epilit.bib

The role of social ecology in Chiropteran pathogen richness and zoonotic potential

Tim C. D. Lucas

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Mathematics and Physics in the Life Sciences and Experimental

Biology

University College London

December 16, 2015

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 work.

Abstract

My research is about stuff.

It begins with a study of some stuff, and then some other stuff and things.

There is a 300-word limit on your abstract.

Acknowledgements

Acknowledge all the things!

Contents

List of Figures

List of Tables

Chapter 1

Introductory Material

Some stuff about things. Some more things. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

General intro

Total mammalian viral diversity. Known mammal viruses?

Bit about how to search for viruses? Deep pcr etc.

Number of known zoonotic viruses/pathogens. Number of known from mammals. Examples of mortality rates and examples of scale of outbreaks.

Some outbreaks caused by bats. Number of known bat viruses, known bat zoonoses.

Bats are second largest order of mammals. Long lived, social and fly.

Specific intro

Pathogen richness

Define. Simple count of pathogens across whole species. Ignores heterogeneit in prevalence, across space and time.

-interaction nonstopmodetionEmpirical evidence from mammals

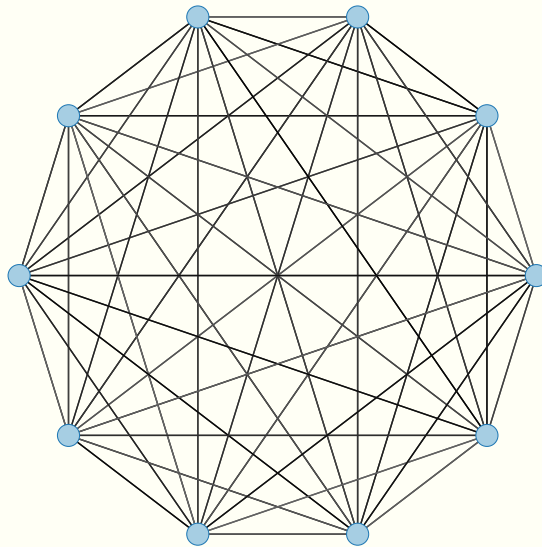# Pathogen richness

Define. Simple count of pathogens across whole species. Ignores heterogeneit in prevalence, across space and time.

-interaction nonstopmodetionEmpirical evidence from mammals

Chapter 2

Does network structure of bat populations promote viral diversity?

Abstract

An increasingly large fraction of emerging diseases come from wild 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 particularly wild host species increases with the diversity of pathogens in that species. However, the factors that control pathogen diversity in wild populations are still unclear.

Host species traits such as population density, longevity, body size and population structure have been shown to correlate with pathogen diversity. Typically it is assumed that well-connected, unstructured populations promote the invasion of new pathogens and therefore increase pathogen diversity. However, this assumption is largely untested; in particular our mechanistic understanding of how population structure affects pathogen diversity, in the presence of inter-pathogen competition, is poor. Greater understanding is needed to clarify exactly by which mechanisms population structure affects pathogen diversity.

It is unknown whether population structure allows invading pathogens to escape from competition by stochastically creating areas of low pathogen prevalence. I hypothesise that low dispersal rates and a low number of connections in a metapopulation network will allow invading pathogens to establish more readily, thus increasing pathogen diversity. I test these hypothesese using population 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 find that neither population connectedness nor dispersal rate affect the probability that a new pathogen will invade into a population. The assumption that factors causing high R_0 allow new pathogens to invade and therefore increase pathogen diversity is not supported. Instead we find that changes in population structure do not affect the probability of invasion of a new pathogen.

This result implies that population structure does not control pathogen diversity. This also implies that population structure is not a useful proxy for pathogen diversity with respect to zoonotic disease surveillance, for example

in bats.

Introduction

General Intro

The diversity of pathogens in a wild animal species strongly affects the chance that a disease from that species will infect humans. As over 50% of emerging infectious diseases have an animal source [,], understanding and predicting this process is a global health priority []. However, the factors that control the diversity of pathogens in a wild animal population are still unclear[].

Specific Intro

A number of host traits have been shown to correlate with pathogen richness including body size [,], population density [,] and range size [,].

However, empirical, correlative studies are often contradictory due to small sample sizes, noisy data and because empirical relationships often do not generalise across taxa (though see [] for a meta-analysis). Furthermore, the correlation between many traits (e.g. []) makes it hard to clearly distinguish which factors are important. Knowing the factors that correlate with pathogen richness also does not tell us how it controls richness. Mechanisms by which a trait could increase pathogen diversity include promoting the evolution of new strains within a species [], reduction of the rate of parasite extinction and an increased probability of pathogen invasion from other host species. These seperate mechanisms have not been examined and it is difficult to see how they could be approached through comparitive methods. We need explicit mechanistic models in order to tease apart these factors.

One host-species trait that has been largely understudied with respect to pathogen diversity is population structure. Population structure has been comprehensively studied with respect to single or competing epidemics in human populations, wildlife populations and technological networks. With respect to single disease epidemics it is clear that changes in population structure that increase R_0 increase the chance of a disease spreading and persisting. However, the assumption that population structures that yield high R_0 will also give high pathogen diversity [] is unfounded.

The processes by which a single disease spreads through a population are

very well studied. One commonly taken assumption is that factors that promote high disease spread automatically promotes high diversity. However, this ignores competitive mechanisms such as cross-immunity and depletion of susceptible hosts. A number of authors have demonstrated a competitive exclusion principle where only the most optimally spreading disease will survive when infection with one pathogen prevents infection with another pathogen [,]. 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 diversity.

Competitive affects are likely to be more or less important depending on the mechanism by which pathogen diversity is created. Competition is likely to be very strong between a newly evolved strain and the existing strain of a pathogen. In this case, there is likely to be strong cross immunity between strains. In the case of pathogens invading from other hosts species, cross immunity is likely to be less strong due to the evolutionary distance between the new and endemic pathogens. However, competition could still occur as disease induced mortality of host individuals reduces host density which promotes pathogen persistance. Finally, it is unclear how competition would affect the persistance of pathogens once they are established, though it seems that cross immunity can keep a pathogen at a low prevalence in the population (e.g. monkey pox [] and monkey malaria []). This would make stochastic extinction more likely.

The important affects of population structure on single disease epidemic spread and competing epidemics shows the important role structure has on disease dynamics. We can therefore expect it to have an important role in the dynamics of multidisease systems as well.

Studies of the role of population structure on pathogen diversity have been in very simple systems. These have been so simple that empirical data cannot easily be applied to them to predict pathogen diversity of real wild animal populations. There is a need for models that can be carefully and fully explored, while still capturing the complexities of the real world.

Analytical models of well mixed have widely different outcomes: infinite pathogen diversity [,] or competitive exclusion have both been predicted [,

]. When competitive exclusion occurs, population structure has sometimes been shown to allow coexistence [,].

Competing epidemics, or two pathogens spreading at the same time in a population, is a well studied area [,]. This area is related to the study of pathogen richness in that they indicate that dynamics of multiple pathogens in a population do depend on population structure. However, the results for short term epidemic competition do not directly transfer to the study of long term disease persistence which is the relative time scale when looking at zoonotic potential of a host species.

