of oestrus in the tropical bat *Saccopteryx bilineata* living in a Costa Rican rain forest'. *J. Trop. Ecol.* **27**(02), 121–131. DOI: 10.1017/S0266467410000696

Grenfell B, Bolker B & Kleczkowski A (1995). 'Seasonality and extinction in chaotic metapopulations'. *Proc. R. Soc. B* **259**(1354), 97–103. DOI: 10.1098/rspb.1995.0015

Griffiths EC, Pedersen AB, Fenton A & Petchey OL (2011). 'The nature and consequences of coinfection in humans'. *Journal of Infection* **63**(3), 200–206. DOI: 10.1016/j.jinf.2011.06.005

Griffiths EC, Pedersen AB, Fenton A & Petchey OL (2014). 'Analysis of a summary network of co-infection in humans reveals that parasites interact most via shared resources'. *Proc. R. Soc. B* **281**(1782), 20132286. DOI: 10.1098/rspb.2013.2286

Halpin K, Hyatt AD, Fogarty R, Middleton D, Bingham J, Epstein JH, Rahman SA, Hughes T, Smith C, Field HE et al. (2011). 'Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission'. *Am. J. Trop. Med. Hyg.* **85**(5), 946–951. DOI: 10.4269/ajtmh.2011.10-0567

Han BA, Park AW, Jolles AE & Altizer S (2015). 'Infectious disease transmission and behavioural allometry in wild mammals'. *J. Anim. Ecol.* **84**(3), 637–646. DOI: 10.1111/1365-2656.12336

Hanski I (1998). 'Metapopulation dynamics'. *Nature* **396**(6706), 41–49. DOI: 10.1038/23876

Harris D, Matias L, Thomas L, Harwood J & Geissler WH (2013). 'Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic'. *J. Acoust. Soc. Am.* **134**(5), 3522–3535. DOI: 10.1121/1.4821207

Hassel-Finnegan HM, Borries C, Larney E, Umponjan M & Koenig A (2008). 'How reliable are density estimates for diurnal primates?' *Int. J. Primatol.* **29**(5), 1175–1187. DOI: 10.1007/s10764-008-9301-6

Hayman DT (2015). 'Biannual birth pulses allow filoviruses to persist in bat populations'. *Proc. R. Soc. B* **282**(1803), 20142591

Heaney LR, Balete DS, Alviola P, Rickart EA & Ruedi M (2012). '*Nyctalus plancyi* and *Falsistrellus petersi* (Chiroptera: Vespertilionidae) from northern Luzon, Philippines: ecology, phylogeny, and biogeographic implications'. *Acta Chiropt.* **14**(2), 265–278. DOI: 10.3161/150811012x661602

Heesterbeek J (2002). 'A brief history of R 0 and a recipe for its calculation'. *Acta Biotheor.* **50**(3), 189–204. DOI: 10.1023/A:1016599411804

Henderson LE & Broders HG (2008). 'Movements and resource selection of the Northern Long-Eared Myotis (*Myotis septentrionalis*) in a forest–agriculture landscape'. *J. Mammal.* **89**(4), 952–963. DOI: 10.1644/07-mamm-a-214.1

Hethcote HW (1994). 'A thousand and one epidemic models'. In: *Frontiers in mathematical biology*. Springer, 504–515. DOI: 10.1007/978-3-642-50124-1_29

Holderied M & Von Helversen O (2003). 'Echolocation range and wingbeat period match in aerial-hawking bats'. *Proc. R. Soc. B* **270**(1530), 2293–2299. DOI: 10.1098/rspb.2003.2487

Huang S, Drake JM, Gittleman JL & Altizer S (2015). 'Parasite diversity declines with host evolutionary distinctiveness: A global analysis of carnivores'. *Evolution* **69**(3), 621–630. DOI: 10.1111/evo.12611

Hughes JM, Wilson ME, Halpin K, Hyatt AD, Plowright RK, Epstein JH, Daszak P, Field HE, Wang L & Daniels PW (2007). 'Emerging viruses: coming in on a wrinkled wing and a prayer'. *Clin. Infect. Dis.* **44**(5), 711–717. DOI: 10.1086/511078

Hulva P, Fornůsková A, Chudárková A, Evin A, Allegrini B, Benda P & Bryja J (2010). 'Mechanisms of radiation in a bat group from the genus Pipistrellus inferred by phylogeography, demography and population genetics'. *Mol. Ecol.* **19**(24), 5417–5431

Humphrey SR (1971). 'Photographic estimation of population size of the Mexican free-tailed bat, *Tadarida brasiliensis*'. *Am. Midl. Nat.* 220–223. DOI: 10.2307/2423705

Hutchinson JMC & Waser PM (2007). 'Use, misuse and extensions of "ideal gas" models of animal encounter.' *Biol. Rev. Camb. Philos. Soc.* **82**(3), 335–359. DOI: 10.1111/j.1469-185x.2007.00014.x

Hutterer R, Ivanova T, Meyer-Cords C & Rodrigues L (2005). *Bat migrations in Europe: a review of banding data and literature*. Bonn, Germany: Federal Agency for Nature Conservation

ITIS (2015). *Integrated Taxonomic Information System*. http://www.itis.gov

IUCN (2010). *Red List Of Threatened Species. Version 2010.1*. www.iucnredlist.org

JabRef Development Team (2015). *JabRef*. URL: http://jabref.sf.net

Jenkins DG, Carey M, Czerniewska J, Fletcher J, Hether T, Jones A, Knight S, Knox J, Long T, Mannino M et al. (2010). 'A meta-analysis of isolation by distance: relic or reference standard for landscape genetics?' *Ecography* **33**(2), 315–320. DOI: 10.1111/j.1600-0587.2010.06285.x

Jirinec V, Campos BR & Johnson MD (2011). 'Roosting behaviour of a migratory songbird on Jamaican coffee farms: landscape composition may affect delivery of an ecosystem service'. *Bird Conserv. Int.* **21**(03), 353–361. DOI: 10.1017/s0959270910000614

Jones KE, Bininda-Emonds OR & Gittleman JL (2005). 'Bats, clocks, and rocks: diversification patterns in Chiroptera'. *Evolution* **59**(10), 2243–2255. DOI: 10.1554/04-635.1

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL & Daszak P (2008). 'Global trends in emerging infectious diseases'. *Nature* **451**(7181), 990–993. DOI: 10.3201/eid0903.020336

Jones KE, Russ JA, Bashta AT, Bilhari Z, Catto C, Csősz I, Gorbachev A, Győrfi P, Hughes A, Ivashkiv I, Koryagina N, Kurali A, Langton S, Collen A, Margiean G, Pandourski I, Parsons S, Prokofev I, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Walters CL, Weatherill A & Zavarzin O (2011). 'Indicator Bats Program: a system for the global acoustic monitoring of bats'. *Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment and Local Action*. Ed. by B Collen, N Pettorelli, JEM Baillie & SM Durant, 211–247. DOI: 10.1002/9781118490747.ch10

Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C et al. (2009). 'PanTHERIA: a species-level database

of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184'. *Ecology* **90**(9), 2648–2648. DOI: 10.1890/08-1494.1

Kamiya T, O'Dwyer K, Nakagawa S & Poulin R (2014). 'What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts'. *Biol. Rev.* **89**(1), 123–134. DOI: 10.1111/brv.12046

Karanth K (1995). 'Estimating tiger (*Panthera tigris*) populations from camera-trap data using capture–recapture models'. *Biol. Conserv.* **71**(3), 333–338. DOI: 10.1016/0006-3207(94)00057-w

Kelly MJ, Betsch J, Wultsch C, Mesa B & Mills LS (2012). 'Noninvasive sampling for carnivores.' In: *Carnivore ecology and conservation: a handbook of techniques*. Ed. by L Boitani & R Powell. New York: Oxford University Press, 47–69. DOI: 10.1093/acprof:oso/9780199558520.003.0004

Kerth G, Mayer F & Petit E (2002). 'Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat (*Myotis bechsteinii*)'. *Mol. Ecol.* **11**(8), 1491–1498. DOI: 10.1046/j.1365-294x.2002.01528.x

Kerth G & Van Schaik J (2012). 'Causes and consequences of living in closed societies: lessons from a long-term socio-genetic study on Bechstein's bats'. *Mol. Ecol.* **21**(3), 633–646. DOI: 10.1111/j.1365-294X.2011.05233.x

Kerth G (2008). 'Causes and consequences of sociality in bats'. *Bioscience* **58**(8), 737–746. DOI: 10.1641/B580810

Kerth G, Perony N & Schweitzer F (2011). 'Bats are able to maintain long-term social relationships despite the high fission–fusion dynamics of their groups'. *Proc. R. Soc. B* **278**(1719), 2761–2767

Kessel S, Cooke S, Heupel M, Hussey N, Simpfendorfer C, Vagle S & Fisk A (2014). 'A review of detection range testing in aquatic passive acoustic telemetry studies'. *Rev. Fish Biol. Fish.* **24**(1), 199–218. DOI: 10.1007/s11160-013-9328-4

Kimura S, Akamatsu T, Dong L, Wang K, Wang D, Shibata Y & Arai N (2014). 'Acoustic capture-recapture method for towed acoustic surveys of echolocating porpoises'. *J. Acoust. Soc. Am.* **135**(6), 3364–3370. DOI: 10.1121/1.4875710

King AM, Adams MJ & Lefkowitz EJ (2011). *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. Vol. 9. San Diego, CA: Elsevier. ISBN: 978-0123846846

Kirupaharan N & Allen LJ (2004). 'Coexistence of multiple pathogen strains in stochastic epidemic models with density-dependent mortality'. *Bull. Math. Biol.* **66**(4), 841–864

Kloepper LN, Linnenschmidt M, Blowers Z, Branstetter B, Ralston J & Simmons JA (2016). 'Estimating colony sizes of emerging bats using acoustic recordings'. *R. Soc. Open Sci.* **3**(3), 160022. DOI: 10.1098/rsos.160022

Knobler S, Lemon S, Mack A, Sivitz L & Oberholtzer K (2004). *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary*. National Academies Press Washington, DC, USA: DOI: 10.1016/j.socscimed.2006.08.004

Krauel JJ & McCracken GF (2013). 'Recent advances in bat migration research'. In: *Bat Evolution, Ecology, and Conservation*. Springer, 293–313. DOI: 10.1007/978-1-4614-7397-8_15

Kurta A & Murray SW (2002). 'Philopatry and migration of banded Indiana bats (*Myotis sodalis*) and effects of radio transmitters'. *J. Mammal.* **83**(2), 585–589

Kuzmin IV, Bozick B, Guagliardo SA, Kunkel R, Shak JR, Tong S & Rupprecht CE (2011). 'Bats, emerging infectious diseases, and the rabies paradigm revisited'. *Emerg. Health Threats J.* **4**. DOI: 10.3402/ehtj.v4i0.7159

Langwig KE, Frick WF, Reynolds R, Parise KL, Drees KP, Hoyt JR, Cheng TL, Kunz TH, Foster JT & Kilpatrick AM (2015). 'Host and pathogen ecology drive the seasonal dynamics of a fungal disease, white-nose syndrome'. *Proc. R. Soc. B* **282**(1799), 20142335. DOI: 10.1098/rspb.2014.2335

Le Galliard JF, Rémy A, Ims RA & Lambin X (2012). 'Patterns and processes of dispersal behaviour in arvicoline rodents'. *Mol. Ecol.* **21**(3), 505–523. DOI: 10.1111/j.1365-294X.2011.05410.x

Lefebvre A, Fiet C, Belpois-Duchamp C, Tiv M, Astruc K & Glélé LA (2014). 'Case fatality rates of Ebola virus diseases: a meta-analysis of World Health Organization data'. *Med. Maladies. Infection.* **44**(9), 412–416. DOI: 10.1016/j.medmal.2014.08.005

Lehmann J, Korstjens AH & Dunbar RI (2010). 'Apes in a changing world—the effects of global warming on the behaviour and distribution of African apes'. *J. Beogeogr.* **37**(12), 2217–2231. DOI: 10.1111/j.1365-2699.2010.02373.x

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Délicat A, Paweska JT, Gonzalez JP & Swanepoel R (2005). 'Fruit bats as reservoirs of Ebola virus'. *Nature* **438**(7068), 575–576

Levin SA (1974). 'Dispersion and population interactions'. *American Naturalist* **108**(960), 207

Lewis T, Gillespie D, Lacey C, Matthews J, Danbolt M, Leaper R, McLanaghan R & Moscrop A (2007). 'Sperm whale abundance estimates from acoustic surveys of the Ionian Sea and Straits of Sicily in 2003'. *J. Mar. Biolog. Assoc. U.K.* **87**(01), 353–357. DOI: 10.1111/j.1748-7692.2005.tb01242.x

Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S & Wang LF (2005). 'Bats are natural reservoirs of SARS-like coronaviruses'. *Science* **310**(5748), 676–679. DOI: 10.1126/science.1118391

Li XZ, Liu JX & Martcheva M (2010). 'An age-structured two-strain epidemic model with super-infection'. *Math. Biosci. Eng.* **7**(1), 123–147

Lim BK & Engstrom MD (2001). 'Bat community structure at Iwokrama forest, Guyana'. *J. Trop. Ecol.* **17**(05), 647–665. DOI: 10.1017/s0266467401001481

Lindenfors P, Nunn CL, Jones KE, Cunningham AA, Sechrest W & Gittleman JL (2007). 'Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density'. *Global. Ecol. Biogeogr.* **16**(4), 496–509. DOI: 10.1111/j.1466-8238.2006.00301.x

Lloyd-Smith JO, Cross PC, Briggs CJ, Daugherty M, Getz WM, Latto J, Sanchez MS, Smith AB & Swei A (2005). 'Should we expect population thresholds for wildlife disease?' *Trends Ecol. Evol.* **20**(9), 511–519. DOI: 10.1016/j.tree.2005.07.004

López-Baucells A, Rocha R, Fernández-Arellano G, Bobrowiec PED, Palmeirim JM & Meyer CF (2014). 'Echolocation of the big red bat *Lasiurus egregius* (Chiroptera: Vespertilionidae) and first record from the Central Brazilian Amazon'. *Stud. Neotrop. Fauna Environ.* **49**(1), 18–25. DOI: 10.1080/01650521.2014.907600

López-Roig M, Bourhy H, Lavenir R & Serra-Cobo J (2014). 'Seroprevalence dynamics of European bat lyssavirus type 1 in a multispecies bat colony'. *Viruses* **6**(9), 3386–3399. DOI: 10.1371/journal.pone.0095610

Lu G, Lin A, Luo J, Blondel DV, Meiklejohn KA, Sun K & Feng J (2013). 'Phylogeography of the Rickett's big-footed bat, *Myotis pilosus* (Chiroptera: Vespertilionidae): a novel pattern of genetic structure of bats in China'. *BMC Evol. Biol.* **13**(1), 241

Luby SP, Hossain MJ, Gurley ES, Ahmed BN, Banu S, Khan SU, Homaira N, Rota PA, Rollin PE, Comer JA et al. (2009). 'Recurrent zoonotic transmission of Nipah virus into humans, Bangladesh, 2001–2007'. *Emerg. Infect. Dis.* **15**(8), 1229. DOI: 10.3201/eid1508.081237

Lucas TCD (2015a). 'MetapopEpi: Functions to run multipathogen, metapopulation epidemiological simulations'. R package version 0.0.1. DOI: 10.5281/zenodo.48942. URL: https://github.com/timcdlucas/metapopepi

Lucas TCD (2015b). *palettetown: Use Pokemon inspired colour palettes*. R package version 0.1.0. DOI: 10.5281/zenodo.48943. URL: http://CRAN.R-project.org/package=palettetown

Lucas TCD, Moorcroft EA, Freeman R, Rowcliffe JM & Jones KE (2015). 'A generalised random encounter model for estimating animal density with remote sensor data'. *Methods Ecol. Evol.* **6**(5), 500–509. DOI: 10.1111/2041-210X.12346

Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, Mills JN, Timonin ME, Willis CKR, Cunningham AA, Fooks AR, Rupprecht CE, Wood JLN & Webb CT (2013). 'A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?' *Proc. R. Soc. B* **280**(1756), 20122753. DOI: 10.1098/rspb.2012.2753

Ma J, Jones G, Zhang S, Shen J, Metzner W, Zhang L & Liang B (2003). 'Dietary analysis confirms that Rickett's big-footed bat (*Myotis ricketti*) is a piscivore'. *J. Zool.* **261**(03), 245–248. DOI: 10.1017/s095283690300414x

MacKenzie DI & Royle JA (2005). 'Designing occupancy studies: general advice and allocating survey effort'. *J. Appl. Ecol.* **42**(6), 1105–1114. DOI: 10.1111/j.1365-2664.2005.01098.x

Maganga GD, Bourgarel M, Vallo P, Dallo TD, Ngoagouni C, Drexler JF, Drosten C, Nakouné ER, Leroy EM & Morand S (2014). 'Bat distribution size or shape as determinant of viral richness in african bats'. *PLoS One* **9**(6), e100172. DOI: 10.1371/journal.pone.0100172

Manor R & Saltz D (2003). 'Impact of human nuisance disturbance on vigilance and group size of a social ungulate'. *Ecol. Appl.* **13**(6), 1830–1834. DOI: 10.1890/01-5354

Marceau V, Noël PA, Hébert-Dufresne L, Allard A & Dubé LJ (2011). 'Modeling the dynamical interaction between epidemics on overlay networks'. *Phys. Rev. E* **84**(2), 026105

Marcoux M, Auger-Méthé M, Chmelnitsky EG, Ferguson SH & Humphries MM (2011). 'Local passive acoustic monitoring of narwhal presence in the Canadian Arctic: a pilot project'. *Arctic* **64**(3), 307–316. DOI: 10.14430/arctic4121

Marjamäki PH, Contasti AL, Coulson TN & McLoughlin PD (2013). 'Local density and group size interacts with age and sex to determine direction and rate of social dispersal in a polygynous mammal'. *Ecol. Evol.* **3**(9), 3073–3082. DOI: 10.1002/ece3.694

Marques TA, Munger L, Thomas L, Wiggins S & Hildebrand JA (2011). 'Estimating North Pacific right whale *Eubalaena japonica* density using passive acoustic cue counting'. *Endanger. Species Res.* **13**(3), 163–172. DOI: 10.3354/esr00325

Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D & Tyack PL (2013). 'Estimating animal population density using passive acoustics'. *Biol. Rev.* **88**(2), 287–309. DOI: 10.1111/brv.12001

Martcheva M & Li XZ (2013). 'Competitive exclusion in an infection-age structured model with environmental transmission'. *J. Math. Anal. Appl.* **408**(1), 225–246. DOI: 10.1016/j.jmaa.2013.05.064

May RM & Anderson RM (1979). 'Population biology of infectious diseases: Part II'. *Nature* **280**(5722), 455–461. DOI: 10.1038/280455a0

May RM & Lloyd AL (2001). 'Infection dynamics on scale-free networks'. *Phys. Rev. E* **64**(6), 066112. DOI: 10.1103/PhysRevE.64.066112

May RM & Nowak MA (1994). 'Superinfection, metapopulation dynamics, and the evolution of diversity'. *J. Theor. Biol.* **170**(1), 95–114. DOI: 10.1006/jtbi.1994.1171

McCallum H, Barlow N & Hone J (2001). 'How should pathogen transmission be modelled?' *Trends Ecol. Evol.* **16**(6), 295–300. DOI: 10.1016/S0169-5347(01)02144-9

McCracken GF & Bradbury JW (1981). 'Social organization and kinship in the polygynous bat *Phyllostomus hastatus*'. *Behav. Ecol. Sociobiol.* **8**(1), 11–34. DOI: 10.1007/bf00302840

Metcalf C, Birger R, Funk S, Kouyos R, Lloyd-Smith J & Jansen V (2015). 'Five challenges in evolution and infectious diseases'. *Epidemics* **10**, 40–44. DOI: 10.1016/j.epidem.2014.12.003

Metheny JD, Kalcounis-Rueppell MC, Bondo KJ & Brigham RM (2008). 'A genetic analysis of group movement in an isolated population of tree-roosting bats'. *Proceedings of the Royal Society of London B: Biological Sciences* **275**(1648), 2265–2272. DOI: 10.1098/rspb.2008.0532

Miller-Butterworth CM, Jacobs DS & Harley EH (2003). 'Strong population substructure is correlated with morphology and ecology in a migratory bat'. *Nature* **424**(6945), 187–191. DOI: 10.1038/nature01742

Morand S (2000). 'Wormy world: comparative tests of theoretical hypotheses on parasite species richness'. In: *Evolutionary Biology of Host-Parasite Relationships*. Ed. by SM R. Poulin & A Skorping. Amsterdam: Elsevier, 63–79. DOI: 10.1017/S0031182001001135

Morand S & Poulin R (1998). 'Density, body mass and parasite species richness of terrestrial mammals'. *Evol. Ecol.* **12**(6), 717–727. DOI: 10.1023/A:1006537600093

Moreno-Valdez A, Honeycutt RL & Grant WE (2004). 'Colony dynamics of *Leptonycteris nivalis* (Mexican long-nosed bat) related to flowering Agave in northern Mexico'. *J. Mammal.* **85**(3), 453–459. DOI: 10.1644/1383942

Munywoki PK, Koech DC, Agoti CN, Kibirige N, Kipkoech J, Cane PA, Medley G & Nokes DJ (2015). 'Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial virus (RSV) shedding'. *Epidemiology and infection* **143**(04), 804–812. DOI: http://dx.doi.org/10.1017/S0950268814001393

Nowak MA & May RM (1994). 'Superinfection and the evolution of parasite virulence'. *Proc. R. Soc. B* **255**(1342), 81–89. DOI: 10.1098/rspb.1994.0012

Nunes A, da Gama MT & Gomes M (2006). 'Localized contacts between hosts reduce pathogen diversity'. *J. Theor. Biol.* **241**(3), 477–487. DOI: 10.1016/j.jtbi.2005.12.010

Nunn CL, Altizer S, Jones KE & Sechrest W (2003). 'Comparative tests of parasite species richness in primates'. *Am. Nat.* **162**(5), 597–614. DOI: 10.1086/378721

Nunn CL, Jordán F, McCabe CM, Verdolin JL & Fewell JH (2015). 'Infectious disease and group size: more than just a numbers game'. *Phil. Trans. R. Soc. B* **370**(1669), 20140111. DOI: 10.1098/rstb.2014.0111

O'Brien TG, Kinnaird MF & Wibisono HT (2003). 'Crouching tigers, hidden prey: Sumatran tiger and prey populations in a tropical forest landscape'. *Anim. Conserv.* **6**(2), 131–139. DOI: 10.1017/s1367943003003172

O'Donnell CF, Richter S, Dool S, Monks JM & Kerth G (2015). 'Genetic diversity is maintained in the endangered New Zealand long-tailed bat (*Chalinolobus tuberculatus*) despite a closed social structure and regular population crashes'. *Conserv. Genet.* 1–12. DOI: 10.1007/s10592-015-0746-9

Oleksy R, Racey PA & Jones G (2015). 'High-resolution GPS tracking reveals habitat selection and the potential for long-distance seed dispersal by Madagascan flying foxes *Pteropus rufus*'. *Glob. Ecol. Conserv.* **3**, 678–692. DOI: 10.1016/j.gecco.2015.02.012

Olival KJ, Weekley CC & Daszak P (2015). 'Are bats really "special" as viral reservoirs? What we know and need to know'. In: *Bats and Viruses: A New Frontier of Emerging Infectious Diseases*. Ed. by LF Wang & C Cowled. New Jersey: Wiley, 281–294. DOI: 10.1002/9781118818824.ch11

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N & Pearse W (2012). *caper: Comparative Analyses of Phylogenetics and Evolution in R*. R package version 0.5. URL: http://CRAN.R-project.org/package=caper

Orr TJ & Zuk M (2013). 'Does delayed fertilization facilitate sperm competition in bats?' *Behav. Ecol. Sociobiol.* **67**(12), 1903–1913. DOI: 10.1007/s00265-013-1598-2

O'Shea TJ, Cryan PM, Cunningham AA, Fooks AR, Hayman DT, Luis AD, Peel AJ, Plowright RK & Wood JL (2014). 'Bat flight and zoonotic viruses'. *Emerg. Infect. Dis.* **20**(5), 741. DOI: 10.3201/eid2005.130539

Owen SF, Menzel MA, Ford WM, Chapman BR, Miller KV, Edwards JW & Wood PB (2003). 'Home-range size and habitat used by the northern myotis (*Myotis septentrionalis*)'. *Am. Midl. Nat.* **150**(2), 352–359. DOI: 10.1674/0003-0031(2003)150[0352:hsahub]2.0.co;2

Paradis E, Claude J & Strimmer K (2004). 'APE: analyses of phylogenetics and evolution in R language'. *Bioinformatics* **20**, 289–290. DOI: 10.1093/bioinformatics/btg412

Pastor-Satorras R, Castellano C, Van Mieghem P & Vespignani A (2015). 'Epidemic processes in complex networks'. *Rev. Mod. Phys.* **87** (3), 925–979. DOI: 10.1103/RevModPhys.87.925. URL: http://link.aps.org/doi/10.1103/RevModPhys.87.925

Pastor-Satorras R & Vespignani A (2001). 'Epidemic spreading in scale-free networks'. *Phys. Rev. Lett.* **86**(14), 3200. DOI: 10.1103/PhysRevLett.86.3200

Peel A, Baker K, Crameri G, Barr J, Hayman D, Wright E, Broder C, Fernández-Loras A, Fooks A, Wang L et al. (2012). 'Henipavirus Neutralising Antibodies in an Isolated Island Population of African Fruit Bats'. *PLoS One* **7**(1), e30346

Peel AJ, Pulliam J, Luis A, Plowright R, O'Shea T, Hayman D, Wood J, Webb C & Restif O (2014). 'The effect of seasonal birth pulses on pathogen persistence

in wild mammal populations'. *Proc. R. Soc. B* **281**(1786), 20132962. DOI: 10.1098/rspb.2013.2962

Peel AJ, Sargan DR, Baker KS, Hayman DT, Barr JA, Crameri G, Suu-Ire R, Broder CC, Lembo T, Wang LF et al. (2013). 'Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses'. *Nat. Commun.* **4**. DOI: 10.1038/ncomms3770

Petit E & Mayer F (1999). 'Male dispersal in the noctule bat (*Nyctalus noctula*): where are the limits?' *Proc. R. Soc. B* **266**(1430), 1717–1722. DOI: 10.1098/rspb.1999.0837

Phillimore AB, Orme CDL, Davies RG, Hadfield JD, Reed WJ, Gaston KJ, Freckleton RP & Owens IP (2007). 'Biogeographical basis of recent phenotypic divergence among birds: a global study of subspecies richness'. *Evolution* **61**(4), 942–957. DOI: 10.1111/j.1558-5646.2007.00068.x

Pilosof S, Morand S, Krasnov BR & Nunn CL (2015). 'Potential parasite transmission in multi-host networks based on parasite sharing'. *PloS one* **10**(3), e0117909. DOI: 10.1371/journal.pone.0117909

Pinheiro J, Bates D, DebRoy S, Sarkar D & R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-122. DOI: 10.1371/journal.pone.0104012. URL: http://CRAN.R-project.org/package=nlme

Plowright RK, Foley P, Field HE, Dobson AP, Foley JE, Eby P & Daszak P (2011). 'Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (Pteropus spp.)' *Proc. R. Soc. B* **278**(1725), 3703–3712. DOI: 10.1098/rspb.2011.0522

Plowright RK, Eby P, Hudson PJ, Smith IL, Westcott D, Bryden WL, Middleton D, Reid PA, McFarlane RA, Martin G et al. (2015). 'Ecological dynamics of emerging bat virus spillover'. *Proc. R. Soc. B* **282**(1798), 20142124. DOI: 10.1098/rspb.2014.2124

Poletto C, Meloni S, Colizza V, Moreno Y & Vespignani A (2013). 'Host mobility drives pathogen competition in spatially structured populations'. *PLoS Comput. Biol.* **9**(8), e1003169. DOI: 10.1371/journal.pcbi.1003169

Poletto C, Meloni S, Van Metre A, Colizza V, Moreno Y & Vespignani A (2015). 'Characterising two-pathogen competition in spatially structured environments'. *Sci. Rep.* **5**. DOI: 10.1038/srep07895

Pons-Salort M, Serra-Cobo J, Jay F, Lopez-Roig M, Lavenir R, Guillemot D, Letort V, Bourhy H & Opatowski L (2014). 'Insights into persistence mechanisms of a zoonotic virus in bat colonies using a multispecies

metapopulation model'. *PLoS One* **9**(4), e95610. DOI: 10.1371/journal.pone.0095610

Popa-Lisseanu AG & Voigt CC (2009). 'Bats on the move'. *J. Mammal.* **90**(6), 1283–1289. DOI: http://dx.doi.org/10.1644/09-MAMM-S-130R2.1

Porter T & Wilkinson G (2001). 'Birth synchrony in greater spear-nosed bats (*Phyllostomus hastatus*)'. *J. Zool.* **253**(3), 383–390. DOI: 10.1017/S0952836901000358

Poulin R (1995). 'Phylogeny, ecology, and the richness of parasite communities in vertebrates'. *Ecol. Monogr.* 283–302. DOI: 10.2307/2937061

Poulin R (2010). 'Latitudinal gradients in parasite diversity: bridging the gap between temperate and tropical areas'. *Neotrop. Helminthol.* **4**(2), 169–177

Poulin R (2014). 'Parasite biodiversity revisited: frontiers and constraints'. *Int. J. Parasitol.* **44**(9), 581–589. DOI: 10.1016/j.ijpara.2014.02.003

Poulin R & Morand S (2000). 'The diversity of parasites'. *Q. Rev. Biol.* 277–293

Purvis A, Gittleman JL, Cowlishaw G & Mace GM (2000). 'Predicting extinction risk in declining species'. *Proc. R. Soc. B* **267**(1456), 1947–1952. DOI: 10.1098/rspb.2000.1234

Qiu Z, Kong Q, Li X & Martcheva M (2013). 'The vector–host epidemic model with multiple strains in a patchy environment'. *J. Math. Anal. Appl.* **405**(1), 12–36. DOI: 10.1016/j.jmaa.2013.03.042

R Development Core Team (2010). *R: A Language And Environment For Statistical Computing*. R Foundation For Statistical Computing. Vienna, Austria. URL: http://www.R-Project.org

Racey PA (2015). 'The uniqueness of bats'. In: *Bats and Viruses: A New Frontier of Emerging Infectious Diseases. Wiley Blackwell*. Ed. by LF Wang & C Cowled, 1–22. DOI: 10.1002/9781118818824.ch1

Rand D, Keeling M & Wilson H (1995). 'Invasion, stability and evolution to criticality in spatially extended, artificial host-pathogen ecologies'. *Proc. R. Soc. B* **259**(1354), 55–63. DOI: 10.1098/rspb.1995.0009

Revell LJ (2010). 'Phylogenetic signal and linear regression on species data'. *Methods Ecol. Evol.* **1**(4), 319–329. DOI: 10.1111/j.2041-210x.2010.00044.x

Richter H & Cumming G (2008). 'First application of satellite telemetry to track African straw-coloured fruit bat migration'. *J. Zool.* **275**(2), 172–176. DOI: 10.1111/j.1469-7998.2008.00425.x

Rifkin JL, Nunn CL & Garamszegi LZ (2012). 'Do animals living in larger groups experience greater parasitism? A meta-analysis'. *Am. Nat.* **180**(1), 70–82. DOI: 10.1086/666081

Rivera-Ortíz F, Aguilar R, Arizmendi M, Quesada M & Oyama K (2015). 'Habitat fragmentation and genetic variability of tetrapod populations'. *Anim. Conserv.* **18**(3), 249–258. DOI: 10.1111/acv.12165

Rogers TL, Ciaglia MB, Klinck H & Southwell C (2013). 'Density can be misleading for low-density species: benefits of passive acoustic monitoring'. *PLoS One* **8**(1), e52542. DOI: 10.1371/journal.pone.0052542

Rohani P, Green C, Mantilla-Beniers N & Grenfell B (2003). 'Ecological interference between fatal diseases'. *Nature* **422**(6934), 885–888. DOI: 10.1038/nature01542

Rohde K (1992). 'Latitudinal gradients in species diversity: the search for the primary cause'. *Oikos*, 514–527. DOI: 10.2307/3545569

Rovero F, Zimmermann F, Berzi D & Meek P (2013). '"Which camera trap type and how many do I need?" A review of camera features and study designs for a range of wildlife research applications.' *Hystrix* **24**(2), 148–156. DOI: 10.4404/hystrix-24.2-8789

Rowcliffe JM & Carbone C (2008). 'Surveys using camera traps: are we looking to a brighter future?' *Anim. Conserv.* **11**(3), 185–186

Rowcliffe J, Field J, Turvey S & Carbone C (2008). 'Estimating animal density using camera traps without the need for individual recognition'. *J. Appl. Ecol.* **45**(4), 1228–1236. DOI: 10.1111/j.1365-2664.2008.01473.x

Royle JA & Nichols JD (2003). 'Estimating abundance from repeated presence-absence data or point counts'. *Ecology* **84**(3), 777–790. DOI: 10.1890/0012-9658(2003)084[0777:eafrpa]2.0.co;2

Sabol BM & Hudson MK (1995). 'Technique using thermal infrared-imaging for estimating populations of gray bats'. *J. Mammal.* **76**(4), 1242–1248. DOI: 10.2307/1382618

Schielzeth H (2010). 'Simple means to improve the interpretability of regression coefficients'. *Methods Ecol. Evol.* **1**(2), 103–113. DOI: 10.1111/j.2041-210x.2010.00012.x

Schmidt BR (2003). 'Count data, detection probabilities, and the demography, dynamics, distribution, and decline of amphibians'. *C. R. Biol.* **326**, 119–124. DOI: 10.1016/s1631-0691(03)00048-9

Schooley RL & Branch LC (2009). 'Enhancing the area-isolation paradigm: habitat heterogeneity and metapopulation dynamics of a rare wetland mammal'. *Ecol. Appl.* **19**(7), 1708–1722. DOI: 10.1890/08-2169.1

Scott M & Lewis J (1987). 'Population dynamics of helminth parasites in wild and laboratory rodents'. *Mammal Review* **17**(2-3), 95–103

Serra-Cobo J, Amengual B, Abellán C & Bourhy H (2002). 'European bat lyssavirus infection in Spanish bat populations'. *Emerg. Infect. Dis.* **8**(4), 413–420. DOI: 10.3201/eid0804.010263

Shump KA & Shump AU (1982). '*Lasiurus borealis*'. *Mammalian species*, 1–6. DOI: 10.2307/3503843

Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C & Ramachandran S (2014). 'Global rise in human infectious disease outbreaks'. *J. R. Soc. Interface* **11**(101), 20140950. DOI: 10.1098/rsif.2014.0950

Smouse PE, Focardi S, Moorcroft PR, Kie JG, Forester JD & Morales JM (2010). 'Stochastic modelling of animal movement'. *Phil. Trans. R. Soc. B* **365**(1550), 2201–2211. DOI: 10.1098/rstb.2010.0078

Speakman J, Racey P, Catto C, Webb P, Swift S & Burnett A (1991). 'Minimum summer populations and densities of bats in NE Scotland, near the northern borders of their distributions'. *J. Zool.* **225**(2), 327–345. DOI: 10.1111/j.1469-7998.1991.tb03820.x

SymPy Development Team (2014). *SymPy: Python library for symbolic mathematics*. URL: http://www.sympy.org

Taylor LH, Latham SM & Mark E (2001). 'Risk factors for human disease emergence'. *Phil. Trans. R. Soc. B* **356**(1411), 983–989. DOI: 10.1093/acprof:oso/9780198525738.003.0035

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BF, De Siqueira MF, Grainger A, Hannah L et al. (2004). 'Extinction risk from climate change'. *Nature* **427**(6970), 145–148. DOI: 10.1038/nature02121

Thomas L & Marques TA (2012). 'Passive acoustic monitoring for estimating animal density'. *Acoustics Today* **8**(3), 35–44. DOI: 10.1121/1.4753915

Turmelle AS & Olival KJ (2009). 'Correlates of viral richness in bats (order Chiroptera)'. *EcoHealth* **6**(4), 522–539. DOI: 10.1007/s10393-009-0263-8