Few studies focus on bats, despite their role in recent zoonoses. Maganga et al. found that distribution fragmentation predicts viral richness [], but [] finds the opposite relationship. While the data set in [] is larger, the analysis in [] is much more focused on fragmentation.

Genetic correlates of population structure have also been used. Turmelle et al. [], in a small analysis, find that high F_{st} (i.e. a structured population) correlates with high richness. However, they do not account for the widely different spatial scales foudn in population structure studies, nor do they deal with the differences between F_{ST}, ϕ_{ST} and other measures appropriately.

How structured a population is can be defined in many ways on many scales. The most relevant scale is that of an epidemiological population. This is the population within which a pathogen can spread in an epidemiologically relevant time period (years or decades). It is therefore closely related to a population as defined by population genetics, but with movement defined on a shorter time scale.

The epidemiological contacts within the population can be examined at the individual level (as in contact network epidemiology) or larger scales. I consider the metapopulation network the most appropriate. Ignoring the metapopulation assumes a fully mixed population which is unlikely. Trying to study the contact network relies detailed individual level detail which is not available. Metapopulation models consider a network of small subpopulations. Within subpopulations, epidemiological contacts are fully mixed and relatively fast. Between subpopulations, epidemiological contacts are dependant on an underlying network structure and relatively slow. The network underlying the metapopulation is made up of nodes representing the sub-

populations, and edges which represent movement between subpopulations. Animals, and therefore infection, can only move between two subpopulations if they are connected by an edge.

There are two factors that affect how structured a population is, given this model framework. Firstly, dispersal is the rate at which individuals move between subpopulations. Secondly, the metapopulation network structure controls population structure. The simplest measure of how structured the network is the average number of edges each node has. In the extremes, all subpopulations could be either connected to all other subpopulations or only connected to one or two other subpopulations. However, other measures that take into account second-order structure in the network are also often used.

Bats (Order Chiroptera) have, over the last decade, become a focus for disease research [,]. Recently they have been implicated in a number of high profile diseases such as Ebola, SARS, Hendra and Nipah [,].

A number of traits have been suggested as predisposing bats towards being reservoirs of zoonotic diseases: high sympatry [], flight [] and longevity [].

Bats have an unusually large variety of social structures. Colony sizes range from 10 to 1 million []. Many bats also have interesting seasonal behaviour such as migration [,], hibernation, birth puleses and swarming behaviours.

The gap

We have very abstract, simplified models that predict zero or infinite diversity depending on specifics. These cannot be easily applied to real data. They also do not easily predict quantitative or even relative diversity as they often predict either zero or infinite diversity with nothing in between.

We need models that can quantitatively or at least relatively predict diversity in a populations. This requires a middle ground of model complexity.

There are no studies that directly model bat pathogen diversity.

Specifically we use these models to test the affects of population structure on the ability of a new pathogen to invade a population. We test two aspects of structure, dispersal rate and connectedness of the metapopulation network.

Term	Definition	Synonyms
Metapopulation	A group of subpopulations with rare movement of animals between them. Closed to outside migration.	Network
Subpopulation	A group of animals. Social interactions within a subpopulation is likely high.	Node, colony
Dispersal	Movement from one subpopulation to another	Migration
Population	A closed group of animals. No epidemiological affects from outside the group on epidemiological timescale (years – decades.)	
Pathogen diversity	The number of species or strains of pathogens in a host	Pathogen richness
Connectedness		

Table 2.1: Glossary of terms

What I did

I have run epidemiological simulations based broadly on real-world bat populations. Although still simplified, the model is complex enough that if good measurements of bat populations could be found, simulations of the real world bat population could be run.

I have studied the invasion of new pathogens as a mechanisms for increasing pathogen richness. In particular I have focussed 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 cross immunity is strong.

I have studied two metrics for population structure, dispersal rate and metapopulation network topology, to test for effects of population structure on pathogen richness.

What I found

Here I show that given the assumptions of a metapopulation, population structure does not affect the rate of invasion of new pathogens.

Methods

Metapopulation model

Two pathogen SIR model

I examine a multipathogen SIR model. This is a compartment model with individuals being classed as susceptible, infected or recovered with immunity (Figure 4.1The SIR model used.). Susceptible individuals are counted in class S. 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 pathogen. This modelling choice allows the model to be easily expanded to include more than two pathogens. The assumption of immediate recovery from all other diseases is likely to be quite accurate for very closely related pathogens as is being studied here as once an acquired immune response is activated, all infections are likely to be cleared quickly.

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 (i.e. $\alpha = 0.5$ means coinfection happens at half the rate of infections). Birth and death rates are assumed to be equal, $b = d$ (see Table 4.1All symbols used. for a list of symbols used). The time scale of the simulations are set by setting b = 0.05 per year, yielding an average host generation time of 20 years.

Metapopulation

The population is divided into a number of subpopulations. This metapopulation is modelled as a network with subpopulations being nodes and dispersal between subpopulations being indicated by edges (Figure 2.2Network topologies used to compare network connectedness). Individuals within a subpopulation interact randomly so that the subpopulation is fully mixed. Dispersal between subpopulations occurs at a rate λ. Individuals can only disperse to subpopulations connected to theirs in the network. The rate of dispersal is not affected by the number of edges a subpopulation has (the degree of the subpopulation). So the dispersal rate from a subpopulation m with degree
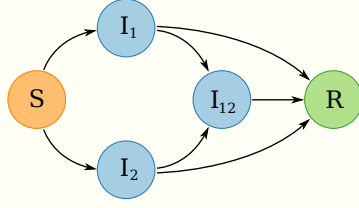
Figure 2.1: Schematic of the SIR model used. Individuals are in one five classes, susceptible (orange, S), infectious with pathogen 1, pathogen 2 or both (blue, I_1, I_2, I_{12}) or recovered and immune from further infection (green, R). Transitions between classes occurs only as indicated by arrows. 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.

k_m to subpopulation n is $\frac{\lambda}{k_m}$. Note this rate is independent of the degree of subpopulation n.

Stochastic simulations

I examine this model using stochastic, continuous time simulations using the Gillespie algorithm. At each step in the simulation we calculate the rate that each possible event might occur. One event is then randomly chosen, weighted by it's rate

$$p(\text{event } i) = \frac{r_i}{\sum_i r_i} \qquad (2.1)$$

where r_i is the rate that event i occurs. Finally, the length of the time step, δ, is drawn from an exponential distribution

$$\delta \sim \text{Exp}\left(\sum_i r_i\right). \qquad (2.2)$$

This means that the total length of each simulation is stochastic. We define the number of events we wish to simulate instead.

We can now write down the rates of all events. I define I_p^+ to be the sum of all classes that are infectious with pathogen p, for example $I_1^+ = I_1 + I_{12}$. Assuming asexual reproduction, that all classes reproduce at the same rate and that individuals are born into the susceptible class we get

$$P\left(S_{nt'} = S_{nt} + 1\right) = b\left(S_{nt} + \sum_q I_{qnt} + R_{nt}\right) \qquad (2.3)$$

where $P(S_{nt'} = S_{nt} + 1)$ is the probability that the number of susceptibles in subpopulation n will increase by 1 (a single birth) in the time interval t to t' and $\sum_q I_{qnt}$ is the sum of all infection classes $q \in 1, 2, 12$. The rates of death, given a death rate d are given by

$$P(S_{nt'} = S_{nt} - 1) = dS_{nt}, \tag{2.4}$$

$$P(I_{qnt'} = I_{qnt} - 1) = dI_{qnt}, \tag{2.5}$$

$$P(R_{nt'} = R_{nt} - 1) = dR_{nt}. \tag{2.6}$$

Infection of a susceptible with either pathogen 1 or 2, $S \to I_p$ where $p \in \{1, 2\}$, is given by

$$P(I_{pnt'} = I_{pnt} + 1, S_{nt'} = S_{nt} - 1) = \beta S_{nt} I_{pnt}^+, \tag{2.7}$$

while coinfection, given a crossimmunity factor α, is given by

$$P(I_{12,nt'} = I_{12,nt} + 1, I_{1nt'} = I_{1nt} - 1) = \alpha \beta I_{1nt} I_{2nt}^+, \tag{2.8}$$

$$P(I_{12,nt'} = I_{12,nt} + 1, I_{2nt'} = I_{2nt} - 1) = \alpha \beta I_{2nt} I_{1nt}^+. \tag{2.9}$$

The probability of migration from colony m (with degree k_m) to colony n, given a dispersal rate λ is given by

$$P(S_{nt'} = S_{nt} + 1, S_{mt'} = S_{mt} - 1) = \frac{\lambda S_{mt}}{k_m - 1}, \tag{2.10}$$

$$P(I_{qnt'} = I_{qnt} + 1, I_{qmt'} = I_{qmt} - 1) = \frac{\lambda I_{qmt}}{k_m - 1}, \tag{2.11}$$

$$P(R_{nt'} = S_{nt} + 1, R_{mt'} = R_{mt} - 1) = \frac{\lambda R_{mt}}{k_m - 1}. \tag{2.12}$$

Finally, recovery from any infectious class occurs at a rate γ

$$P(I_{qnt'} = I_{qnt} - 1, R_{nt'} = R_{nt} + 1) = \gamma I_{qnt}. \tag{2.13}$$

In each simulation the population was seeded with 10 sets of 200 infected individuals of disease 1. There 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. Disease 1 was then allowed to spread and reach equilibrium. After 3×10^5 events, 5 individuals infected with disease 2 were added to one colony. Visual inspection of preliminary simulations was used to decide on 3×10^5 as being long enough for the epidemic to reach an equilibrium state. After another 5×10^5

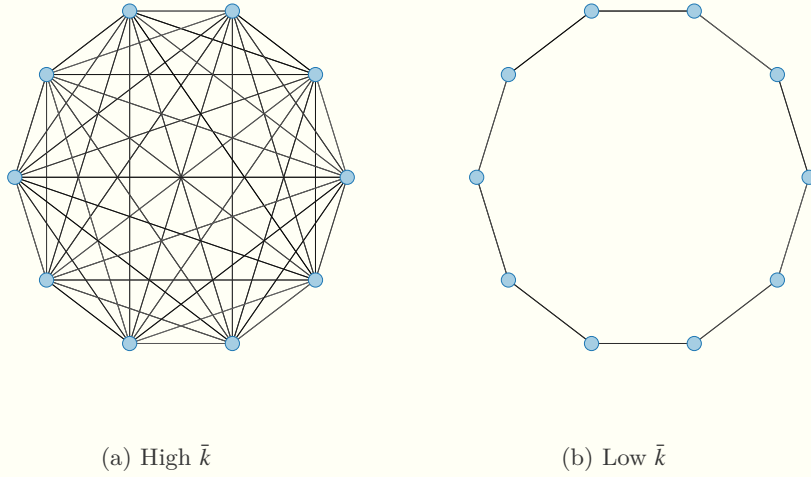(a) High $\bar{k}$ (b) Low $\bar{k}$

Figure 2.2: The two network topologies used to test whether network connectedness influences a viruses ability to invade. Blue circles are subpopulations of 3000 individuals. Dispersal only occurs between subpopulations connected by an edge (black line). The dispersal rate is held constant between the two topologies.

events the invasion of disease 2 is considered successful if any individuals are still infected with pathogen 2. Again visual inspection of preliminary simulations was used to determine whether this was long enough to be confident that an invading disease

Dispersal

The values used for the independant variables are chosen to highlight the affects of these variables. Dispersal values are $\lambda = 0.1, 0.01$ and 0.001 dispersals per individual per year. $\lambda = 0.1$ relates to individuals moving between colonies on average twice per lifetime. Therefore exclusively juvenile dispersal would have dispersal rates similar to this. Otherwise it relates to dispersal being a rare event with animals often staying in a colony for many years. $\lambda = 0.01$ relates to 20% of individuals dispersing once in their lifetime. This value is therefore close to male-biased dispersal, with female philopatry. Finally, $\lambda = 0.001$ relates to 2% of individuals dispersing in their lifetime. This therefore relates to a population that does not habitually disperse.

Network structure

The network structure is synthetically created to be either fully or minimally connected (See Figure 2.2Network topologies used to compare network connectedness). 10 subpopulations was selected as a trade off between computation time and a network complicated enough that structure might have an effect. This value is artificially small compared to wildlife populations.

Parameter selection

The fixed parameters used are chosen to roughly reflect realistic wild bat populations. The death rate d is set as 0.05 per year giving a generation time of 20 years. The birth rate b is set to be equal to d 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 very difficult to directly estimate infection durations in wild populations. But it seems that these infections might be long lasting [].

Cross immunity is set to 0.1 so that an individual infected with one disease is 90% less likely to be infected with another. This is a rather arbitrary value. However, the rationale of the model is that the invading species might be a newly speciated strain of the endemic species. Furthermore, the model assumes complete cross immunity after infection. Therefore cross immunity is likely to be very strong.

The population size of each subpopulation is set to 3000. This is appropriate for many bat species [], especially the large, frugivorous Pteropodidae that have been particularly associated with recent zoonotic diseases.

Four values of the transmission rate β are used, 0.1, 0.2, 0.3 and 0.4. All simulations are run under all four transmission rates as this is such a fundamental parameter. Given the recovery, birth and death rates we can calculate an approximation of R_0 that ignores spatial structure. That is, this is R_0 for the local, within-subpopulation dynamics. Furthermore, it is R_0 for the first pathogen; R_0 of the invading pathogen will be lower due to competition. We can calculate that $R_0 \approx \frac{\beta b}{d(d+\gamma)}$. For our four values of $\beta = 0.1, 0.2, 0.3, 0.4$ we therefore get $R_0 \approx 13.3, 33.3, 66.6$. These values are very high in part to again find a reasonable trade off between the number of simulations and the

reasonableness of the parameters. $R_0 \approx 13.3$ is similar to a highly contagious disease such as measles or pertussis.

For the simulations where an invading pathogen is added to the populations the number of invading pathogens added is set to 10. This is a trade off between getting a reasonable proportion of invasions, while still retaining the stochastic nature of invasion.