Turner K & Garnett G (2002). 'The impact of the phase of an epidemic of sexually transmitted infection on the evolution of the organism'. *Sex. Transm. Infect.* **78**, i20–i30. DOI: 10.1136/sti.78.suppl_1.i20

Van de Bovenkamp R, Kuipers F & Van Mieghem P (2014). 'Domination-time dynamics in susceptible-infected-susceptible virus competition on networks'. *Phys. Rev. E* **89**(042818), 042818

Vespignani A (2008). 'Reaction-diffusion processes and epidemic metapopulation models in complex networks'. *Eur. Phys. J. B* **64**(3), 349–353. DOI: 10.1140/epjb/e2008-00302-y

Vitone ND, Altizer S & Nunn CL (2004). 'Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates'. *Evol. Ecol. Res.* **6**(2), 183–199

Vögeli M, Lemus JA, Serrano D, Blanco G & Tella JL (2011). 'An island paradigm on the mainland: host population fragmentation impairs the community of avian pathogens'. *Proc. R. Soc. B* **278**(1718), 2668–2676. DOI: 10.1098/rspb.2010.1227

Vonhof MJ, Russell AL & Miller-Butterworth CM (2015). 'Range-wide genetic analysis of little brown bat (*Myotis lucifugus*) populations: Estimating the risk of spread of white-nose syndrome'. *PLoS One* **10**(7). DOI: 10.1371/journal.pone.0128713

Vonhof M & Russell AL. 'Genetic approaches to the conservation of migratory bats: a case study of the eastern red bat (*Lasiurus borealis*)'

Walters CL, Collen A, Lucas T, Mroz K, Sayer CA & Jones KE (2013). 'Challenges of using bioacoustics to globally monitor bats'. In: *Bat Evolution, Ecology, and Conservation*. Ed. by RA Adams & SC Pedersen. Springer, 479–499. DOI: 10.1007/978-1-4614-7397-8_23

Walters CL, Freeman R, Collen A, Dietz C, Brock Fenton M, Jones G, Obrist MK, Puechmaille SJ, Sattler T, Siemers BM, Parsons S & Jones KE (2012). 'A continental-scale tool for acoustic identification of European bats'. *J. Appl. Ecol.* **49**(5), 1064–1074. DOI: 10.1111/j.1365-2664.2012.02182.x

Wang LF, Walker P & Poon L (2011). 'Mass extinctions, biodiversity and mitochondrial function: are bats 'special' as reservoirs for emerging viruses?' *Curr. Opin. Virol.* DOI: 10.1016/j.coviro.2011.10.013

Wang W, Chen M, Min Y & Jin X (2016). 'Structural diversity effects of multilayer networks on the threshold of interacting epidemics'. *Physica A* **443**, 254–262. DOI: 10.1016/j.physa.2015.09.064

Wardeh M, Risley C, McIntyre MK, Setzkorn C & Baylis M (2015). 'Database of host-pathogen and related species interactions, and their global distribution'. *Sci. Data* **2**. DOI: 10.1038/sdata.2015.49

Wasserman TN, Cushman SA, Littell JS, Shirk AJ & Landguth EL (2013). 'Population connectivity and genetic diversity of American marten (*Martes americana*) in the United States northern Rocky Mountains in a climate change context'. *Conserv. Genet.* **14**(2), 529–541. DOI: 10.1007/s10592-012-0336-z

Webb N & Tidemann C (1996). 'Mobility of Australian flying-foxes, Pteropus spp.(Megachiroptera): evidence from genetic variation'. *Proc. R. Soc. B* **263**(1369), 497–502

Webber QM, Brigham RM, Park AD, Gillam EH, O'Shea TJ & Willis CK (2016). 'Social network characteristics and predicted pathogen transmission in summer colonies of female big brown bats (*Eptesicus fuscus*)'. *Behav. Ecol. Sociobiol.* 1–12. DOI: 10.1007/s00265-016-2093-3

Whittingham MJ, Stephens PA, Bradbury RB & Freckleton RP (2006). 'Why do we still use stepwise modelling in ecology and behaviour?' *J. Anim. Ecol.* **75**(5), 1182–1189. DOI: 10.1111/j.1365-2656.2006.01141.x

Whittingham MJ, Swetnam RD, Wilson JD, Chamberlain DE & Freckleton RP (2005). 'Habitat selection by yellowhammers *Emberiza citrinella* on lowland farmland at two spatial scales: implications for conservation management'. *J. Appl. Ecol.* **42**(2), 270–280. DOI: 10.1111/j.1365-2664.2005.01007.x

Wickham H (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. ISBN: 978-0-387-98140-6. URL: http://had.co.nz/ggplot2/book

Wickham H (2015). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 0.2.0. URL: http://CRAN.R-project.org/package=rvest

Wilke CO (2015). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 0.6.0. URL: https://CRAN.R-project.org/package=cowplot

Wilmer JW, Moritz C, Hall L & Toop J (1994). 'Extreme population structuring in the threatened ghost bat, Macroderma gigas: evidence from mitochondrial DNA'. *Proc. R. Soc. B* **257**(1349), 193–198. DOI: 10.1098/rspb.1994.0115

Wilmer JW, Hall L, Barratt E & Moritz C (1999). 'Genetic structure and male-mediated gene flow in the ghost bat (*Macroderma gigas*)'. *Evolution*, 1582–1591

Wilson DE & Reeder DM (2005). *Mammal species of the world: a taxonomic and geographic reference*. Vol. 12. JHU Press. ISBN: 9780801882210

Witsenburg F, Clément L, López-Baucells A, Palmeirim J, Pavlinić I, Scaravelli D, Ševčík M, Dutoit L, Salamin N, Goudet J et al. (2015). 'How a haemosporidian parasite of bats gets around: the genetic structure of a parasite, vector and host compared'. *Mol. Ecol.* **24**(4), 926–940

Wolfe ND, Eitel MN, Gockowski J, Muchaal PK, Nolte C, Tassy Prosser A, Ndongo Torimiro J, Weise SF & Burke DS (2000). 'Deforestation, hunting and the ecology of microbial emergence'. *Global Change & Human Health* **1**(1), 10–25. DOI: 10.1023/A:1011519513354

Woolhouse ME & Gowtage-Sequeria S (2006). 'Host range and emerging and reemerging pathogens'. In: *Ending the War Metaphor: The Changing Agenda for Unraveling the Host-Microbe Relationship: Workshop Summary*. Vol. 192. National Academies Press

World Bank (2014). *Update on the economic impact of the 2014 Ebola epidemic on Liberia, Sierra Leone, and Guinea*. Washington, DC. DOI: 10.1211/pj.2014.20066832. URL: http://documents.worldbank.org/curated/en/2014/12/20454884/update-economic-impact-2014-ebola-epidemic-liberia-sierra-leone-guinea

World Bank (2015). 'World Development Indicators: Size of the economy'. URL: http://wdi.worldbank.org/table/1.1

Wright SJ & Hubbell SP (1983). 'Stochastic extinction and reserve size: a focal species approach'. *Oikos*, 466–476. DOI: 10.2307/3544106

Wu Q, Zhang H, Small M & Fu X (2013). 'Threshold analysis of the susceptible-infected-susceptible model on overlay networks'. *Commun. Nonlinear Sci. Numer. Simul.*

Xie Y (2015). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.11. URL: http://CRAN.R-project.org/package=knitr

Yapp W (1956). 'The theory of line transects'. *Bird study* **3**(2), 93–104. DOI: 10.1080/00063655609475840

Yu G (2015). 'ggtree: an R package for versatile annotation and visualization of phylogenetic tree'

Zahn A, Rottenwallner A & Güttinger R (2006). 'Population density of the greater mouse-eared bat (*Myotis myotis*), local diet composition and availability of foraging habitats'. *J. Zool.* **269**(4), 486–493. DOI: 10.1111/j.1469-7998.2006.00081.x

Zero VH, Sundaresan SR, O'Brien TG & Kinnaird MF (2013). 'Monitoring an endangered savannah ungulate, Grevy's zebra (*Equus grevyi*): choosing a method for estimating population densities'. *Oryx* **47**(03), 410–419. DOI: 10.1017/s0030605312000324

Zhang L, Jones G, Zhang J, Zhu G, Parsons S, Rossiter SJ & Zhang S (2009). 'Recent surveys of bats (Mammalia: Chiroptera) from China. I.

Rhinolophidae and Hipposideridae'. *Acta Chiropt.* **11**(1), 71–88. DOI: 10.3161/150811010x504626

Zunino GE, Kowalewski MM, Oklander LI & Gonzalez V (2007). 'Habitat fragmentation and population size of the black and gold howler monkey (*Alouatta caraya*) in a semideciduous forest in northern Argentina'. *Am. J. Primatol.* **69**(9), 966–975. DOI: 10.1002/ajp.20389

A Appendix: A comparative test of the role of population structure in determining pathogen richness