Results

Invasion

Dispersal

The proportion of invasions was not different across dispersal rates across 1.2×10^3 simulations run (Figure 2.3ANetwork structure and diversity). This was true at all transmission levels (χ^2 test. $\beta = 0.1$: $\chi^2 = 1.98$, df $= 2$, $p = 0.37$. $\beta = 0.2$: $\chi^2 = 9.44$, df $= 2$, $p = 8.9 \times 10^{-3}$. $\beta = 0.3$: $\chi^2 = 0.29$, df $= 2$, $p = 0.87$, $\beta = 0.4$: $\chi^2 = 0.83$, df $= 2$, $p = 0.66$).

Network structure

I ran 800 simulations over 4 transmission values ($\beta = 0.1, 0.2, 0.3, 0.4$). The proportion of invasions was not different between highly connected and largely unconnected metapopulations (Figure 2.3BNetwork structure and diversity). This was true at all transmission levels (χ^2 test. $\beta = 0.1$: $\chi^2 = 5.23 \times 10^{-31}$, df $= 1$, $p = 1$. $\beta = 0.2$: $\chi^2 = 0.02$, df $= 1$, $p = 0.88$. $\beta = 0.3$: $\chi^2 = 1.53$, df $= 1$, $p = 0.22$. $\beta = 0.4$: $\chi^2 = 0.26$, df $= 1$, $p = 0.61$).

Transmission

Inline with theory, increasing the transmission rate increased the probability of invasion (Figure 2.3ANetwork structure and diversity–2.3BNetwork structure and diversity). This is true for all three dispersal values (χ^2 test. $\lambda = 0.001$: $\chi^2 = 243.56$, df $= 3$, $p = 1.62 \times 10^{-52}$. $\lambda = 0.01$: $\chi^2 = 265.4$, df $= 3$, $p = 3.05 \times 10^{-57}$. $\lambda = 0.1$: $\chi^2 = 244.95$, df $= 3$, $p = 8.1 \times 10^{-53}$) and both network structures (χ^2 test. Fully connected: $\chi^2 = 267.3$, df $= 3$, $p = 1.18 \times 10^{-57}$. Minimally connected: $\chi^2 = 267.3$, df $= 3$, $p = 1.18 \times 10^{-57}$).

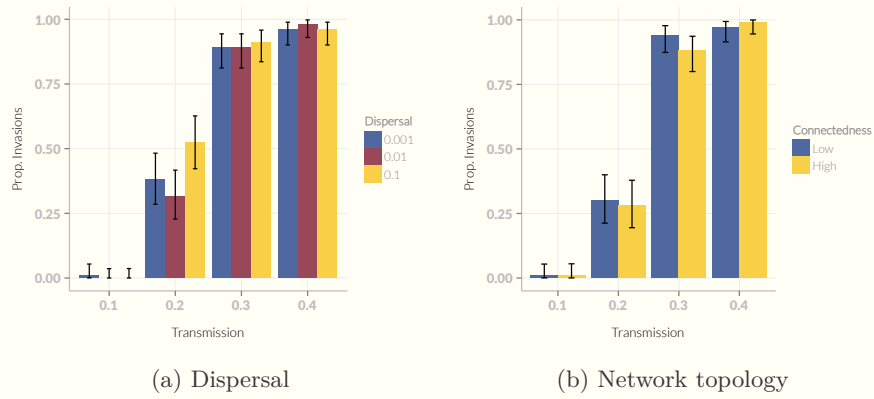(a) Dispersal   (b) Network topology

Figure 2.3: The probability of successful invasion. For three different transmission rates, the probability of invasion does not change between different a) dispersal rates or b) network structures. Error bars are 95% confidence intervals. 100 simulations were run for each treatment. Other parameters are kept constant at: $N = 10$, $b = d = 0.05$, $\gamma = 0.1$, $\alpha = 0.1$. When dispersal is varied, the population structure is fully connected. When population structure is varied, $\lambda = 0.01$.

Discussion

Restate the gap and the main result

Empirical studies on the role of population structure on the are equivocal and cannot examine the specific mechanisms by which pathogen communities are created and maintained. I have used mechanistic, metapopulation models to test whether increased population structure can promote pathogen richness by facilitating invasion of new pathogens.

Link results to consequences

Population structure does not affect pathogen richness

Probably because dynamics are dominated by local processes. This goes against many predictions that increasing R_0 increases pathogen richness. Further work could examine reduced col ony sizes to test when global structure become more important. Measures of population structure should not be used to predict zoonotic potential.

Dispersal does not affect pathogen richness

Network connectedness does not affect pathogen richness

This is in direct contrast to []. However, the model in [] is a contact network, so increasing the connectedness increases the chance of succesful transmission events for the first few transmission generations. This lends support to the idea that I found now affect of connectedness due to the dominance of local dynamics.

Network connectedness can be seen as a function of average dispersal distance, density and colony size. A high density species with small colony sizes must have colonies relatively close together. Therefore colonies would be more likely to be connected for a given dispersal distance.

Discuss 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 probably a

fairly reasonable assumption given that I am modelling a newly evolved strain. However, further work could relax this assumption using a model similar to [] 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 [] contains 3^ρ classes for a system with ρ pathogen species. This quickly becomes computationally restrictive.

Identical strains

Many papers on pathogen richness have focussed on the evolution of pathogen traits and have considered a trade off between transmission rate and virulence [,] or infectious period []. However, here we are 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.
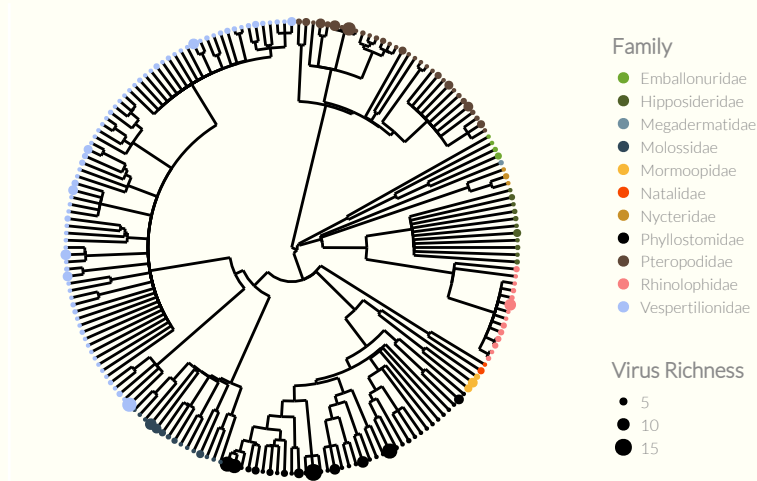
Appendix

	Explanation	Units	Value
S	Susceptible individuals		
I_q	Infectious with diseases q		
I_p^+	Sum of classes infected with pathogen p		
N	Number of colonies		10
$\bar{n}$	Mean colony starting size		3000
β	Transmission rate	Transmission events per year per individual	2, 5, 10
γ	Recovery rate	Recovery events per year per individual.	1
λ	Dispersal	Dispersal events per day per individual	0.001–0.1
b	Birth rate	Births per year per individual	0.05
d	Death rate	Deaths per year per individual	0.05
d_I	Infectious death rate	Additional deaths per day per individual	
ρ	No. pathogens		2
p	Pathogen index i.e. $p \in \{1,2\}$ for pathogens 1 and 2		
q	Disease class i.e., $q \in \{1,2,12\}$		
$\mathscr{V}$	Neighbourhood of a node		
t,t'	Time and time plus waiting time i.e., $t+\delta$	Days	
k_i	Degree of node i		
δ	Waiting time until next event	Days	
α	Cross immunity	Proportion	0.1
n,m	Colony index		
$\boldsymbol{A}_{mn}$	Adjacency matrix.	Distance	
μ	Maximum distance for edge to exist	km	40, 100
σ	Invading pathogen seed size		10
r_i	The rate that event i occurs.	Days^{-1}	

Table 2.2: All symbols used.

Chapter 3

Does social structure affect viral diversity in wild bat populations?

Family
- Emballonuridae
- Hipposideridae
- Megadermatidae
- Molossidae
- Mormoopidae
- Natalidae
- Nycteridae
- Phyllostomidae
- Pteropodidae
- Rhinolophidae
- Vespertilionidae

Virus Richness
- 5
- 10
- 15

Abstract

It is still unclear what factors determine the number of pathogens a wild species carries. But once understood, these factors could provide a way to prioritise surveillance of wild populations for zoonotic disease.

The pattern of contacts between individuals (i.e. population structure) has long been known to strongly affect epidemic processes. Theory suggests that population structure can promote pathogen richness while the ecological literature generally assumes it will decrease richness. Previous studies in wild populations have had contradictory results and the different measures of population structure have different shortcomings.

Here I use comparative data to test whether population structure influences pathogen richness in bats as they have been associated with a number of important, recent zoonotic outbreaks. Unlike previous studies I use two mea-

sures of population structure: a novel measure, number of subspecies, and a more careful application of genetic measures which have been used previously.

I find that both of these measures are associated with pathogen richness but with effects in opposite directions suggesting that population structure has no clear effect.

The results conflict with each other and with other studies which suggests that tests of population structure are sensitive to the exact measurements and data used. Given the conflicting results in the literature and unclear results here, it seems likely that population structure does not strongly affect pathogen richness in bats.

The use of larger datasets and and multiple measurements of population structure is therefore important to ensure the robustness of results. Given the weakness of any association between population structure and pathogen richness in bats, this is not a useful metric for prioritising zoonotic disease surveillance.

Introduction

General Intro

The number of pathogen species carried by a host species has important consequences for the ecology of the host and the probability that the host will be a reservoir of a zoonotic pathogen. However, the factors that affect pathogen richness are poorly understood.

Specific Intro

Theoretical background

Single pathogen models show that increasing population structure simply slows disease spread and makes establishment less likely [colizza2007invasion, vespignani2008reaction]. In the ecological literature this is often taken as predicting that increased population structure will decrease pathogen richness []. However, models of competition between multiple pathogens show that in unstructured populations a competitive exclusion process occurs but that splitting the population into two patches allows coexistence [,].

Previous Studies

Three studies have used comparative data to test for an association between population structure and viral richness. A study on 15 African bats found a positive relationship between distribution fragmentation and viral richness [] while a study on 20 South-East Asian bats found the opposite relationship []. A global study on 33 bats found a positive relationship between F_{ST} — a measure of genetic structure — and viral richness []. However, this study included measures using mtDNA which only measures female dispersal which may have biased the results many bat species show female philopatry [,]. Furthermore, this study used measures of F_{ST} irrespective of the study scale with studies covering from tens [] to thousands [] of kilometers. As isolation by distance has been shown in a number of bat species [,] this could bias results further. Finally, when a global F_{ST} value is not given they use the mean of all pairwise F_{ST} between sites. It is not clear that this is correct as from global F_{ST} we expect migration rates of $M = \frac{1-F_{ST}}{8F_{ST}}$ while from F_{ST} between pairs of populations we expect migration rates of $M = \frac{1-F_{ST}}{16F_{ST}}$ where M is the absolute number of diploid inviduals dispersing per generation []. As it is in fact the movement of individuals that is epidemiologically relavent, using these studies is probably not correct without attempting to correct for these difference.

Studies on single pathogens, notably rabies, have also shown that for virulent pathogens, space can allow persistence where a well mixed population ould experience a single, large epidemic then pathogen extinction [blackwood2013resolving, colizza, pons2014insights].

Rates

Choice of measure of population structure

A number of measurments of population structure have been used and each has it's own shortcomings. In particular, the better, more direct measurements tend to be very work intensive which consequently means data is available for few species.

The ideal measurement of population structure is direct measurement of dispersal rates and distance. These are incredibly difficult to obtain, especially over large scales. Due white nose syndrom, some very large mark-recapture

studies have been conducted, but recapture rates are low. Further, these large studies have been in species the live in a fe large colonies, so recapture rates should be higher than in less social species.

As direct measurement of dispersal are difficult, genetic data is often used. Measurement such as F_{ST} are used to calculate migration. There are strong model assumptions under the conversion from F_{ST} to migration. However, the main issue with this measure is the effort required for each study and the subsequent low number of measurements. Further, there are differences in the scales of the studies and the genetic regions being sequenced. This differences should not be ignored.

For a population to evolve distinct phenotypic or genetic traits, such that they can be classed as a subspecies, there must be limited migration between populations. The number of subspecies a species has therefore reflects the level of population structure in that species. The value of this measurement is available for every bat species. However, it is likely biased, with well studied species being likely to have more recognised subspecies. Further, this is a very course measure and it is important to consider whether it is measuring migration at a timescale and rate that is epidemiologically relevant.

The final measurement that has been used is derived from the shape of the species' range, typically from IUCN [iucn] maps. The ratio between the perimeter of the range and the area (or similar values) are calculated. Range maps are very course for many species. Furthermore there is a potential bias with island living species being given sea based edges where continental species might be assumed to live everywhere in between locations where it is known to live, without considering the different terrestrial habitats in these areas.

The gap

There is a lack of studies using multiple measures of population structure and larger datasets to robustly estimate the importance of population structure. Furthermore, the

What I did

Here I use two measures of population structure — the number of subspecies and gene flow — to robustly test for an association between population struc-
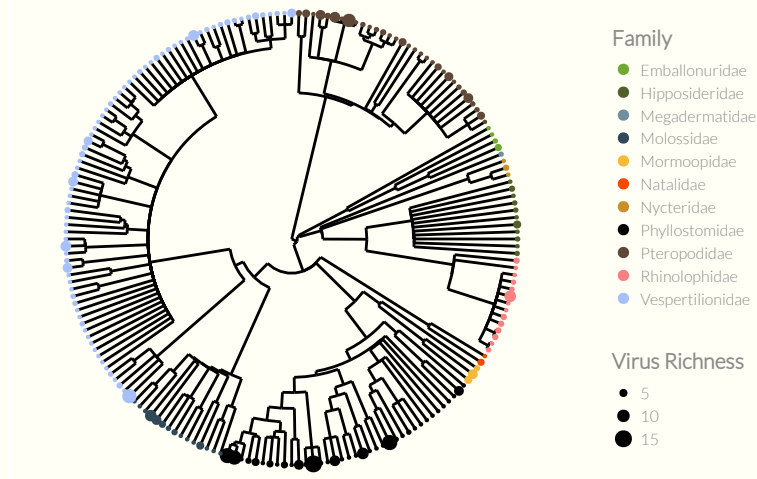
Figure 3.1: Pruned phylogeny with dot size showing number of pathogens and colour showing family.

ture and pathogen richness. Furthermore, I use a much larger dataset for one of these analyses, further promoting robustness of results.