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Megadermatidae	*Macroderma gigas*	0		0.25	124.37	1078.95	769	13	3609.23	3148	(Wilmer et al. 1999)
Miniopteridae	*Miniopterus natalensis*	0		0.08	10.43	3784.83	180	5	6657.28	1706	(Miller-Butterworth et al. 2003)
Miniopteridae	*Miniopterus schreibersii*	12		1.86	11.46	3707.97	3090	64	7050.77	2649	(Witsenburg et al. 2015)
Phyllostomidae	*Macrotus waterhousii*	0		2.33	16.27	802.74	594	15	4298.38	935	(Burns & Broders 2014)
Pteropodidae	*Rousettus obliviosus*	0		13.76	45.32	1.67	52	0	174.92	160	(Goodman et al. 2010)
Pteropodidae	*Thoopterus nigrescens*	0		0.14	66.12	182.87	52	0	1388.27	620	(Burns & Broders 2014)
Vespertilionidae	*Myotis ciliolabrum*	0		12.38	4.89	1388.27	592	5	2150.93	473	(Burns & Broders 2014)
Vespertilionidae	*Myotis macropus*	0		0.44	9.8	1326.85	214	0	3644.70	883	(Burns & Broders 2014)
Hipposideridae	*Hipposideros speoris*	2	1	0.07	10.39	1225.28	603	9	2801.74	1175	(Chinnasamy et al. 2013)
Phyllostomidae	*Desmodus rotundus*	7	1	1.29	33.16	17726.52	6810	265	9314.10	2252	(Burns & Broders 2014)
Phyllostomidae	*Macrotus californicus*	1	1	1.26	11.83	643.20	808	15	2139.83	590	(Burns & Broders 2014)
Pteropodidae	*Cynopterus sphinx*	3	6	5.08	44.71	6456.51	1620	53	6821.48	3915	(Burns & Broders 2014)
Pteropodidae	*Pteropus alecto*	6	4	5.31	610.13	1354.21	1530	49	5064.99	2961	(Webb & Tidemann 1996)
Pteropodidae	*Pteropus conspicillatus*	2	2	2.92	760.71	219.53	465	15	3294.68	993	(Fox 2006)
Pteropodidae	*Pteropus poliocephalus*	3	1	8.80	702.78	249.32	1950	54	1844.19	721	(Webb & Tidemann 1996)
Pteropodidae	*Pteropus scapulatus*	7	1	4.34	380.35	3038.11	707	19	4053.21	2625	(Burns & Broders 2014)
Pteropodidae	*Rousettus leschenaultii*	11	3	17.73	84.88	6764.79	848	0	6795.10	3828	(Burns & Broders 2014)
Pteropodidae	*Rousettus madagascariensis*	1	1	31.12	65.72	292.74	135	2	1483.04	1366	(Burns & Broders 2014)
Rhinolophidae	*Rhinolophus ferrumequinum*	9	7	2.78	22.59	9747.14	5070	85	12178.06	9670	(Burns & Broders 2014)
Vespertilionidae	*Lasiurus borealis*	1	1	19.61	12.33	4881.50	3460	28	4185.94	2198	(Vonhof & Russell n.d.)
Vespertilionidae	*Myotis lucifugus*	8	5	7.64	7.8	12040.88	9720	298	6671.21	5173	(Vonhof et al. 2015)
Vespertilionidae	*Myotis myotis*	6	2	3.45	25.59	3874.58	8880	137	4377.75	2786	(Burns & Broders 2014)
Vespertilionidae	*Myotis ricketti*	3	1	2.38	22.5	987.18	421	11	2741.15	2272.57	(Lu et al. 2013)
Vespertilionidae	*Nyctalus noctula*	8	4	20.71	28.48	8033.97	4670	37	10238.26	4015	(Burns & Broders 2014)
Emballonuridae	*Rhynchonycteris naso*	1	1		4.14	11508.83	607	2			
Emballonuridae	*Saccolaimus flaviventris*	1	1		45.25	6053.68	342	3			
Emballonuridae	*Taphozous melanopogon*	4	5		25.99	5624.37	503	15			
Emballonuridae	*Taphozous perforatus*	2	4		24.43	2114.92	264	5			
Hipposideridae	*Aselliscus stoliczkanus*	1	1		6.09	1362.32	128	9			
Hipposideridae	*Hipposideros armiger*	5	4		49.99	4475.41	656	26			
Hipposideridae	*Hipposideros bicolor*	1	7		8.39	1584.43	371	4			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Hipposideridae	*Hipposideros caffer*	3	4		9.46	6348.39	458	6			
Hipposideridae	*Hipposideros cineraceus*	1	2		3.84	2203.63	164	0			
Hipposideridae	*Hipposideros commersoni*	1	1		89.99	513.16	429	0			
Hipposideridae	*Hipposideros diadema*	2	15		46.9	3318.84	472	4			
Hipposideridae	*Hipposideros lankadiva*	1	1		44.76	712.59	188	2			
Hipposideridae	*Hipposideros larvatus*	2	5		19.95	3287.88	398	7			
Hipposideridae	*Hipposideros pomona*	2	3		6.2	2840.90	226	5			
Megadermatidae	*Megaderma lyra*	2	2		39.27	6042.09	1370	32			
Molossidae	*Chaerephon plicatus*	2	5		21.83	2462.34	62	0			
Molossidae	*Chaerephon pumilus*	4	1		10.98	7240.27	177	0			
Molossidae	*Cynomops abrasus*	1	4		35.39	10671.17	143	2			
Molossidae	*Cynomops planirostris*	1	1		12.84	11450.26	138	0			
Molossidae	*Eumops auripendulus*	1	2		28.49	14106.77	343	2			
Molossidae	*Eumops glaucinus*	1	2		36.2	11385.12	545	5			
Molossidae	*Eumops perotis*	1	3		50.97	11695.63	859	5			
Molossidae	*Molossops neglectus*	1	1		11	6515.61	166	0			
Molossidae	*Molossus molossus*	7	7		13.7	15767.72	2220	68			
Molossidae	*Molossus rufus*	9	1		120	14800.28	629	11			
Molossidae	*Mops condylurus*	3	4		26.59	9123.92	310	5			
Molossidae	*Nyctinomops laticaudatus*	1	5		13.12	13101.23	511	11			
Molossidae	*Nyctinomops macrotis*	1	1		16.38	15743.81	641	8			
Molossidae	*Tadarida brasiliensis*	5	9		12.61	13833.40	5400	131			
Molossidae	*Tadarida teniotis*	1	2		28.07	3743.23	1160	7			
Mormoopidae	*Mormoops megalophylla*	3	4		16.09	3738.69	806	9			
Mormoopidae	*Pteronotus davyi*	5	3		9.52	3471.86	680	2			
Mormoopidae	*Pteronotus parnellii*	7	9		19.59	8592.98	2400	128			
Natalidae	*Natalus stramineus*	1	6		5.68	4.47	675	4			
Natalidae	*Natalus tumidirostris*	4	3		6.3	1593.55	19	0			
Nycteridae	*Nycteris gambiensis*	3	1		7.13	1860.95	62	0			
Nycteridae	*Nycteris nana*	1	1		6.98	2938.13	42	0			
Nycteridae	*Nycteris thebaica*	1	8		9.2	14397.36	625	7			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Phyllostomidae	*Anoura caudifer*	1	1		10.81	8856.50	177	0			
Phyllostomidae	*Anoura geoffroyi*	9	3		15.15	7791.08	1140	11			
Phyllostomidae	*Artibeus cinereus*	3	1		12.7	3466.55	463	3			
Phyllostomidae	*Artibeus fimbriatus*	1	1		63.89	1480.43	477	4			
Phyllostomidae	*Artibeus jamaicensis*	12	13		43.63	1932.23	3950	68			
Phyllostomidae	*Artibeus lituratus*	13	3		59.3	14235.95	2800	64			
Phyllostomidae	*Artibeus phaeotis*	5	4		11.69	3737.37	354	0			
Phyllostomidae	*Carollia brevicauda*	2	1		14.85	10749.19	22	0			
Phyllostomidae	*Carollia perspicillata*	15	1		19.23	13796.36	4040	126			
Phyllostomidae	*Carollia subrufa*	2	1		15.84	212.26	230	0			
Phyllostomidae	*Chrotopterus auritus*	1	1		78.26	13116.79	955	4			
Phyllostomidae	*Diaemus youngi*	1	1		36.71	13059.18	725	11			
Phyllostomidae	*Diphylla ecaudata*	1	1		28.11	8092.44	1010	13			
Phyllostomidae	*Glossophaga commissarisi*	1	3		9.15	3716.73	406	7			
Phyllostomidae	*Glossophaga morenoi*	1	3		8.54	327.07	146	0			
Phyllostomidae	*Glossophaga soricina*	8	5		9.97	14858.41	3460	90			
Phyllostomidae	*Leptonycteris curasoae*	1	1		25.27	840.36	1280	19			
Phyllostomidae	*Leptonycteris nivalis*	1	1		24.26	949.99	678	5			
Phyllostomidae	*Lonchophylla robusta*	1	1		13.72	1116.13	248	0			
Phyllostomidae	*Lonchophylla thomasi*	1	1		7.09	8383.35	282	0			
Phyllostomidae	*Lonchorhina aurita*	1	2		15.38	11577.00	509	3			
Phyllostomidae	*Lophostoma brasiliense*	1	1		9.76	8095.48	226	0			
Phyllostomidae	*Lophostoma silvicolum*	2	4		32.29	12138.52	398	2			
Phyllostomidae	*Micronycteris megalotis*	1	1		6.4	12094.01	806	0			
Phyllostomidae	*Mimon crenulatum*	1	4		13.91	11136.33	596	3			
Phyllostomidae	*Phyllostomus discolor*	2	2		36.7	12498.39	1610	54			
Phyllostomidae	*Phyllostomus hastatus*	13	2		91.44	12629.83	2380	38			
Phyllostomidae	*Platyrrhinus helleri*	4	2		13.44	11318.77	523	2			
Phyllostomidae	*Platyrrhinus lineatus*	1	2		24.34	5807.57	685	12			
Phyllostomidae	*Sturnira lilium*	7	8		20.19	15289.33	2440	36			
Phyllostomidae	*Sturnira ludovici*	1	3		21	2234.37	493	2			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Phyllostomidae	*Tonatia bidens*	1	1		27.7	3599.57	508	0			
Phyllostomidae	*Trachops cirrhosus*	1	3		36.9	12642.87	1230	16			
Phyllostomidae	*Uroderma bilobatum*	4	3		16.28	12796.60	1210	6			
Phyllostomidae	*Vampyrodes caraccioli*	1	2		35.89	8345.54	344	0			
Pteropodidae	*Cynopterus brachyotis*	5	8		33.87	2699.82	1120	33			
Pteropodidae	*Dobsonia anderseni*	1	1		233.99	46.53	30	0			
Pteropodidae	*Dobsonia moluccensis*	1	1		447.64	878.25	168	2			
Pteropodidae	*Eidolon dupreanum*	3	1		297.58	460.03	159	5			
Pteropodidae	*Eidolon helvum*	7	3		254.61	11803.17	1640	68			
Pteropodidae	*Eonycteris spelaea*	5	4		58.7	3525.63	1030	11			
Pteropodidae	*Epomophorus gambianus*	1	2		134.59	3791.20	295	4			
Pteropodidae	*Epomophorus wahlbergi*	1	1		93.59	4999.93	511	13			
Pteropodidae	*Epomops dobsonii*	1	1		122.08	1367.02	22	0			
Pteropodidae	*Epomops franqueti*	2	1		118.99	4539.60	515	7			
Pteropodidae	*Hypsignathus monstrosus*	2	1		336.97	2860.59	697	9			
Pteropodidae	*Macroglossus minimus*	1	4		16.3	3590.25	551	5			
Pteropodidae	*Micropteropus pusillus*	3	1		25.38	5380.92	313	4			
Pteropodidae	*Myonycteris torquata*	1	1		44.92	4524.08	442	5			
Pteropodidae	*Pteropus admiralitatum*	1	4		306.46	75.83	39	0			
Pteropodidae	*Pteropus dasymallus*	1	5		491.86	38.78	350	3			
Pteropodidae	*Pteropus giganteus*	2	4		824.85	4003.76	1510	49			
Pteropodidae	*Pteropus hypomelanus*	4	16		435.61	524.21	591	28			
Pteropodidae	*Pteropus lylei*	5	1		319.75	106.83	346	9			
Pteropodidae	*Pteropus neohibernicus*	1	2		1017.37	672.55	85	0			
Pteropodidae	*Pteropus rufus*	2	1		792	185.37	310	9			
Pteropodidae	*Pteropus vampyrus*	1	6		1027.54	1940.63	1340	40			
Pteropodidae	*Rousettus aegyptiacus*	8	6		134	3888.09	2680	121			
Pteropodidae	*Rousettus amplexicaudatus*	1	5		74.37	4281.50	436	3			
Pteropodidae	*Syconycteris australis*	1	7		17.6	1054.00	618	4			
Rhinolophidae	*Rhinolophus affinis*	1	9		13.7	5505.76	457	8			
Rhinolophidae	*Rhinolophus blasii*	2	4		10.29	3410.28	469	3			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Rhinolophidae	*Rhinolophus eloquens*	2	2		19.15	711.92	59	0			
Rhinolophidae	*Rhinolophus euryale*	2	2		9.25	2989.98	1300	3			
Rhinolophidae	*Rhinolophus hildebrandtii*	1	1		25.99	3184.50	106	2			
Rhinolophidae	*Rhinolophus hipposideros*	1	6		4.57	6256.45	3450	19			
Rhinolophidae	*Rhinolophus lepidus*	1	5		5.46	3509.10	203	2			
Rhinolophidae	*Rhinolophus macrotis*	3	6		6.18	2338.40	202	3			
Rhinolophidae	*Rhinolophus mehelyi*	1	2		14.03	1978.08	747	5			
Rhinolophidae	*Rhinolophus pearsonii*	3	2		11.55	3176.33	83	0			
Rhinolophidae	*Rhinolophus pusillus*	3	9		5.15	3970.82	259	9			
Rhinolophidae	*Rhinolophus rouxii*	3	2		12.25	864.10	159	0			
Rhinolophidae	*Rhinolophus sinicus*	3	2		11.4	2169.00	256	15			
Vespertilionidae	*Antrozous pallidus*	1	7		22.24	4367.60	2650	28			
Vespertilionidae	*Barbastella barbastellus*	1	2		8.31	3803.51	2540	11			
Vespertilionidae	*Corynorhinus townsendii*	1	5		10.3	4596.89	2250	3			
Vespertilionidae	*Eptesicus brasiliensis*	1	4		9.2	12912.90	600	2			
Vespertilionidae	*Eptesicus diminutus*	1	2		5.99	2385.96	207	0			
Vespertilionidae	*Eptesicus furinalis*	1	4		7.7	16292.71	726	6			
Vespertilionidae	*Eptesicus fuscus*	7	12		17.49	13169.33	9780	405			
Vespertilionidae	*Eptesicus serotinus*	4	10		23.09	12118.86	3980	56			
Vespertilionidae	*Euderma maculatum*	1	1		16.17	2104.99	1110	5			
Vespertilionidae	*Glauconycteris argentata*	1	1		9.3	3996.26	34	0			
Vespertilionidae	*Histiotus montanus*	1	3		12.5	4225.09	410	4			
Vespertilionidae	*Histiotus velatus*	1	1		11.32	2635.15	320	5			
Vespertilionidae	*Ia io*	1	1		49.3	1506.49	6810	2			
Vespertilionidae	*Lasionycteris noctivagans*	1	1		11.02	10110.81	2480	33			
Vespertilionidae	*Lasiurus blossevillii*	1	4		22	19046.13	1010	2			
Vespertilionidae	*Lasiurus cinereus*	2	3		27.06	21071.17	3800	35			
Vespertilionidae	*Lasiurus ega*	1	5		12.2	15726.17	751	6			
Vespertilionidae	*Lasiurus egregius*	1	1		14.2	1322.64	92	0			
Vespertilionidae	*Lasiurus intermedius*	1	2		22.96	1748.79	471	0			
Vespertilionidae	*Lasiurus seminolus*	1	1		9.88	1193.52	384	0			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Vespertilionidae	*Lasiurus xanthinus*	1	1		7.18	1737.40	177	0			
Vespertilionidae	*Miniopterus inflatus*	1	2		14.9	2349.94	164	4			
Vespertilionidae	*Miniopterus magnater*	2	2		14.14	3160.92	133	5			
Vespertilionidae	*Miniopterus pusillus*	3	1		8.95	1106.94	122	4			
Vespertilionidae	*Miniopterus schreibersii*	12	16		11.46	3707.97	3090	64			
Vespertilionidae	*Miniopterus tristis*	1	5		15.17	733.06	48	0			
Vespertilionidae	*Murina leucogaster*	2	2		7.54	616.31	447	12			
Vespertilionidae	*Myotis albescens*	1	1		5.69	14595.01	545	4			
Vespertilionidae	*Myotis austroriparius*	1	1		7.35	832.90	751	5			
Vespertilionidae	*Myotis blythii*	2	4		23.82	6194.60	1500	6			
Vespertilionidae	*Myotis californicus*	2	4		4.39	3984.79	1290	6			
Vespertilionidae	*Myotis chinensis*	1	1		41.99	2263.22	141	3			
Vespertilionidae	*Myotis dasycneme*	2	1		15.16	5424.34	1750	10			
Vespertilionidae	*Myotis daubentonii*	2	7		7.63	14654.91	3670	53			
Vespertilionidae	*Myotis evotis*	2	6		6.91	3163.30	1410	5			
Vespertilionidae	*Myotis grisescens*	1	1		10.84	853.02	1520	7			
Vespertilionidae	*Myotis horsfieldii*	1	5		6.05	2562.28	103	0			
Vespertilionidae	*Myotis leibii*	1	1		5.22	1423.17	725	0			
Vespertilionidae	*Myotis levis*	1	2		5.49	1393.86	259	3			
Vespertilionidae	*Myotis macrodactylus*	1	3		7.48	682.55	319	6			
Vespertilionidae	*Myotis mystacinus*	2	3		5.9	4972.35	2530	15			
Vespertilionidae	*Myotis nattereri*	2	2		7.25	5867.92	2940	27			
Vespertilionidae	*Myotis nigricans*	3	4		11	14380.23	1710	23			
Vespertilionidae	*Myotis occultus*	1	1		8.79	996.79	193	0			
Vespertilionidae	*Myotis riparius*	1	1		4.57	14110.14	440	0			
Vespertilionidae	*Myotis septentrionalis*	4	1		7.82	4946.39	1520	13			
Vespertilionidae	*Myotis thysanodes*	1	4		8.49	3472.25	1260	8			
Vespertilionidae	*Myotis velifer*	1	5		9.82	1953.95	1720	19			
Vespertilionidae	*Myotis volans*	2	4		8.71	4711.97	1080	7			
Vespertilionidae	*Myotis yumanensis*	1	6		5.15	4063.52	1230	7			
Vespertilionidae	*Nyctalus leisleri*	1	2		12.47	5170.69	2220	9			

Table A.1 Raw data for both analyses. Range Length is the distance between furthest apart points in the species range. Dmax is the distance between furthest apart F_{ST} sampling locations. The references are for the F_{ST} data only.

Family	Binomial	Virus Sp.	Subsp.	Gene Flow	Mass (g)	Range Size ($\times 10^9 km^2$)	Scholar	PubMed	Range Length (km)	Dmax (km)	Reference
Vespertilionidae	*Nyctalus plancyi*	1	2		19.7	1917.71	75	6			
Vespertilionidae	*Nycticeius humeralis*	1	3		9.12	2883.26	1440	9			
Vespertilionidae	*Philetor brachypterus*	1	1		12	1317.15	84	0			
Vespertilionidae	*Pipistrellus abramus*	3	1		5.87	3842.25	859	22			
Vespertilionidae	*Pipistrellus ceylonicus*	1	7		8.05	2290.09	175	0			
Vespertilionidae	*Pipistrellus hesperus*	1	2		3.56	2655.91	950	0			
Vespertilionidae	*Pipistrellus kuhlii*	1	3		6.07	10723.12	1360	12			
Vespertilionidae	*Pipistrellus nathusii*	3	1		7.44	5886.19	2250	11			
Vespertilionidae	*Pipistrellus pipistrellus*	6	2		5.3	11497.40	6420	83			
Vespertilionidae	*Pipistrellus pygmaeus*	1	1		5.3	1954.26	1880	12			
Vespertilionidae	*Pipistrellus subflavus*	1	4		5.74	4264.01	1680	0			
Vespertilionidae	*Plecotus auritus*	3	5		8.19	6533.49	5430	34			
Vespertilionidae	*Rhogeessa parvula*	1	1		4.37	448.91	147	0			
Vespertilionidae	*Scotophilus kuhlii*	7	7		20.31	5226.24	377	5			
Vespertilionidae	*Scotophilus nigrita*	2	2		27.34	2195.99	142	0			
Vespertilionidae	*Tylonycteris pachypus*	2	5		4.1	4170.18	428	8			
Vespertilionidae	*Vespertilio murinus*	2	2		15.42	15732.73	2280	9			
Vespertilionidae	*Vespertilio sinensis*	1	5		24.3	1566.06	216	0			

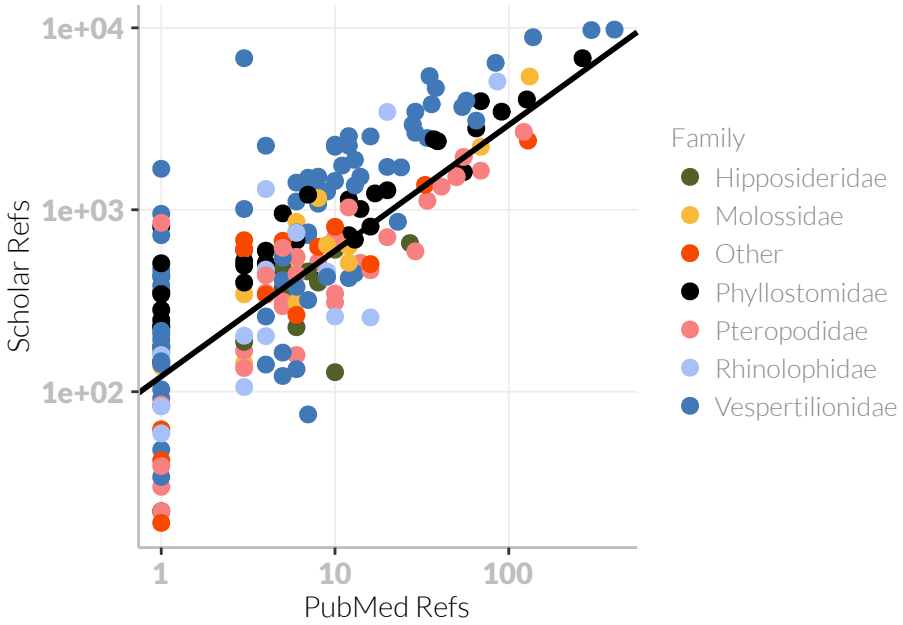


Figure A.1 Logged number of references on Scholar and PubMed, with a fitted phylogenetic linear model. Colours indicate family. (pgls: $t = 19.32$, df $= 194$, $p < 10^{-5}$.)

Interactive plots

Table A.2 Model selection results for number of subspecies analysis. $\bar{AICc}$ is the mean AICc score across inlinenBoots resamplings of the null random variable. ΔAICc is the model's $\bar{AICc}$ score minus $\min(\bar{AICc})$. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights. log(Scholar)*NSubspecies implies the interaction term between study effort and number of subspecies.

Model	$\bar{AICc}$	ΔAICc	w	$\sum w$
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass) + log(RangeSize)	882.33	0.00	0.38	0.38
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass)	883.71	1.39	0.19	0.57
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass) + rand	884.56	2.24	0.12	0.70
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies	885.47	3.14	0.08	0.78
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(RangeSize)	885.51	3.18	0.08	0.86
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(RangeSize) + rand	886.27	3.94	0.05	0.91
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + rand	886.28	3.95	0.05	0.96
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize)	889.26	6.93	0.01	0.97
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize) + rand	890.13	7.80	0.01	0.98
log(Scholar) + NSubspecies + log(Mass)	890.66	8.34	0.01	0.99
log(Scholar) + NSubspecies + log(Mass) + rand	891.59	9.26	0.00	0.99
log(Scholar) + NSubspecies	892.30	9.98	0.00	0.99
log(Scholar) + NSubspecies + log(RangeSize)	892.31	9.98	0.00	1.00
log(Scholar) + NSubspecies + log(RangeSize) + rand	893.15	10.82	0.00	1.00
log(Scholar) + NSubspecies + rand	893.19	10.86	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize)	897.19	14.86	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize) + rand	898.05	15.72	0.00	1.00
log(Scholar) + log(Mass)	898.36	16.03	0.00	1.00
log(Scholar) + log(RangeSize)	899.13	16.80	0.00	1.00
log(Scholar)	899.20	16.87	0.00	1.00
log(Scholar) + log(Mass) + rand	899.26	16.94	0.00	1.00
log(Scholar) + log(RangeSize) + rand	899.95	17.63	0.00	1.00
log(Scholar) + rand	900.06	17.73	0.00	1.00
NSubspecies + log(Mass) + log(RangeSize) + rand	906.60	24.27	0.00	1.00
NSubspecies + log(Mass) + log(RangeSize)	907.03	24.70	0.00	1.00
NSubspecies + log(RangeSize)	914.05	31.73	0.00	1.00
NSubspecies + log(RangeSize) + rand	914.85	32.52	0.00	1.00
NSubspecies + log(Mass)	920.11	37.79	0.00	1.00
NSubspecies + log(Mass) + rand	921.06	38.73	0.00	1.00
NSubspecies	923.37	41.04	0.00	1.00
NSubspecies + rand	924.26	41.94	0.00	1.00
log(Mass) + log(RangeSize)	924.61	42.28	0.00	1.00
log(Mass) + log(RangeSize) + rand	924.68	42.35	0.00	1.00
log(RangeSize)	931.53	49.20	0.00	1.00
log(RangeSize) + rand	932.35	50.02	0.00	1.00
log(Mass)	941.11	58.78	0.00	1.00
log(Mass) + rand	942.07	59.75	0.00	1.00
Intercept only	943.72	61.39	0.00	1.00
rand	944.64	62.31	0.00	1.00

Table A.3 Model selection results for effective gene flow analysis. $\bar{AICc}$ is the mean AICc score across inlinenBoots resamplings of the null random variable. ΔAICc is the model's $\bar{AICc}$ score minus $\min(\bar{AICc})$. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights.