What I found

Methods

To measure pathogen richness I used data from []. These simply include known infections of a bat species with a pathogen species. Only species with at least one pathogen were included in the analysis. Rows with host species that were not identified to species level were removed. Many viruses were not identified to species level or their identified species was not in the ICTV virus taxonomy []. I counted a virus if it was the only virus, for that host species, in the lowest taxonomic level identified in the ICTV taxonomy. That is, if a host carries an unknown Paramyxoviridae virus, then it must carry at least one Paramyxoviridae virus. If a host carries an unknown Paramyxoviridae virus and a known Paramyxoviridae virus, then it is hard to confirm that the unknown virus is not another record of the known virus. In this case, this would be counted as one virus species.

I used two measures of population structure. F_{ST} and the number of subspecies. The number of subspecies was counted using the Wilson and Reeder taxonomy []. F_{ST} and other measures were collated from the literature. Studies are from a wide range of spatial scales, from local (~ 10km) to continental. As

F_{ST} inevitably increases with spatial scale 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 (measured with http://www.distancefromto.net/ if not stated) to the width of the IUCN species range and only used studies if this ratio was greater than 0.2. To allow comparison between different measures (F_{ST}, ϕ_{ST}) and data from different molecular regions I converted all data to diploid gene flow. WILL ADD EXTRA METHODS LATER. These two measures of population structure were analysed separately as the number of subspecies has 196 data points while there is only F_{ST} data for ~ 30 bat species.

To control for study bias I collected the number of Pubmed and Google Scholar citations for each bat species including synonyms from ITIS [] via the taxize package []. The counts were scraped using the rvest package []. I log transformed these variables as they were strongly right skewed. The log number of citations on Pubmed and Google scholar were highly correlated (pgls: $t = 19.32$, df $= 194$, $p = 0$). The results here are for analyses using only Google Scholar citations.

Measures of body mass are taken from Pantheria [] and primary literature [,   ]. Pipistrellus pygmaeus was assigned the same mass as P. pipistrellus as they indistinguishable by mass. Body mass measurements were log transformed due to the strong right skew. Distribution size was estimated by downloading range maps for all species from IUCN [iucn] and were also logged due to right skew.

To control for phylogenetic nonindependance I used the best-supported phylogeny from [] which is the supertree from [] with names updated to match the Wilson & Reeder taxonomy []. Phylogenetic manipulation was performed using the ape package []. The importance of the phylogeny on each variable separately was estimated using [].

F_{ST} studies are conducted at a range of spatial scales, but F_{ST} often increases with distance studied [,   ]. To minimise the effects of this I only used data from studies that cover 20% of the diameter of the species range. This is a largely arbitrary value that could be considered to reflect a "global" estimate of F_{ST} while keeping a reasonable number of datapoints available. I calculated the diameter of the species range by finding the furthest apart points in the
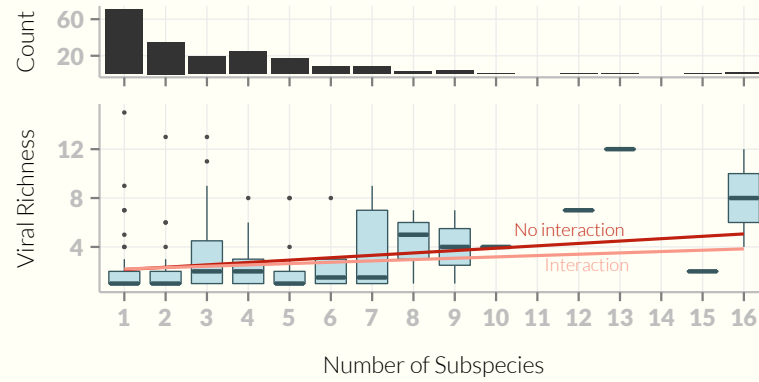
Figure 3.2: Number of virus species against number of subspecies. Data within a number of subspecies are plotted as boxplots with the dark bar showing the median, the box showing the interquartile range, vertical lines showing the range and outliers shown as seperate points. Regression lines are from multivariate phylogenetic models with other independant variables set at their median value. The models shown are thos with (pink) and without (red) an interaction between study effort and number of subspecies.
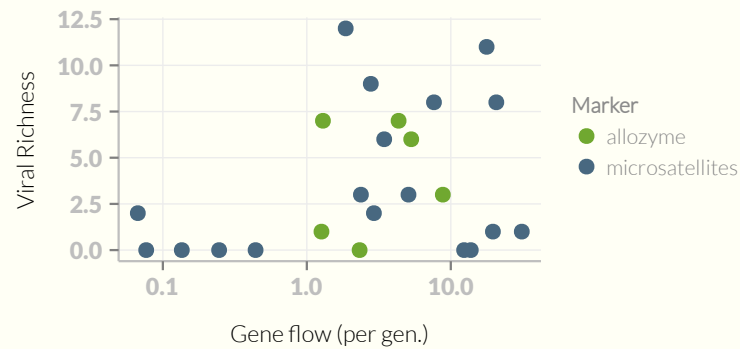

Figure 3.3: Gene flow per generation (on a log scale) against viral richness with the genetic marker used shown with colour.

IUCN species range [iucn] even if the range is split into multiple polygons. The width covered by each study was the distance between the most distant sampling sites. When this was not explicit in the paper, the centre of the lowest level of geographic area was used.

Statistical analysis

Statistical analysis for both dependant variables was conducted using a information theory/model averaging approach [] specifically following [,]. I chose a credible set of models including all combinations of independent variables. In the analysis using the number of subspecies dependant variable I also in-
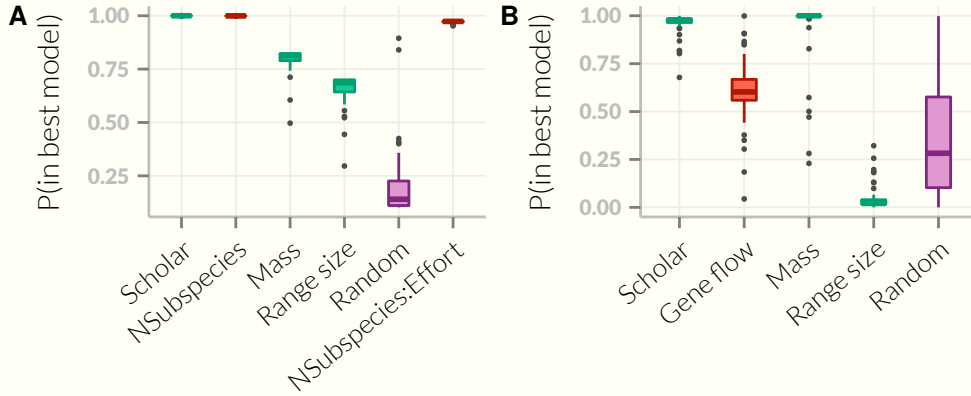
Figure 3.4: Akaika variable weights for both analyses. The probability that each variable will be in the best model if the data were recollected is shown for each of the bootstrap analyses. The purple "Random" box is a uniform random variable used as a null. Population structure (Number of subspecies and Gene flow), shown in red, is likely to be in the best model in both analyses.

cluded an interaction term between study effort and number of subspecies as I believe a priori that this interaction may be present. The interaction was only included in models with both study effort and number of subspecies as an individual term.