Model	$\bar{AICc}$	ΔAICc	w	$\sum w$
log(Scholar) + Gene Flow + log(Mass)	70.57	0.00	1.00	1.00
log(RangeSize)	104.66	34.09	0.00	1.00
log(Mass)	105.62	35.06	0.00	1.00
Gene Flow + log(Mass)	107.45	36.88	0.00	1.00
rand	110.97	40.40	0.00	1.00
log(Mass) + rand	113.96	43.40	0.00	1.00
Gene Flow + log(Mass) + rand	116.98	46.41	0.00	1.00
Gene Flow + rand	118.94	48.37	0.00	1.00
log(Scholar) + log(Mass) + rand	120.29	49.72	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + rand	122.68	52.11	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize) + rand	124.08	53.52	0.00	1.00
log(Scholar) + rand	125.67	55.10	0.00	1.00
log(RangeSize) + rand	126.07	55.50	0.00	1.00
log(Scholar) + log(Mass)	126.52	55.95	0.00	1.00
log(Scholar)	126.62	56.05	0.00	1.00
log(Mass) + log(RangeSize) + rand	127.51	56.94	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize)	128.01	57.44	0.00	1.00
log(Scholar) + log(RangeSize)	128.11	57.54	0.00	1.00
log(Scholar) + log(RangeSize) + rand	128.94	58.38	0.00	1.00
log(Scholar) + Gene Flow	129.13	58.56	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + log(RangeSize) + rand	129.14	58.57	0.00	1.00
log(Scholar) + Gene Flow + rand	129.18	58.61	0.00	1.00
Gene Flow + log(RangeSize) + rand	129.29	58.72	0.00	1.00
log(Mass) + log(RangeSize)	129.50	58.93	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + log(RangeSize)	130.80	60.24	0.00	1.00
log(Scholar) + Gene Flow + log(RangeSize)	130.92	60.35	0.00	1.00
Gene Flow + log(Mass) + log(RangeSize) + rand	131.18	60.61	0.00	1.00
Gene Flow + log(Mass) + log(RangeSize)	131.85	61.28	0.00	1.00
log(Scholar) + Gene Flow + log(RangeSize) + rand	132.59	62.02	0.00	1.00
Gene Flow + log(RangeSize)	133.12	62.56	0.00	1.00
Gene Flow	135.79	65.22	0.00	1.00
Intercept only	136.23	65.66	0.00	1.00

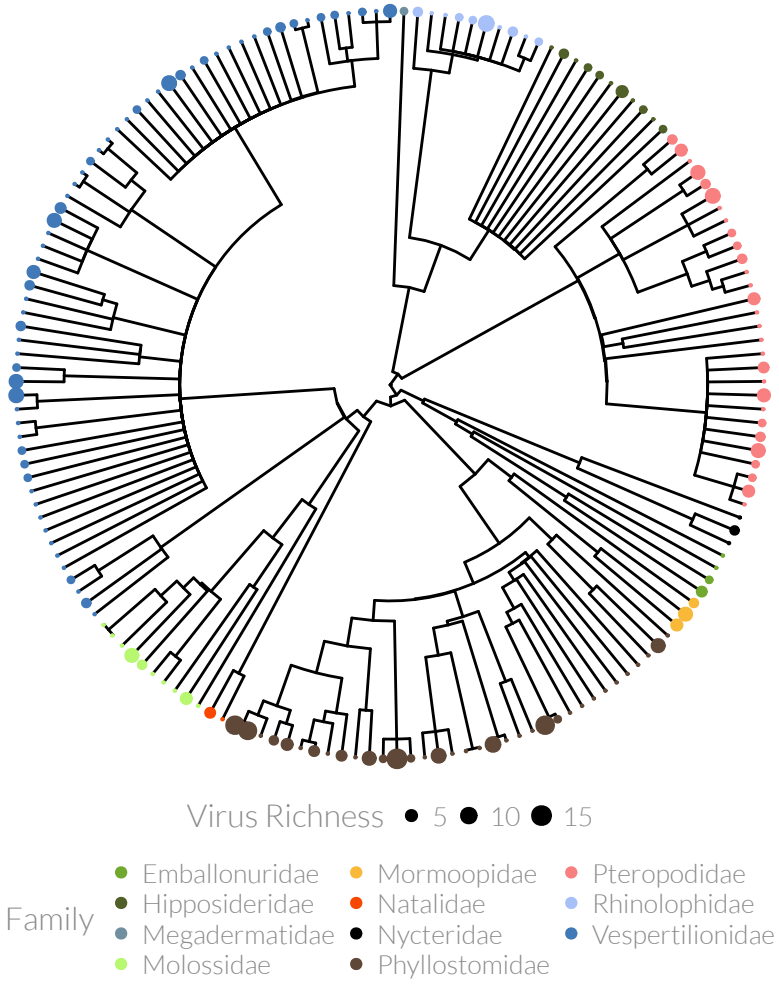


Figure A.2 The distribution of viral richness on the alternate phylogeny. The phylogeny is from Jones et al. (2005) (version 2) pruned to include all species used in either the number of subspecies or gene flow analysis. Dot size shows the number of known viruses for that species and colour shows family. The red scale bar shows 25 million years. There is no clear association between phylogeny and virus richness (pgls: $\lambda = 0.04$, $p = 0.21$).

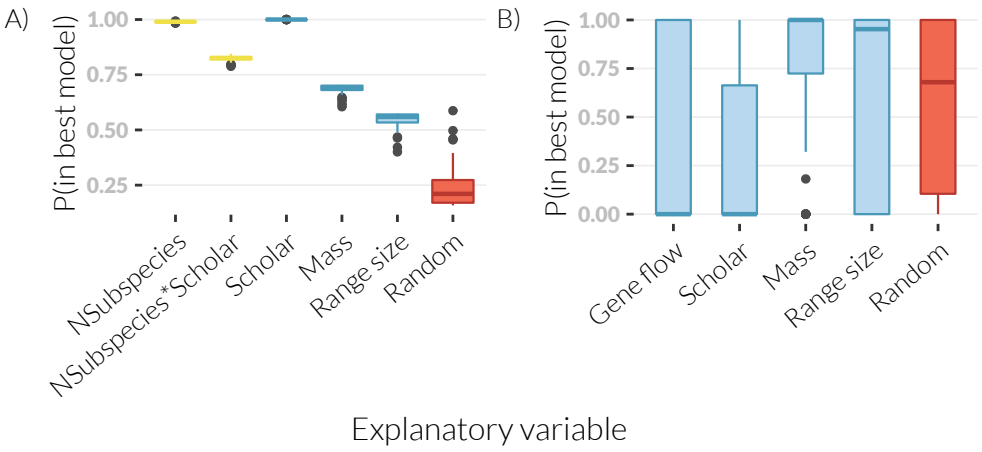


Figure A.3 Akaike variable weights for both analyses using the phylogeny from (Jones et al. 2005). The probability that each variable is in the best model (amongst the models test) is shown, with the boxplots showing the variation amongst the models over 50 resamplings of the uniformly random "null" variable. The three bars of the boxplot show the median values and upper and lower quartiles of the data, vertical lines show the range and points display outliers. The purple "Random" box is the uniformly random variable.

Table A.4 Model selection results for number of subspecies analysis using phylogeny from (Jones et al. 2005). $\bar{AIC}c$ is the mean AICc score across inlinenBoots resamplings of the null random variable. ΔAICc is the model's $\bar{AIC}c$ score minus $\min(\bar{AIC}c)$. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights. log(Scholar)*NSubspecies implies the interaction term between study effort and number of subspecies.

Model	$\bar{AIC}c$	ΔAICc	w	$\sum w$
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass) + log(RangeSize)	756.44	0.00	0.21	0.21
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass)	756.90	0.46	0.17	0.38
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies	757.64	1.19	0.12	0.49
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(Mass) + rand	758.04	1.59	0.09	0.59
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(RangeSize)	758.35	1.90	0.08	0.67
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + rand	758.79	2.34	0.07	0.73
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize)	759.28	2.83	0.05	0.79
log(Scholar)*NSubspecies + log(Scholar) + NSubspecies + log(RangeSize) + rand	759.50	3.06	0.05	0.83
log(Scholar) + NSubspecies + log(Mass)	759.92	3.47	0.04	0.87
log(Scholar) + NSubspecies + log(Mass) + log(RangeSize) + rand	760.33	3.89	0.03	0.90
log(Scholar) + NSubspecies	760.76	4.31	0.02	0.92
log(Scholar) + NSubspecies + log(Mass) + rand	760.99	4.54	0.02	0.94
log(Scholar) + NSubspecies + log(RangeSize)	761.34	4.90	0.02	0.96
log(Scholar) + NSubspecies + rand	761.83	5.39	0.01	0.98
log(Scholar) + NSubspecies + log(RangeSize) + rand	762.42	5.98	0.01	0.99
log(Scholar) + log(Mass) + log(RangeSize)	765.17	8.73	0.00	0.99
log(Scholar) + log(Mass)	765.74	9.29	0.00	0.99
log(Scholar)	766.00	9.55	0.00	0.99
log(Scholar) + log(Mass) + log(RangeSize) + rand	766.21	9.76	0.00	1.00
log(Scholar) + log(RangeSize)	766.49	10.05	0.00	1.00
log(Scholar) + log(Mass) + rand	766.78	10.33	0.00	1.00
log(Scholar) + rand	767.04	10.60	0.00	1.00
log(Scholar) + log(RangeSize) + rand	767.54	11.10	0.00	1.00
NSubspecies + log(Mass) + log(RangeSize)	778.19	21.74	0.00	1.00
NSubspecies + log(Mass) + log(RangeSize) + rand	779.22	22.78	0.00	1.00
NSubspecies + log(RangeSize)	784.68	28.23	0.00	1.00
NSubspecies + log(RangeSize) + rand	785.76	29.31	0.00	1.00
NSubspecies + log(Mass)	789.77	33.32	0.00	1.00
log(Mass) + log(RangeSize) + rand	790.42	33.98	0.00	1.00
log(Mass) + log(RangeSize)	790.49	34.04	0.00	1.00
NSubspecies + log(Mass) + rand	790.85	34.41	0.00	1.00
NSubspecies	792.53	36.09	0.00	1.00
NSubspecies + rand	793.64	37.19	0.00	1.00
log(RangeSize)	796.89	40.44	0.00	1.00
log(RangeSize) + rand	797.96	41.52	0.00	1.00
log(Mass)	804.51	48.06	0.00	1.00
log(Mass) + rand	804.78	48.33	0.00	1.00
Intercept only	806.58	50.13	0.00	1.00
rand	807.66	51.22	0.00	1.00

Table A.5 Model selection results for effective gene flow analysis using phylogeny from (Jones et al. 2005). $\bar{AICc}$ is the mean AICc score across inlinenBoots resamplings of the null random variable. ΔAICc is the model's $\bar{AICc}$ score minus $\min(\bar{AICc})$. w is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w$ is the cumulative sum of the Akaike weights.

Model	$\bar{AICc}$	ΔAICc	w	$\sum w$
log(Mass) + log(RangeSize)	106.05	0.00	1.00	1.00
log(Scholar) + log(Mass) + rand	119.34	13.30	0.00	1.00
Gene Flow + log(Mass)	120.15	14.11	0.00	1.00
log(Mass) + rand	122.83	16.78	0.00	1.00
log(Mass)	123.09	17.04	0.00	1.00
log(Scholar) + log(Mass)	124.22	18.18	0.00	1.00
Gene Flow + log(Mass) + log(RangeSize) + rand	124.52	18.48	0.00	1.00
log(Mass) + log(RangeSize) + rand	124.68	18.64	0.00	1.00
rand	125.42	19.37	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize) + rand	125.69	19.64	0.00	1.00
log(Scholar) + rand	126.04	19.99	0.00	1.00
Gene Flow + log(Mass) + rand	126.52	20.48	0.00	1.00
log(Scholar)	126.56	20.52	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + rand	126.90	20.86	0.00	1.00
log(Scholar) + log(RangeSize) + rand	127.83	21.78	0.00	1.00
log(Scholar) + log(RangeSize)	127.96	21.91	0.00	1.00
log(Scholar) + log(Mass) + log(RangeSize)	128.01	21.96	0.00	1.00
Gene Flow + log(RangeSize) + rand	128.07	22.02	0.00	1.00
Gene Flow + rand	128.37	22.33	0.00	1.00
log(RangeSize) + rand	129.02	22.98	0.00	1.00
log(Scholar) + Gene Flow	129.09	23.04	0.00	1.00
log(Scholar) + Gene Flow + log(Mass)	129.18	23.13	0.00	1.00
log(Scholar) + Gene Flow + rand	130.46	24.42	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + log(RangeSize)	130.80	24.76	0.00	1.00
log(Scholar) + Gene Flow + log(RangeSize)	130.81	24.76	0.00	1.00
log(Scholar) + Gene Flow + log(RangeSize) + rand	130.89	24.84	0.00	1.00
log(RangeSize)	131.22	25.17	0.00	1.00
Gene Flow + log(Mass) + log(RangeSize)	131.85	25.80	0.00	1.00
log(Scholar) + Gene Flow + log(Mass) + log(RangeSize) + rand	132.97	26.93	0.00	1.00
Gene Flow + log(RangeSize)	133.17	27.12	0.00	1.00
Gene Flow	135.91	29.86	0.00	1.00
Intercept only	136.23	30.18	0.00	1.00

Table A.6 Estimated variable weights (probability that a variable is in the best model) and their estimated coefficients for both number of subspecies and gene flow analyses using phylogeny from (Jones et al. 2005). . The coefficients for the number of subspecies variable are also separated for models with and without the interaction term because this term strongly changes the coefficient and because the coefficient can only be usefully interpreted when estimated without the interaction. However, there are no weights for these separated terms as they are not directly compared in the model selection framework.

	Number of Subspecies		*Gene flow*	
Variable	*Pr*	Coefficient	*Pr*	Coefficient
Number of subspecies				
Total	0.99	0.35		
Models without interaction term		0.5		
Models with interaction term		0.38		
Number of subspecies*log(Scholar)	0.78	0.50		
Gene flow			1.1×10^{-3}	-0.9
log(Scholar)	1.00	1.01	1.66×10^{-3}	3.17
log(Mass)	0.62	0.47	1	-0.4
log(Range size)	0.45	0.33	1	3.9
Random	0.29	-4.77×10^{-3}	2×10^{-3}	0.2

B Appendix: Understanding how population structure affects pathogen richness in a mechanistic model of bat populations

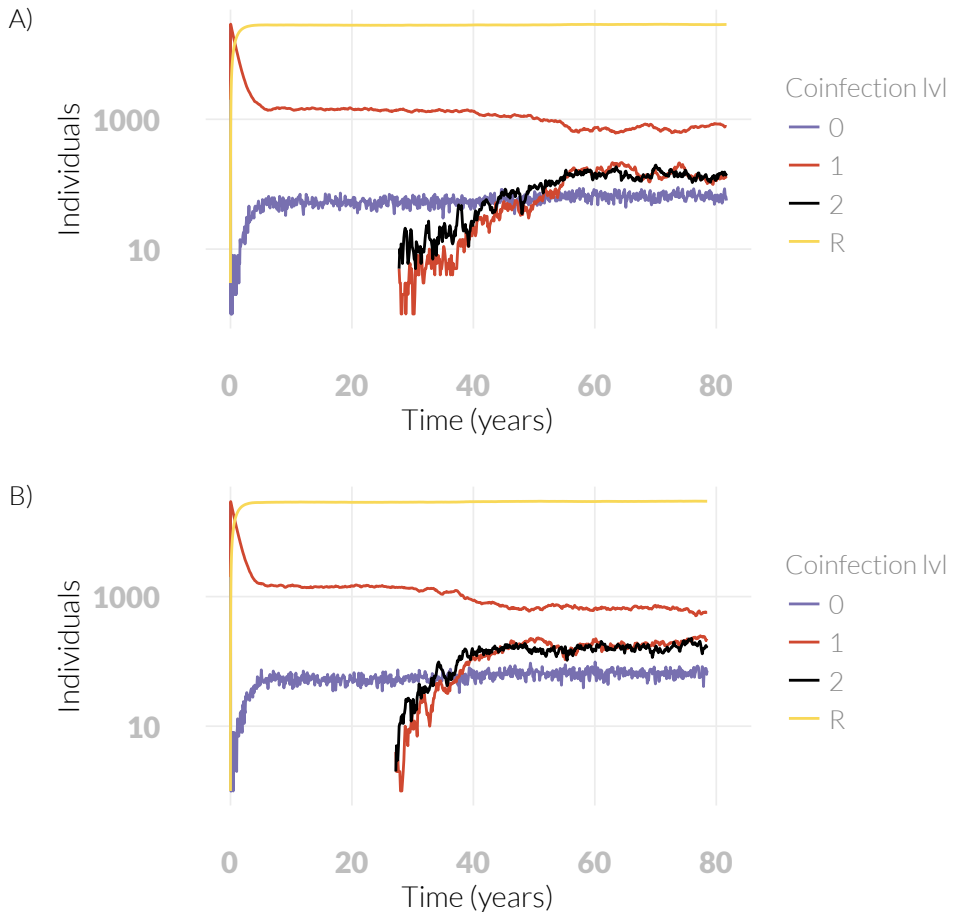


Figure B.1 Two examples (A and B) of a succesful invasion plotted on a logged y-axis. The lines are coloured such that blue represents susceptibles, brown represents individuals infected with one pathogen (the two seperate brown lines are pathogen 1 and 2), black represents co-infected individuals and yellow represents recovered and immune individuals. Pathogen 2 is seeded after $3 imes 10^4$ events (approximately 30 years). Simulations are run on a fully-connected network. Parameter values are: dispersal rate = 0.1, transmission rate = 0.2. All other parameters are as stated in Table 3.1.

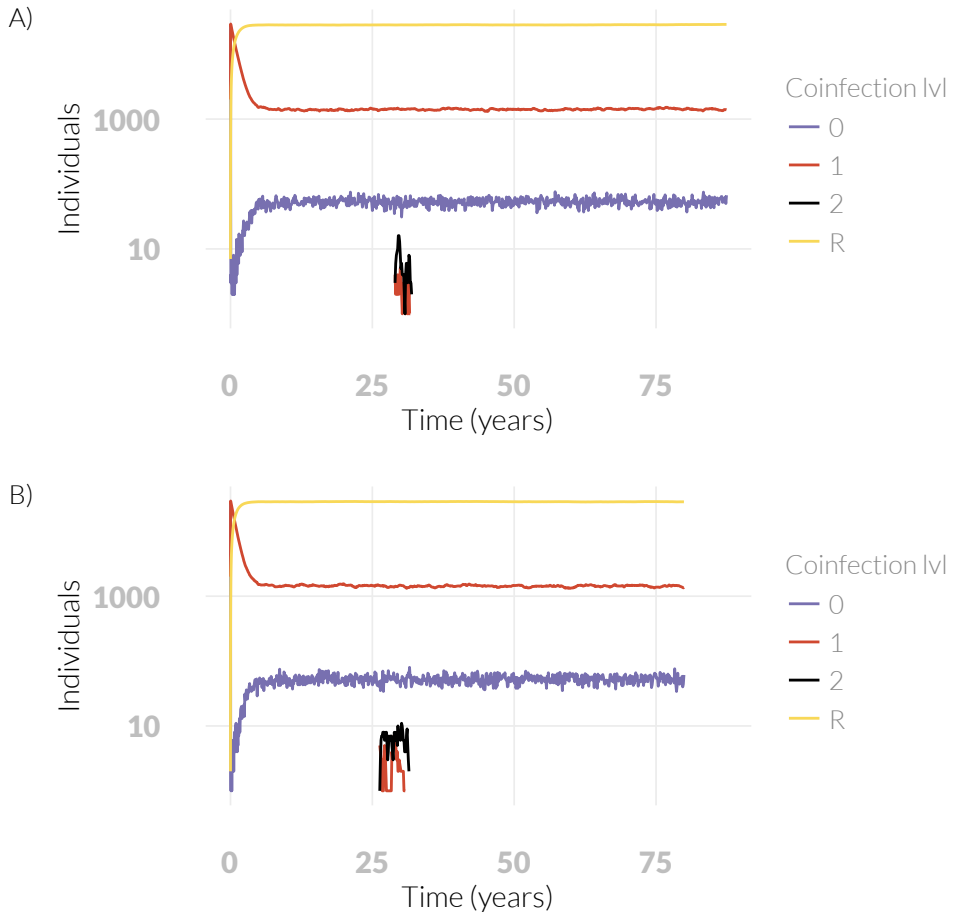


Figure B.2 Two examples (A and B) of an unsuccesful invasion plotted on a logged y-axis. The lines are coloured such that blue represents susceptibles, brown represents individuals infected with one pathogen (the two seperate brown lines are pathogen 1 and 2), black represents co-infected individuals and yellow represents recovered and immune individuals. Pathogen 2 is seeded after $3imes10^4$ events (approximately 30 years). It can be seen that after seeding pathogen 2, there is a very short period before extinction as opposed to a long fade out of disease. Simulations are run on a fully-connected network. Parameter values are: dispersal rate = 0.1, transmission rate = 0.2. All other parameters are as stated in Table 3.1.

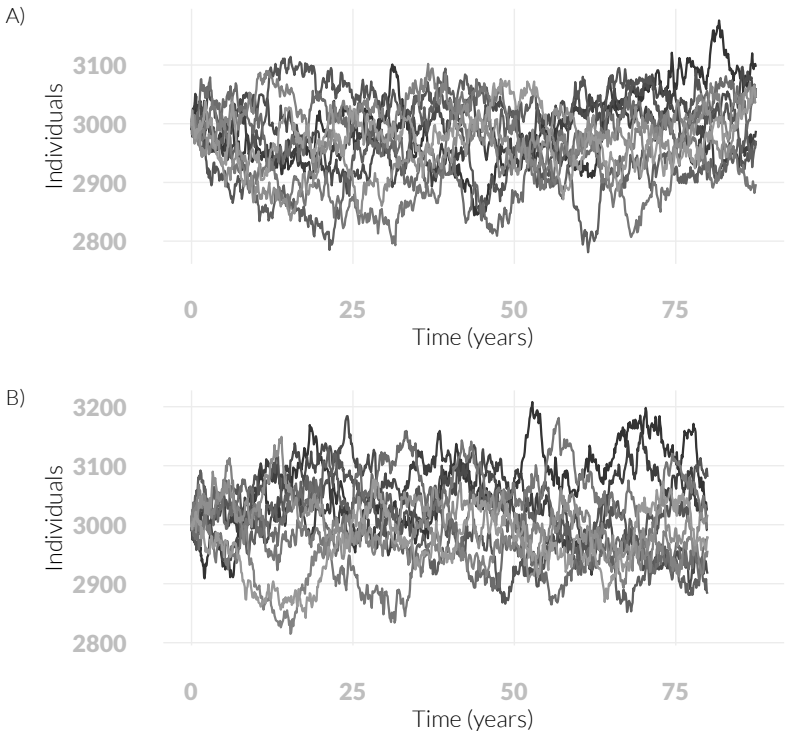


Figure B.3 Two examples (A and B) of the change in colony sizes throughout a simulation (note the truncated y-axis). The size of each colony changes as a random walk. However, given the length of the simulations, there is little risk of colonies going extinct or becoming very large. Birth and death rate are equal and set to 0.05, giving a generation time of 20 years. The metapopulation network is fully-connected and the dispersal rate is 0.1 per year. The starting colony size is 3,000.

C Appendix: A generalised random encounter model for estimating animal density with remote sensor data

C.1 Table of symbols

Table C.1 List of symbols used to describe the gREM and simulations. '-' means the quantity has no units.

Symbol	Description	Units
θ	Sensor width	rad
α	Animal signal width	rad
x_i	Focal angle, $i \in \{1, 2, 3, 4\}$	rad
r	Detection distance	m
$\bar{p}$	Average profile width	m
p	A specific profile width	m
v	Velocity	$\mathrm{m\,s^{-1}}$
t	Time	s
z	Number of detections	-
D	Animal density	$\mathrm{m^{-2}}$
T	Step length	s
N	Number of steps per simulation	-
d	Distance moved in a time step	m
S	Probability of remaining stationary	-
A	Maximum turning angle	rad

C.2 Supplementary Methods

C.2.1 Introduction

These supplementary methods derive all the models used. For continuity, the gas model derivation is included here as well as in the main text. The calculation of all integrals used in the gREM is included in the Python script S3.

C.2.2 Gas model

Following (Yapp 1956), we derive the gas model where sensors can capture animals in any direction and animal signals are detectable from any direction ($\theta = 2\pi$ and $\alpha = 2\pi$). We assume that animals are in a homogeneous environment, and move in straight lines of random direction with velocity v. We allow that our stationary sensor can capture animals at a detection distance r and that if an animal moves within this detection zone they are captured with a probability of one, while animals outside the zone are never captured.

In order to derive animal density, we need to consider relative velocity from the reference frame of the animals. Conceptually, this requires us to imagine that all animals are stationary and randomly distributed in space, while the sensor moves with velocity v. If we calculate the area covered by the sensor during the survey period we can estimate the number of animals the sensor should capture. As a circle moving across a plane, the area covered by the sensor per unit time is $2rv$. The number of expected captures, z, for a survey period of t, with an animal density of D is $z = 2rvtD$. To estimate the density, we rearrange to get $D = z/2rvt$.

C.2.2.1 gREM derivations for different detection and signal widths

Different combinations of θ and α would be expected to occur (*e.g.*, sensors have different detection widths and animals have different signal widths). For different combinations θ and α, the area covered per unit time is no longer given by $2rv$. Instead of the size of the sensor detection zone having a diameter of $2r$, the size changes with the approach angle between the sensor and the animal. For any given signal width and detector width and depending on the angle that the animal approaches the sensor, the width of the area within which an animal can be detected is called the profile, p. The size of the profile (averaged across all approach angles) is defined as the average profile $\bar{p}$. However, different combin-

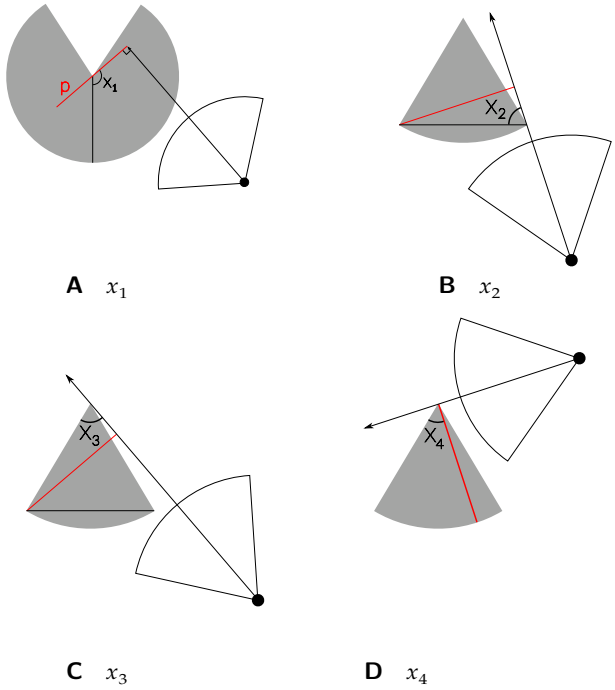


Figure C.1 The location of the focal angles $x_{i\in[1,4]}$. x_1 is used in SE and NE models (including the gas model). $x_2 - x_4$ are used in NW and SW models. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line. (a) Animal is directly approaching the sensor at $x_1 = \frac{\pi}{2}$. (b) Animal is directly approaching the sensor at $x_2 = \frac{\pi}{2}$. x_2 then decreases until the profile is perpendicular to the edge of the detection region. (c) When the profile is perpendicular to the edge of the detection region, $x_3 = \theta$. (d) x_4 measures the angle between the left side of the detection region and the profile.

ations of θ and α need different equations to calculate $\bar{p}$. This $\bar{p}$ is the only thing that changes

We have identified the parameter space for the combinations of θ and α for which the derivation of the equations are the same (defined as sub-models in the gREM) (Figure 5.2). For example, the gas model becomes the simplest gREM sub-model (upper right in Figure 5.2) and the REM from (Rowcliffe et al. 2008) is another gREM sub-model where $\theta < \pi/2$ and $\alpha = 2\pi$.

Models with $\theta = 2\pi$ are described first (the gas model described above and SE1). Then models with $\theta > \pi$ are described (NE then SE). Finally models with

$\theta < \pi$ (NW then SW) are described.

C.2.3 Model SE1

SE1 is very similar to the gas model except that because $\alpha \leq \pi$ the profile width is no longer $2r$ but is instead limited by the width of the animal signal. We therefore get a profile width of $2r \sin(\alpha/2)$ instead.

$$\bar{p}_{\text{SE1}} = \frac{1}{\pi} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} 2r \sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 \tag{C.1}$$

$$\bar{p}_{\text{SE1}} = 2r \sin\left(\frac{\alpha}{2}\right) \tag{C.2}$$

This profile is integrated over the interval $[\frac{\pi}{2}, \frac{3\pi}{2}]$ which is π radians of rotation starting with the animal moving directly towards the sensor (Figure C.1a).

C.2.4 Models NE1–3

When the detection zone is not a circle, we have more complex profiles and need to explicitly write functions for the width of the profile for every approach angle. We then use these functions to find the average profile width $\bar{p}$ for all approach angles by integrating across all 2π angles of approach and dividing by 2π.

There are three submodels within quadrant NE (Figure 5.2). Note that NE1 covers the area $\alpha = 2\pi$ as well as the triangle below it as these two models are specified exactly the same, rather than happening to have equal results.

These models have up to five profiles.

1. The profile width starts, from $x_1 = \frac{\pi}{2}$ as $2r$.

2. At $x_1 = \theta/2$, the right hand side of the profile cannot be r wide as the corner of the 'blind spot' limits its size to being $r \cos(x_1 - \theta/2)$ wide (Figure C.2A).

3. The third profile is only found in NE3. If $\alpha < 4\pi - 2\theta$, then at $x_1 = \theta/2 + \pi/2$, when the profile is perpendicular to the edge of the blind spot, the whole right side of the profile is invisible to the sensor (Figure C.2B). This gives a profile size of just r.

4. At some point, the sensor can detect animals once they have passed the blind spot giving a profile width of $r + r \cos(x_1 + \theta/2)$ (Figure C.2C). From

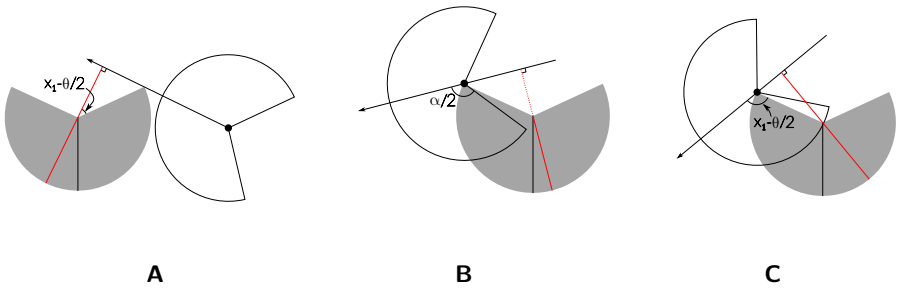


Figure C.2 Three of the integrals in NE models. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line. Dashed red lines indicate areas where animals cannot be detected. (a) The second integral in NE with width $r + r\cos(x_1 - \theta/2)$. (b) The third integral in NE3. $\alpha/2$ is labelled. As it is small, animals to the right of the detector cannot be detected. (c) After further rotation, $\alpha/2$ is now bigger than the angle shown and animals to the right of the detector can again be detected.

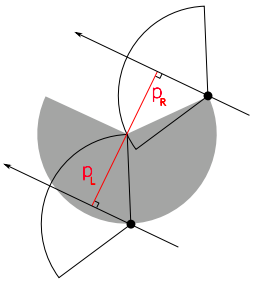


Figure C.3 The second integral in SE. The right side of the profile (p_R) is limited by the size of the sensor region while the left side of the profile (p_L) is limited by the size of the signal width. The full profile has width $p = r\sin(\alpha/2) + r\cos(\theta/2 - x_1)$. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line.

$x_1 = \pi$, if the animal signal is wide enough to be detected in this area, this is the wider profile. This then defines the split between NE1 and NE2. In NE1, with $\alpha > 3\pi - \theta$, the animal signal is wide enough that at $x_1 = \pi$ the animal can immediately be detected past the blind spot and so this profile is used. In NE2, with $\alpha < 3\pi - \theta$, the latter profile is reached at $5\pi/2 - \theta/2 - \alpha/2$.

5. Finally, common to all three models, at $x_1 = 2\pi - \theta/2$ the profile becomes a full $2r$ once again.

C.2.4.1 Model NE1

Submodel NE1 exists within the area bounded by $\alpha \leq 2\pi$, $\theta \leq 2\pi$ and $\alpha \geq 3\pi - \theta$ (Figure 5.2). It has four profiles; it does not include the r profile at $x_1 = \pi$ (profile described in point (3) in Section C.2.4). Furthermore, θ is wide enough that the $r + r\cos(x_1 + \theta/2)$ profile starts at π. This then gives us

$$\begin{aligned}\bar{p}_{\text{NE1}} =& \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r\,\mathrm{d}x_1 + \int_{\frac{\theta}{2}}^{\pi} r\cos\left(\frac{\theta}{2} - x_1\right) + r\,\mathrm{d}x_1 \right. \\ &\left. + \int_{\pi}^{2\pi-\frac{\theta}{2}} r\cos\left(\frac{\theta}{2} + x_1\right) + r\,\mathrm{d}x_1 + \int_{2\pi-\frac{\theta}{2}}^{\frac{3\pi}{2}} 2r\,\mathrm{d}x_1 \right) \qquad \text{(C.3)}\end{aligned}$$

$$\bar{p}_{\text{NE1}} = \frac{r}{\pi}\left(\theta + 2\sin\left(\frac{\theta}{2}\right)\right) \qquad \text{(C.4)}$$

C.2.4.2 Model NE2

Model NE2 is bounded by $\alpha \leq 3\pi - \theta$, $\alpha \geq 4\pi - 2\theta$ and $\alpha \geq \pi$ (Figure 5.2). It is the same as NE1 except that the third profile starts at $5\pi/2 - \theta/2 - \alpha/2$ instead of at π which is reflected in the different bounds in the second and third integral.

$$\begin{aligned}\bar{p}_{\text{NE2}} =& \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r\,\mathrm{d}x_1 + \int_{\frac{\theta}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\cos\left(\frac{\theta}{2} - x_1\right) + r\,\mathrm{d}x_1 \right. \\ &\left. + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{2\pi-\frac{\theta}{2}} r\cos\left(\frac{\theta}{2} + x_1\right) + r\,\mathrm{d}x_1 + \int_{2\pi-\frac{\theta}{2}}^{\frac{3\pi}{2}} 2r\,\mathrm{d}x_1 \right) \qquad \text{(C.5)}\end{aligned}$$

$$\bar{p}_{\text{NE2}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right) + \cos\left(\frac{\alpha}{2} + \theta\right)\right) \qquad \text{(C.6)}$$

C.2.4.3 Model NE3

Model NE3 is bound by $\alpha \leq 4\pi - 2\theta$, $\alpha \geq \pi$ and $\theta \geq \pi$ (Figure 5.2). It is the same as NE2 except that it contains the extra profile with width r (third integral).

$$\bar{p}_{\text{NE3}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r\,\mathrm{d}x_1 + \int_{\frac{\theta}{2}}^{\frac{\theta}{2}+\frac{\pi}{2}} r\cos\left(\frac{\theta}{2} - x_1\right) + r\,\mathrm{d}x_1 \right.$$
$$\left. + \int_{\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\,\mathrm{d}x_1 + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{2\pi-\frac{\theta}{2}} r\cos\left(\frac{\theta}{2} + x_1\right) + r\,\mathrm{d}x_1 + \int_{2\pi-\frac{\theta}{2}}^{\frac{3\pi}{2}} 2r\,\mathrm{d}x_1 \right) \quad \text{(C.7)}$$

$$\bar{p}_{\text{NE3}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right) + 1\right) \quad \text{(C.8)}$$

C.2.5 Models SE2–4

Quadrant SE contains three submodels excluding SE1 (Figure 5.2). The differences between these three models are similar to the differences between the models in NE. There are four possible profiles.

1. As α is less than π the profile is smaller than $2r$, even when the sensor width is a full diameter. The profile width starts as $2r\sin(\alpha/2)$.
2. Similar to NE, at a certain point the blind spot of the sensor area limits the profile width on one side. This gives a profile width of $r\sin(\alpha/2) + r\cos(x_1 - \theta/2)$ (Figure C.3).
3. Also similar to NE, there can be a point where the right side of the profile is 0 giving a profile width of $r\sin(\alpha/2)$.
4. If $\alpha \leq 2\pi - \theta$, then at $x_1 = \theta/2 + \pi/2 + \alpha/2$ the profile width becomes 0. This inequality distinguishes between SE3 and SE4.
5. The third profile $r\sin(\alpha/2)$ starts at $\theta/2 + \pi/2$ while at $5\pi/2 - \alpha/2 - \theta/2$ the profile returns to size $2r\sin(\alpha/2)$. If $\theta/2 + \pi/2 \geq 5\pi/2 - \alpha/2 - \theta/2$ we go straight into the $2r\sin(\alpha/2)$ profile and miss the $r\sin(\alpha/2)$ profile. SE2 and SE3 are separated by this inequality which simplifies to $\alpha \leq 4\pi - 2\theta$.