I fitted phylogenetic regressions using nlme [] to all models. In each case I simultaneously fitted the λ parameter as this avoids mispecifying the model []. κ and δ were constrained to one.

I calculated AICc for each model and then calculated Akaiki weights. This value can be interpreted as the probability that a model would be the best model if the data were recollected. For each variable, the sum of the Akaiki weights of models containing that variable are summed. This value can be interpreted as the probability that the given variable is in the best model. Following [] I included a uniformally random variable as a null variable as even unimportant variables can have Akaiki weights notably greater than zero. The whole analysis was run 50 times, resampling the random variable each time. I calculated the average AICc, $\overline{\text{AICc}}$, by averaging AICc scores within models. ΔAICc was calculated as $\min(\overline{\text{AICc}}) - \overline{\text{AICc}}$, not the mean of the individual ΔAICc scores, to guarantee that the best model has ΔAICc = 0.

Results

Number of Subspecies

After data cleaning there was data for 196 bat species in 11 families. The number of described virus species for a bat host ranged up to 15 viruses in Carollia perspicillata. Figure 3.1Pruned phylogeny with dot size showing number of pathogens and colour showing family shows the phylogeny used and the number of viruses for each species. The mean number of viruses across families is fairly constant with a lower range of 1.67 for Nycteridae. The highest mean is Mormoopidae with 5 virus species per bat species, but this is based on a sample size of 3. The Phyllostomidae have the second highest mean (n = 37) of 3.49.

The small change in mean pathogen richness across families and the lack of clear pattern in Figure 3.1Pruned phylogeny with dot size showing number of pathogens and colour showing family implies that viral richness is not strongly phylogenetic. This is corroborated by the small estimated size of λ ($\lambda = 0.04$, $p = 0.21$). This fact implies that other factors must control pathogen richness. It also implies that pathogens are not directly inherited down the phylogeny, although this is to be expected by the fast evolution of viruses.

Of the explanatory variables, the number of subspecies has no phylogenetic autocorrelation ($\lambda = 10^{-6}$, $p = 1$), study effort and distribution size have weak but significant autocorrelation (Study Effort: $\lambda = 0.1$, $p = 9.12 \times 10^{-3}$, Distribution size: $\lambda = 0.32$, $p = 8.82 \times 10^{-8}$) and mass is strongly phylogenetic ($\lambda = 0.93$, $p = 0$). The parameter λ is fitted and governs how important the phylogeny is in the model. Across all models the mean value of λ is 0.08 implying the residuals from the model are weakly phylogenetic. A small number of models (0.35%) had negatively phylogenetically distributed residuals.

Summing the Akaiki weights of all models that contain a given variable gives a probability, Pr, that the variable would be in the best model (Figure 3.7Akaika variable weights for F_{ST} analysis.a) if the data were recollected []. The number of subspecies is very likely in the best model ($Pr > 0.99$) as is the interaction between number of species and study effort ($Pr = 0.97$) compared to the benchmark random variable which has $Pr = 0.2$. When models

with the interaction term are removed, on average (mean weighted by Akaiki weights) there is a positive relationship between the number of subspecies and viral richness ($\beta = 0.19$, variance $= 1.24 \times 10^{-4}$). Models with an interaction between number of subspecies and study effort have a positive interaction term ($\beta = 0.14$, variance $= 3.94 \times 10^{-6}$) and a negative linear term ($\beta = -0.78$, variance $= 2.37 \times 10^{-4}$) which in total is a positive relationship for most of the range of the study effort data. This supports the hypothesis that population structure promotes pathogen richness. The strong support for a positive interaction term implies that strong population structure only predicts high known pathogen richness in the presence of high study effort. One interpretation of this is that population structure alone does not predict high known richness; reasonable study effort is still needed to turn the expected high richness into known and recorded viral richness. Another interpretation is that having few subspecies does not predict low viral richness unless the species has been suitable sampled; the low number of subspecies is probably due to a lack of study rather than an accurate measurement. As seen in Figure 3.7Akaika variable weights for F_{ST} analysis.a, study effort is very likely in the best model ($\beta = 0.35$, $Pr > 0.99$) body mass and range size are also probably in the best model ($\beta = 0.38$, $Pr = 0.8$ and $\beta = 0.26$, $Pr = 0.66$).

Gene Flow

Due to the low number of studies and the restrictive requirements imposed on study design, there are only data for 24 bat species in 7 families. The number of described virus species for a bat host ranged up to 12 viruses in Miniopterus schreibersii. Figure 3.5Pruned phylogeny with dot size showing number of pathogens and colour showing family shows the phylogeny used and the number of viruses for each species.

As with the Number of Subspecies dataset, there is no phylogenetic signal in the number of virus species ($\lambda = 10^{-6}$, $p = 1$) Gene flow also has no phylogenetic autocorrelation ($\lambda = 10^{-6}$, $p = 1$). Due to the low sample size, significance tests are unlikely to have much power. However, study effort has some phylogenetic autocorrelation ($\lambda = 0.15$, $p = 0.56$) while distribution size and mass seem to show phylogenetic signal (Distribution size: $\lambda = 0.67$, $p = 0.53$, Mass: $\lambda = 0.79$, $p = 2.69 \times 10^{-3}$).

Table 3.1: Estimated variable weights (probability that a variable is in the best model) and their estimated coefficients.

(rl)2-3(rl)4-5 Variable	Number of Subspecies		Gene flow	
	Weight	Coefficient	Weight	Coefficient
Number of subspecies				
Total	1.00	-0.75		
Models without interaction term		0.19		
Models with interaction term		-0.78		
Number of subspecies:log(Scholar)	0.97	0.14		
Gene flow			0.61	0.02
log(Scholar)	1.00	0.35	0.96	0.07
log(Mass)	0.80	0.38	0.94	0.13
log(Range size)	0.66	0.26	0.05	0.21
Random	0.20	0.17	0.37	12.56

While less strongly supported than the number of subspecies, gene flow is likely in the best model ($Pr = 0.61$) compared to the benchmark random variable which has $Pr = 0.37$ as shown in Figure 3.7Akaika variable weights for F_{ST} analysis.b. On average (mean weighted by Akaiki weights) there is a negative relationship between gene flow and viral richness ($\beta = $ NA, variance $=$ NA) despite the apparent positive relationship in Figure 3.6Gene flow per generation (on a log scale) against viral richness with the genetic marker used shown with colour. This supports the hypothesis that population structure promotes viral richness. Possibly due to the smaller sample size, or a weaker relationship, this coefficient is much more varied than the number of subspecies coeficient with 24.25% of models estimating a positive relationship. As in the number of subspecies analysis, study effort is very likely in the best model ($Pr = 0.96$) as is body mass ($Pr = 0.94$). In contrast to the number of subspecies analysis, range size is almost certainly not in the best model with $Pr = 0.05$ being much less than the random variable.