C.2.5.1 Model SE2

SE2 is bounded by $\alpha \geq 4\pi - 2\theta$, $\alpha \leq \pi$ and $\theta \leq 2\pi$ (Figure 5.2). As $\alpha \geq 4\pi - 2\theta$, there is no $r\sin(\alpha/2)$ profile. As $\alpha \leq 4\pi - 2\theta$, the profile returns to $2r\sin(\alpha/2)$ rather than going to 0. These integrals relate to profiles (1), (2) and (5) in Section C.2.5.

$$\bar{p}_{\text{SE2}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 + \int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) + r\cos\left(\frac{\theta}{2}-x_1\right) \mathrm{d}x_1 + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{3\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 \right) \tag{C.9}$$

$$\bar{p}_{\text{SE2}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + \cos\left(\frac{\alpha}{2}+\theta\right)\right) \tag{C.10}$$

C.2.5.2 Model SE3

SE3 is bounded by $4\pi - 2\theta \leq \alpha \leq 4\pi - 2\theta$ and $\alpha \leq \pi$ (Figure 5.2). Therefore there is a $r\sin(\alpha/2)$ profile but no $0r$ profile. This relates to profiles (1), (2), (3) and (5) above.

$$\bar{p}_{\text{SE3}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 + \int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) + r\cos\left(\frac{\theta}{2}-x_1\right) \mathrm{d}x_1 + \int_{\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{3\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 \right) \tag{C.11}$$

$$\bar{p}_{\text{SE3}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.12}$$

C.2.5.3 Model SE4

Finally SE4 is bounded by $\alpha \leq 4\pi - 2\theta$, $\alpha \leq \pi$ and $\theta \leq \pi$ (Figure 5.2). It is the same as SE3 except that the profile becomes 0 rather than returning to $2r\sin(\alpha/2)$. This relates to profiles (1), (2), (3) and (4) above though profile (4) with width 0 is not shown.

$$\bar{p}_{\text{SE4}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 + \int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) + r\cos\left(\frac{\theta}{2}-x_1\right) \mathrm{d}x_1 + \int_{\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{\alpha}{2}+\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 \right) \tag{C.13}$$

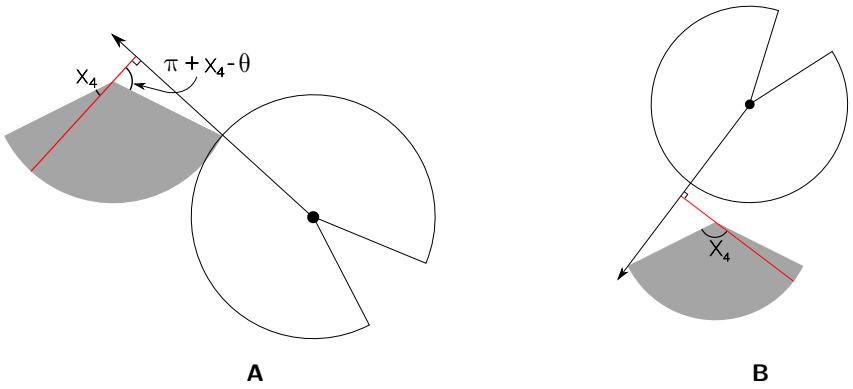


Figure C.4 The second and fourth profiles of NW1. The left side of of both profiles is of width r while the right side differs. (a) The right side of the profile is $r\cos(\pi+x_4-\theta) = -r\cos(\theta-x_4)$ (b) The right side is $r\cos(\pi-x_4) = -r\cos x_4$ respectively. In both images the sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line.

$$\bar{p}_{\text{SE4}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.14}$$

C.2.6 Model NW1

NW1 is the first model with $\theta < \pi$. Whereas previously the focal angle has always been x_1, we now use different focal angles. x_2 and x_3 correspond to γ_1 and γ_2 in (Rowcliffe et al. 2008) while x_4 is new. They are described in Figure C.1b–d.

There are five different profiles in NW1.

1. x_2 has an interval of $[\pi/2, \theta/2]$ which is from the angle of approach being directly towards the sensor until the profile is parallel to the left hand radius of the sensor sector (Figure C.1B). During this interval the profile width is $2r\sin(\theta/2)\sin(x_2)$ which is calculated using the equation for the length of a chord . Note that while rotating anti-clockwise (as usual) x_2 decreases in size.

2. From here, we examine focal angle x_4 (note that x_3 is used in later models, but is not relevant here). The left side of the profile is a full radius while the right side is limited to $-r\cos(x_4 - \theta)$ (Figure C.4A).

3. At $x_4 = \theta - \pi/2$, the profile is perpendicular to the edge of the sensor area. Here, the right side of the profile is $0r$ giving a profile size of r.

4. When $x_4 = \pi/2$ the angle of approach is from behind the sensor, but we can once again be detected on the right side of the sensor (Figure C.4B). Therefore the width of the profile is $r - r\cos(x_4)$.

5. Finally, we have the x_2 profile, but from behind.

$$\bar{p}_{\text{NW1}} = \frac{1}{\pi}\left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{0}^{\theta-\frac{\pi}{2}} r - r\cos(-x_4+\theta)\,\mathrm{d}x_4 + \int_{\theta-\frac{\pi}{2}}^{\frac{\pi}{2}} r\,\mathrm{d}x_4 + \int_{\frac{\pi}{2}}^{\theta} r - r\cos(x_4)\,\mathrm{d}x_4 + \int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 \right) \quad \text{(C.15)}$$

$$\bar{p}_{\text{NW1}} = \frac{r}{\pi}(\theta + 2) \quad \text{(C.16)}$$

C.2.7 Models NW2–4

The models NW2–4 have the five potential profiles in NW1 but not all profiles occur in each model, and the angle at which transitions occur are different. Furthermore, there is one extra profile possible.

1. When approaching the sensor from behind, there is a period where the profile is r wide as in NW1 profile (3).

2. At some point after profile (1) animals to the right of the sensor can be detected again. If this occurs in the x_4 region, the profile width becomes $r - r\cos(x_4)$ as in NW1.

3. However, as α is now less than 2π, animals to the right of the sensor may be undetectable until we are in the second x_2 region. In this case, when we first enter the second x_2 region, the profile has a width of $r\cos(x_2 - \theta/2)$. This occurs only if $\alpha \leq 3\pi - 2\theta$. This inequality is found by noting that

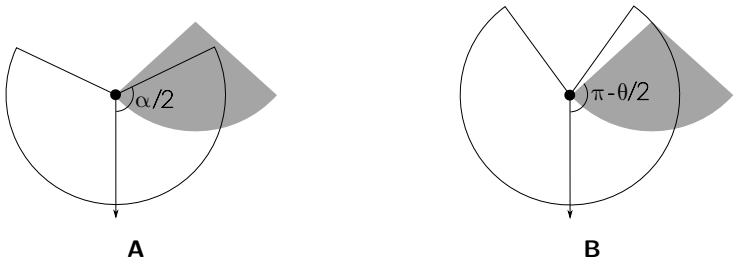


Figure C.5 Profile sizes when an animal approaches from behind in models NW2–4. If α is relatively large, animals can be detected when approaching from behind. Otherwise animals cannot be detected. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. (a) If $\alpha/2$ is less than $\pi - \theta/2$, as is the case here, then the width of the profile when an animal approaches directly from behind is zero. (b) If $\alpha/2 > \pi - \theta/2$ the profile width from behind is $2r \sin(\theta/2) \sin(x_2)$.

animals to the right of the sensor can be detected again at $x_4 = 3\pi/2 - \alpha$ but the x_2 region starts at $x_4 = \theta$. The new profile in x_2 will only occur if $\theta < 3\pi/2 - \alpha/2$ which is rearranged to find the inequality above. This defines the boundary between NW2 and NW3.

4. As $\alpha \leq 2\pi$ it is possible that when the angle of approach is from directly behind the sensor the animal will not be detected at all. This is the case if $\alpha/2 \leq \pi - \theta/2$ (Figure C.5). This inequality (simplified as $\alpha \leq 2\pi - \theta$) defines the boundary between NW3 and NW4.

C.2.7.1 Model NW2

NW2 is bounded by $\alpha \geq 3\pi - 2\theta$, $\alpha \leq 2\pi$ and $\theta \leq \pi$ (Figure 5.2).

NW2 has all five profiles as found in NW1. However, the change from the r profile (third integral) to the $r - r\cos(x_4)$ profile (fourth integral) occurs at $x_4 = 3\pi/2 - \alpha/2$ instead of at $x_4 = \theta$.

$$\bar{p}_{\text{NW2}} = \frac{1}{\pi}\left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2)\, \mathrm{d}x_2 + \int_{0}^{\theta-\frac{\pi}{2}} r - r\cos(-x_4 + \theta)\, \mathrm{d}x_4 \right.$$

$$+\int_{\theta-\frac{\pi}{2}}^{\frac{3\pi}{2}-\frac{\alpha}{2}} r\,\mathrm{d}x_4+\int_{\frac{3\pi}{2}-\frac{\alpha}{2}}^{\theta} r-r\cos\left(x_4\right)\,\mathrm{d}x_4+\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2\Bigg) \tag{C.17}$$

$$\bar{p}_{\mathrm{NW2}}=\frac{r}{\pi}\left(\theta-\cos\left(\frac{\alpha}{2}\right)+1\right) \tag{C.18}$$

C.2.7.2 Model NW3

NW3 is bounded by $\alpha \leq 3\pi-2\theta$, $\alpha \geq 2\pi-\theta$ and $\theta \geq \pi/2$ (Figure 5.2).

NW3 does not have the fourth integral from NW2 as animals are not detectable to the right of the sensor until after the x_4 region has ended and the x_2 region has begun. Therefore the second x_4 integral has an upper limit of θ and the profile after has a width of $r\cos(x_2-\theta/2)$ and is integrated with respect to x_2. The final integral starts at $x_4=3\pi/2-\alpha/2-\theta/2$ and has the full width of $2r\sin(x_2)\sin(\theta/2)$.

$$\bar{p}_{\mathrm{NW3}}=\frac{1}{\pi}\Bigg(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2+\int_{0}^{\theta-\frac{\pi}{2}} r-r\cos\left(-x_4+\theta\right)\,\mathrm{d}x_4+\int_{\theta-\frac{\pi}{2}}^{\theta} r\,\mathrm{d}x_4$$
$$+\int_{\frac{\theta}{2}}^{\frac{3\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\cos\left(\frac{\theta}{2}-x_2\right)\mathrm{d}x_2+\int_{\frac{3\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2\Bigg) \tag{C.19}$$

$$\bar{p}_{\mathrm{NW3}}=\frac{r}{\pi}\left(\theta-\cos\left(\frac{\alpha}{2}\right)+1\right) \tag{C.20}$$

C.2.7.3 Model NW4

Finally, NW4 is bounded by $\alpha \geq \pi$, $\theta \geq \pi/2$ and $\alpha \leq 2\pi-\theta$ (Figure 5.2). NW4 is the same as NW3 except that the final profile width is zero and this profile is reached at $\alpha/2+\theta/2-\pi/2$.

$$\bar{p}_{\mathrm{NW4}}=\frac{1}{\pi}\Bigg(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2+\int_{0}^{\theta-\frac{\pi}{2}} r-r\cos\left(-x_4+\theta\right)\,\mathrm{d}x_4$$
$$+\int_{\theta-\frac{\pi}{2}}^{\theta} r\,\mathrm{d}x_4+\int_{\frac{\theta}{2}}^{\frac{\alpha}{2}+\frac{\theta}{2}-\frac{\pi}{2}} r\cos\left(\frac{\theta}{2}-x_2\right)\mathrm{d}x_2\Bigg) \tag{C.21}$$

$$\bar{p}_{\mathrm{NW4}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.22}$$

C.2.8 Model REM

REM is the model from (Rowcliffe et al. 2008). It has $\alpha = 2\pi$ and $\theta \leq \pi/2$ (Figure 5.2). It has three profile widths, two of which are repeated, once as the animal approaches from in front of the sensor and once as the animal approaches from behind the sensor.

1. Starting with an approach direction of directly towards the sensor, and examining focal angle x_2, the profile width is $2r\sin(x_2)\sin(\theta/2)$.
2. When the profile is perpendicular to the radius on the right hand of the sector sensor region, we instead examine x_3 where the profile width is $r\sin(x_3)$.
3. At $x_3 = \pi/2$ the profile becomes simply r and this continues for θ radians of x_4.
4. The x_3 profile is then repeated with an approach direction from behind the sensor.
5. Finally the x_2 profile is repeated, again with an approach direction from behind the sensor.

$$\begin{aligned}\bar{p}_{\mathrm{REM}} = \frac{1}{\pi}\Bigg(&\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 \\ &+ \int_{0}^{\theta} r\,\mathrm{d}x_4 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2\Bigg)\end{aligned} \tag{C.23}$$

$$\bar{p}_{\mathrm{REM}} = \frac{r}{\pi}(\theta + 2) \tag{C.24}$$

C.2.9 Models NW5–7

In the models NW5–7, the sensor has $\theta \leq \pi/2$ as in the REM. As $\alpha \geq \pi$ a lot of the profiles are similar to the REM. Specifically, the first three profiles are always the same as the first three profiles of the REM. This is because when an animal is moving towards the sensor, the $\alpha \geq \pi$ signal is no different to a 2π signal. However, when approaching the sensor from behind, things are slightly different. The animal can only be detected by the sensor if the signal width is large enough that it can be detected once it has passed the sensor.

1. Starting with an approach direction of directly towards the sensor, and examining focal angle x_2, the profile width is $2r\sin(x_2)\sin(\theta/2)$.
2. When the profile is perpendicular to the radius edge of the sector sensor region, we instead examine x_3 where the profile width is $r\sin(x_3)$.
3. At $x_3 = \pi/2$ the profile becomes simply r and this continues for θ radians of x_4.
4. If $\alpha \leq 2\pi + 2\theta$, the animal becomes undetectable during this profile when x_3 has decreased in size to $\pi - \alpha/2$. This inequality marks the boundary between NW7 and NW6.
5. If instead $\alpha \geq 2\pi + 2\theta$ then the animal does not become undetectable during the x_3 focal angle. Instead the profile has width greater than zero for the whole of the x_3 angle. The x_2 profile starts with width $r\cos(x_2 - \theta/2)$ as only animals approaching to the left of the sensor are detectable.
6. During this second x_2 profile the signal width needed for animals to be detected to the left of the detector is increasing while the angle needed for animals to be detected to the right of the detector is decreasing. Therefore, either the left side becomes undetectable, making both sides undetectable (this occurs if $\alpha \leq 2\pi - \theta$ as in NW6)
7. or the right becomes detectable (if $\alpha \geq 2\pi - \theta$ as in NW5), making both sides detectable and giving a profile width of $2r\sin(x_2)\sin(\theta/2)$.

C.2.9.1 Model NW5

NW5 is bounded by $\alpha \geq 2\pi - \theta$, $\alpha \leq 2\pi$ and $\theta \leq \pi/2$ (Figure 5.2).

It is the same as REM except that it includes the extra profile in x_2 (the fifth integral) where only animals approaching to the left of the profile are detected.

$$\begin{aligned}\bar{p}_{\mathrm{NW5}} =& \frac{1}{\pi}\Bigg(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int_{0}^{\theta} r\,\mathrm{d}x_4 \\ &+ \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{3\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r\cos\left(\frac{\theta}{2}-x_2\right)\mathrm{d}x_2 \\ &+ \int_{\frac{3\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2\Bigg) \end{aligned} \tag{C.25}$$

$$\bar{p}_{\mathrm{NW5}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right)+1\right) \tag{C.26}$$

C.2.9.2 Model NW6

NW6 is bounded by $\alpha \leq 2\pi - \theta$, $\alpha \geq 2\pi + 2\theta$ and $\theta \leq \pi/2$ (Figure 5.2).

NW6 is the same NW5 except that as $\alpha \leq 2\pi - \theta$, animals that approach from directly behind the detector are not detected. Therefore at $x_2 = \alpha/2 + \theta/2 - \pi/2$ the profile width goes to zero and therefore the last integral in NW5 is not included.

$$\begin{aligned}\bar{p}_{\mathrm{NW6}} =& \frac{1}{\pi}\Bigg(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 \\ &+ \int_{0}^{\theta} r\,\mathrm{d}x_4 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\alpha}{2}+\frac{\theta}{2}-\frac{\pi}{2}} r\cos\left(\frac{\theta}{2}-x_2\right)\mathrm{d}x_2\Bigg) \end{aligned} \tag{C.27}$$

$$\bar{p}_{\mathrm{NW6}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right)+1\right) \tag{C.28}$$

C.2.9.3 Model NW7

NW7 is bounded by $\alpha \geq 2\pi + 2\theta$, $\alpha \geq \pi$ and $\theta \geq 0$ (Figure 5.2).

It is similar to NW6 but does not include the last integral as during the x_3 profile, at $x_3 = \pi - \alpha/2$ the signal width is too small for any animals to be detected, so the profile width goes to zero.

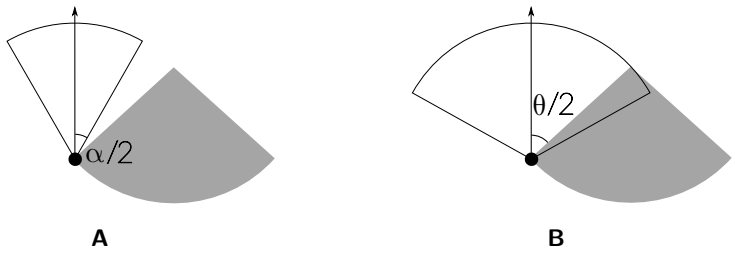


Figure C.6 The first profile in SW models is limited by either α or β depending on whether $\alpha < \beta$. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. (a) As $\alpha/2 < \theta/2$ the profile width is limited by the signal width rather than the sensor region. The profile width is $2r\sin(\alpha/2)$ (b) As $\alpha/2 > \theta/2$ the profile width is limited by the sensor region, not the signal width. The profile width is $2r\sin(\theta/2)\sin(x_2)$.

$$\bar{p}_{\mathrm{NW7}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int_{0}^{\theta} r\,\mathrm{d}x_4 + \int_{\pi-\frac{\alpha}{2}}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 \right) \tag{C.29}$$

$$\bar{p}_{\mathrm{NW7}} = \frac{r}{\pi}\left(\theta - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.30}$$

C.2.10 Model SW1–3

The models in SW1–3 are described with the two focal angles used in models NW2–4, x_2 and x_4. As $\alpha \leq \pi$ an animal can never be detected if it is approaching the detector from behind. This makes these models simpler in that they go through the x_2 and x_4 profiles only once each.

There are five potential profile sizes.

1. At the beginning of x_2, with an approach direction directly towards the sensor, the parameter that limits the width of the profile can either be the sensor width, in which case the profile width is $2r\sin(\theta/2)\sin(x_2)$.

2. Or the signal width can be the limiting parameter, in which case the profile width is instead $2r\sin(\alpha/2)$ (Figure C.6)

3. The next potential profile in x_2 has a width of $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ as the right side of the profile is limited by the width of the sensor region while the left side is limited by the signal width. However, the angle at which the profile starts depends on whether the first profile was 1) or 2) above. If the first profile is profile 1) then the profile is limited on both sides by the sensor region and then the left side of the profile becomes limited by the signal width. This happens at $x_2 = \pi/2 - \alpha/2 + \theta/2$. If however the first profile was 2) then the first profile is limited by the signal width. We move into the new profile when the right side of the profile becomes limited by the sensor region. This occurs at $x_2 = \pi/2 + \alpha/2 - \theta/2$.

4. In the x_4 region the left side of the profile is always $r\sin(\alpha/2)$ while the right side is either 0, giving a profile of $r\sin(\alpha/2)$.

5. Or limited by the sensor giving a profile of size $r\sin(\alpha/2) - r\cos(x_4 - \theta)$.

C.2.10.1 Model SW1

SW1 is bounded by $\alpha \geq \theta$, $\alpha \leq \pi$ and $\theta \leq \pi$ (Figure 5.2).

As α is large the first profile is limited by the size of the sensor region giving it a width of $2r\sin(\theta/2)\sin(x_2)$. It is the only one of the three SW models to start in this way. Later on, still with x_2 as the focal angle the left side of the profile does become limited by the signal width. So at $x_2 = \pi/2 - \alpha/2 + \theta/2$ the profile width becomes $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$.

As we enter the x_4 region, the profile remains limited by the signal on the left and by the sensor on the right, giving a profile width of $r\sin(\alpha/2) - r\cos(x_4 - \theta)$. Finally, at $x_4 = \theta - \pi/2$ the right side of the profile becomes zero and the profile is width is $r\sin(\alpha/2)$.

$$\begin{aligned}\bar{p}_{\text{SW1}} =& \frac{1}{\pi}\left(\int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int_{\frac{\theta}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right)\mathrm{d}x_2\right.\\ &\left. + \int_{0}^{\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos(\theta - x_4)\,\mathrm{d}x_4 + \int_{\theta-\frac{\pi}{2}}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_4\right) \qquad \text{(C.31)}\\ \bar{p}_{\text{SW1}} =& \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \qquad \text{(C.32)}\end{aligned}$$

C.2.10.2 Model SW2

SW2 is bounded by $\theta \geq \pi/2$, $\alpha \leq \theta$ and $\alpha \geq 2\theta - \pi$ (Figure 5.2).

SW2 is largely similar to SW1. However, as $\alpha \leq \theta$ the first profile is limited by α and not by the detection region. Therefore the first profile has width $2r\sin(\alpha/2)$. This also means the transition to the second profile occurs at $x_2 = \pi/2 + \alpha/2 - \theta/2$ instead of $x_2 = \pi/2 - \alpha/2 + \theta/2$.

$$\begin{aligned}\bar{p}_{\text{SW2}} =& \frac{1}{\pi}\left(\int_{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_2 + \int_{\frac{\theta}{2}}^{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right) \mathrm{d}x_2 \right. \\ &\left. + \int_{0}^{\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\theta - x_4\right) \mathrm{d}x_4 + \int_{\theta-\frac{\pi}{2}}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_4 \right) \end{aligned} \tag{C.33}$$

$$\bar{p}_{\text{SW2}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.34}$$

C.2.10.3 Model SW3

SW3 is bounded by $\alpha \leq 2\theta - \pi$ and $\theta \leq \pi$ (Figure 5.2).

SW3 is similar to SW2 except that the profile does not become limited by sensor at all during the the x_4 regions. Therefore, at $x_4 = 0$ the profile is still of width $2r\sin(\alpha/2)$. Only at $x_4 = \theta - \pi/2 - \alpha/2$ does the profile become limited on the right by the sensor region.

$$\begin{aligned}\bar{p}_{\text{SW3}} =& \frac{1}{\pi}\left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_2 + \int_{0}^{-\frac{\pi}{2}+\theta-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_4 \right. \\ &\left. + \int_{-\frac{\pi}{2}+\theta-\frac{\alpha}{2}}^{\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\theta - x_4\right) \mathrm{d}x_4 + \int_{\theta-\frac{\pi}{2}}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_4 \right) \end{aligned} \tag{C.35}$$

$$\bar{p}_{\text{SW3}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.36}$$

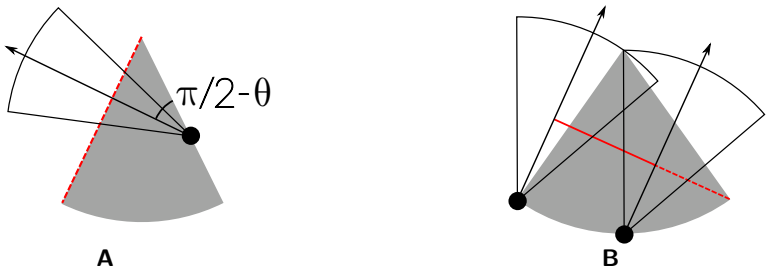


Figure C.7 Description of two profiles in SW models. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line. Dashed red lines indicate areas where animals cannot be detected. (a) At $x_4 = 0$, if $\alpha/2 < \pi/2 - \theta$ then $\alpha/2$ is too small for an animal to be detected at all during the x_4 profile (shown with dashed red). This inequality simplifies to $\alpha < \pi - 2\theta$. (b) The right of the profile is limited by the signal width, not the sensor. On the left, the profile is limited by the sensor and not the signal. Overall the profile width is $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$.

C.2.11 Model SW4–9

As $\alpha < \pi$, animals approaching the sensor from behind can never be detected, so unlike REM, the second x_2 and x_3 profiles are always zero. The six models are split by three inequalities that relate to the models as follows.

1. Models with $\alpha \leq \pi - 2\theta$ have no x_4 profile. This is because at $x_4 = 0$, the signal width is already too small to be detected as can be seen in Figure C.7A where $\alpha/2 < \pi/2 - \theta$ which simplifies to give the previous inequality.

2. Models with $\alpha \leq \theta$ are limited by α in the first, x_2 region (Figure C.6), rather than being limited by θ. Therefore this first profile is of width $2r\sin(\alpha/2)$ rather than $2r\sin(\theta/2)\sin(x_2)$.

3. Finally, models with $\alpha \leq 2\theta$ have a second profile in x_2 where to one side of the sensor α is the limiting factor of profile width, while on the other side θ is (Figure C.7B). This gives a width of $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$. This profile does not occur in models with $\alpha \geq 2\theta$.

C.2.11.1 Model SW4

SW4 is bounded by $\alpha \leq \theta$, $\alpha \geq \pi - 2\theta$ and $\theta \leq \pi/2$ (Figure 5.2). Therefore it does contain a x_4 profile, starts with an α limited profile and does contain the $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ profile in x_2.

$$\begin{aligned}\bar{p}_{\text{SE4}} =& \frac{1}{\pi}\left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 + \int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) + r\cos\left(\frac{\theta}{2} - x_1\right) \mathrm{d}x_1 \right. \\ &\left. + \int_{\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{\alpha}{2}+\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_1 \right) \qquad \text{(C.37)} \\ \bar{p}_{\text{SE4}} =& \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \qquad \text{(C.38)}\end{aligned}$$

C.2.11.2 Model SW5

SW5 is the only model with a tetrahedral bounding region. It is bounded by $\alpha \geq \theta$, $\alpha \geq \pi - 2\theta$, $\alpha \leq 2\theta$ and $\theta \leq \pi/2$ (Figure 5.2). Therefore it does contain a x_4 profile, but starts with a θ limited profile. It does contain the $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ profile in x_2.

$$\begin{aligned}\bar{p}_{\text{SW5}} =& \frac{1}{\pi}\left(\int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right) \mathrm{d}x_2 \right. \\ &+ \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2} + x_2\right) \mathrm{d}x_2 \\ &\left. + \int_{\theta}^{\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_3 + \int_{0}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_4 \right) \qquad \text{(C.39)} \\ \bar{p}_{\text{SW5}} =& \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \qquad \text{(C.40)}\end{aligned}$$

C.2.11.3 Model SW6

SW6 is bounded by $\alpha \geq \pi - 2\theta$, $\alpha \geq 2\theta$ and $\alpha \leq \pi$ (Figure 5.2). It starts with a θ limited profile and has a x_4 profile. However, it does not contain the $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ profile.

$$\bar{p}_{\text{SW6}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2 + \int_{\theta}^{\frac{\alpha}{2}} r\sin\left(x_3\right)\,\mathrm{d}x_3 + \int_{\frac{\alpha}{2}}^{\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_3 + \int_{0}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_4 \right) \tag{C.41}$$

$$\bar{p}_{\text{SW6}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.42}$$

C.2.11.4 Model SW7

SW7 is bounded by $\alpha \leq \pi - 2\theta$, $\alpha \leq \theta$ and $\alpha < 0$ (Figure 5.2). Therefore it does not contain a x_4 profile. It starts with an α limited profile and contains the $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ profile in x_2.

$$\bar{p}_{\text{SW7}} = \frac{1}{\pi}\left(\int_{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_2 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right)\mathrm{d}x_2 + \int_{\theta}^{\frac{\alpha}{2}+\theta} r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_3 \right) \tag{C.43}$$

$$\bar{p}_{\text{SW7}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.44}$$

C.2.11.5 Model SW8

SW8 is bounded by $\alpha \leq \pi - 2\theta$, $\alpha \geq \theta$ and $\alpha \leq 2\theta$ (Figure 5.2). It starts with a θ limited profile. It does contain the $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$ profile in x_2 but does not have a x_4 profile.

$$\bar{p}_{\text{SW8}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right)\,\mathrm{d}x_2 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right)\mathrm{d}x_2 + \int_{\theta}^{\frac{\alpha}{2}+\theta} r\sin\left(\frac{\alpha}{2}\right)\mathrm{d}x_3 \right) \tag{C.45}$$

$$\bar{p}_{\text{SW8}} = \frac{r}{\pi}\left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.46}$$

C.2.11.6 Model SW9

Finally, SW9, the last model, is bounded by y $\alpha \leq \pi - 2\theta$, $\alpha \geq 2\theta$ and $\theta \geq 0$ (Figure 5.2). Therefore it starts with a θ limited profile. However it does not contain the extra x_2 profile nor a x_4 profile.

$$\bar{p}_{\text{SW9}} = \frac{1}{\pi}\left(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin\left(x_2\right) \, \mathrm{d}x_2 + \int_{\theta}^{\frac{\alpha}{2}} r \sin\left(x_3\right) \, \mathrm{d}x_3 + \int_{\frac{\alpha}{2}}^{\frac{\alpha}{2}+\theta} r \sin\left(\frac{\alpha}{2}\right) \mathrm{d}x_3 \right) \tag{C.47}$$

$$\bar{p}_{\text{SW9}} = \frac{r}{\pi}\left(\theta \sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1\right) \tag{C.48}$$

C.3 Supplementary Information: Simulation model results of the gREM precision

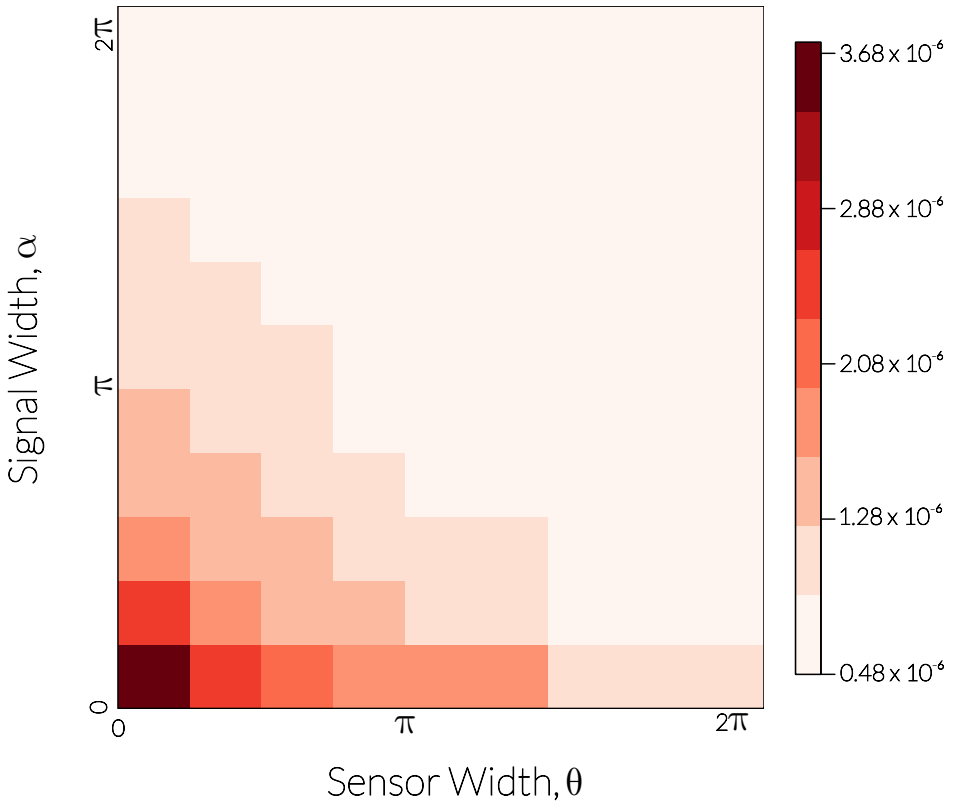


Figure C.1 Simulation model results of the gREM precision given a range of sensor and signal widths, shown by the standard deviation of the error between the estimated and true densities. Standard deviations are shown from deep red to pink, representing high to low values between 0.483×10^{-6} to 3.74×10^{-6}.

C.4 Supplementary Information: Impact of parameter error

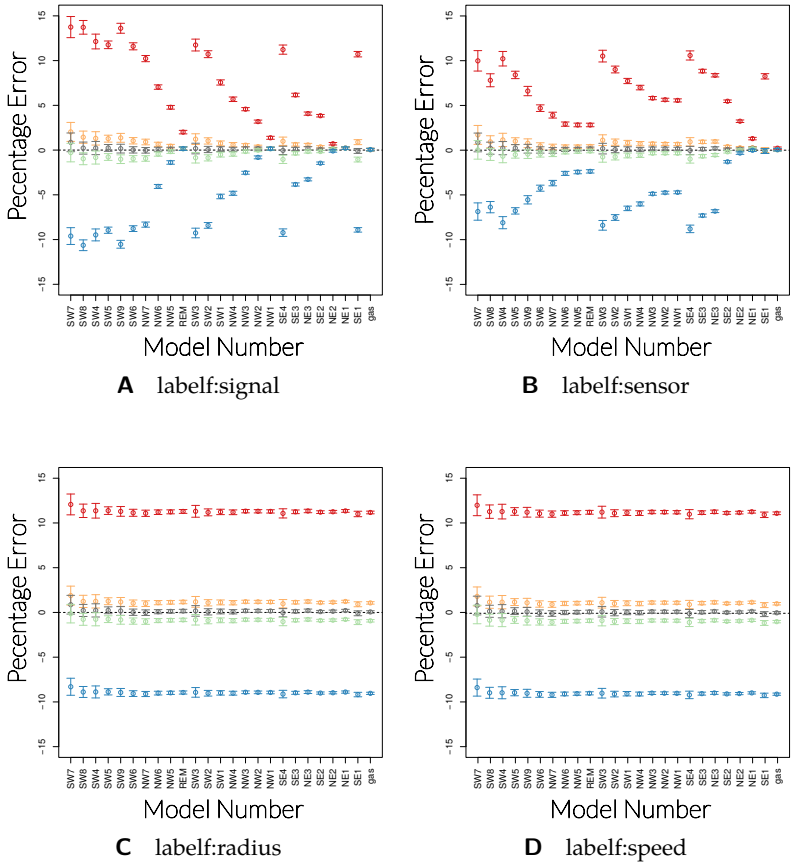


Figure C.2 Model sensitivity (for all gREM submodels) to error in estimates of a) signal width α, b) sensor width θ, c) detection distance r and d) animal movement speed v. Estimates are -10% (red), -1% (orange), 0% (grey), +1% (green) and +10% (blue) of the true parameter value. The black dashed line indicates zero error in density estimates. The error bars 95% confidence intervals across all simulations.

D Colophon

This thesis was set using LaTeX, XeLaTeX and BibLaTeX. The formatting is defined by the *phdthesis* class by Robert Stanley. The TeX Gyre Pagella typeface is used in the main text while Lato Light and **Lato Black** are used in the figures. Chapters 2, 3 and 4 are entirely reproducible *knitr* documents (Xie 2015). Code for the simulations in Chapter 5 is not combined into a *knitr* document but code for creating figures is. All code will be made available on Github at . Plots were created with a combination of *Inkscape*, *ggplot2* (Wickham 2009), *palettetown* (Lucas 2015b), *ggtree* (Yu 2015), *cowplot* (Wilke 2015) and base *R* (R Development Core Team 2010). References were handled with *JabRef* (JabRef Development Team 2015).