Across all models the mean value of λ is -2.88 and a large number of individual models (66.58%) had negatively phylogenetically distributed residuals implying the residuals from the model are strongly negatively phylogenetic.

Table 3.2: Model selection results for number of subspecies analysis. $\bar{\text{AIC}}$c is the mean AICc score across 50 resamplings of the null random variable. ΔAICc is the $\bar{\text{AIC}}$c score minus the lowest score. w_i is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w_i$ is the cumulative sum of the Akaike weights.

Model	$\bar{\text{AIC}}$c	ΔAICc	w_i	$\sum w_i$
log(Scholar)*NSubspecies + rand	882	0.00	0.38	0.38
log(Scholar)*NSubspecies	884	1.39	0.19	0.57
log(Scholar)*NSubspecies + rand + log(Mass)	885	2.24	0.12	0.70
log(Scholar)*NSubspecies + log(Mass)	885	3.14	0.08	0.78
log(Scholar)*NSubspecies + log(Mass) + log(RangeSize)	886	3.18	0.08	0.86
log(Scholar)*NSubspecies + rand + log(RangeSize)	886	3.94	0.05	0.91
log(Scholar)*NSubspecies + log(RangeSize)	886	3.95	0.05	0.96
log(Scholar) + NSubspecies	889	6.93	0.01	0.97
log(Scholar) + NSubspecies + log(Mass)	890	7.80	0.01	0.98

Table 3.3: Model selection results for gene flow analysis. The top two models are considerably better than the other models; there is a 98% chance that one of them is the best model. As body mass and study effort are in both these models these variables are very likely in the best model. Gene flow is in the top model, implying it may well be in the best model. $\bar{\text{AIC}}$c is the mean AICc score across 50 resamplings of the null random variable. ΔAICc is the $\bar{\text{AIC}}$c score minus the lowest score. w_i is the Akaike weight and can be interpreted as the probability that the model is the best model (of those in the plausible set). $\sum w_i$ is the cumulative sum of the Akaike weights. '*' indicates the interaction term and both linear terms.

Model	$\bar{\text{AIC}}$c	ΔAICc	w_i	$\sum w_i$
log(Scholar) + log(Gene flow) + log(Mass)	70	0.00	0.60	0.60
log(Scholar) + log(Mass)	71	0.94	0.38	0.98
log(Mass) + log(Range size)	77	6.73	0.02	1.00
log(Scholar) + log(Gene flow) + log(Mass) + rand	94	24.15	0.00	1.00

Due to the small sample size this is probably due to a small number of data points with large residuals being distant on the tree.

Discussion

By analysing data on two measures of population structure, and using larger datasets than previous studies, I have tested the hypothesis that population structure promotes pathogen richness in bats. I have found that a positive effect of the number of subspecies and a negative effect of gene flow are likely to be in the best models for explaining viral richness. Study effort is also clearly supported confirming the expectation that additional study of a bat species yields more known viruses infecting that species.

Appendix

Chapter 4

Does population size affect viral diversity in bat populations?

Abstract

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

Population density is known to increase pathogen richness while theory suggests that population structure may also play a role. However, these factors, along with population abundance, are intrinsically linked; reducing density reduces contacts between individuals. In group living species, this is particularly true, with group size and the number of groups and distribution size contributing to total animal density. As these factors are completely interdependant, 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 population structure, group size or population size (abundance).

Here I use metapopulation SIR models to test whether it is density per se that increases the ability of a new pathogen to invade as apposed to colony size, population abundance or population structure.

Introduction

General Intro

Why is pathogen diversity important?

We know some factors that correlate with pathogen diversity

But we do not understand the mechanistic processes

Specific Intro

The gap

What I did

What I found

Methods

Metapopulation model

Two pathogen SIR model

We examine a multpathogen SIR model. This is a compartment model with individuals being classed as susceptible, infected or recovered with immunity (Figure 4.1The SIR model used.). Susceptible individuals are counted in class S. 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 a pathogen moves an individual straight into the recovered class, even if the individual is infected with both pathogen. This modelling choice allows the model to be easily expanded to included more than two pathogens. The assumption of immediate recovery from all other diseases is likely to be quite accurate for very closely related pathogens as is being studied here as once an acquired immune response is activated, all infections are likely to be cleared quickly.

The coinfection rate is adjusted compared to the first infection rate by a factor α. Birth and death rates are assumed to be equal, $b = d$.

Metapopulation

The population is divided into a number of subpopulations. This metapopulation is modelled as a network with subpopulations being nodes and dispersal between subpopulations being indicated by edges (Figure 2.2Network topologies used to compare network connectedness) Individuals with a subpopulation interact randomly so that the subpopulation is fully mixed. However, dispersal between subpopulations occurs at a rate λ. Individuals can only disperse to subpopulations connected to theirs in the network. The rate of dispersal is not affected by the number of edges a subpopulation has (the degree of the subpopulation). So the dispersal rate from a subpopulation m with degree k_m to subpopulation n is $\frac{\lambda}{k_m}$. Note this rate is independent of the degree of subpopulation n.
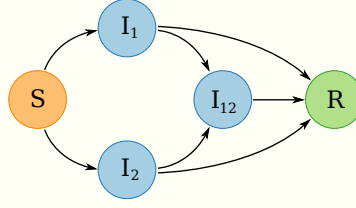
Figure 4.1: The SIR model used.

Stochastic simulations

We examine this model using stochastic, continuous time simulations using the Gillespie algorithm. At each step in the simulation we calculate the rate that each possible event might occur. One event is then randomly chosen, weighted by it's rate

$$p(\text{event } i) = \frac{r_i}{\sum_i r_i} \tag{4.1}$$

where r_i is the rate that event i occurs. Finally, the length of the time step, δ, is drawn from an exponential distribution

$$\delta \sim \text{Exp}\left(\sum_i r_i\right). \tag{4.2}$$

This means that the length of each simulation is stochastic. We define the number of events we wish to simulate instead.

We can now write down the rates of all events. I define I_p^+ to be the sum of all classes that are infectious with pathogen p, for example $I_1^+ = I_1 + I_{12}$. Assuming asexual reproduction, that all classes reproduce at the same rate and that individuals are born into the susceptible class we get

$$P(S_{nt'} = S_{nt} + 1) = b\left(S_{nt} + \sum_q I_{qnt} + R_{nt}\right) \tag{4.3}$$

where $P(S_{nt'} = S_{nt} + 1)$ is the probability that the number of susceptibles in subpopulation n will increase by 1 (a single birth) the short time interval t to t' and $\sum_q I_{qnt}$ is the sum of all infection classes $q \in 1, 2, 12$. The rates of death, given a death rate d are given by

$$P(S_{nt'} = S_{nt} - 1) = dS_{nt} \tag{4.4}$$

$$P(I_{qnt'} = I_{qnt} - 1) = dI_{qnt} \tag{4.5}$$

$$P(R_{nt'} = R_{nt} - 1) = dR_{nt}. \tag{4.6}$$

Infection of a susceptible with either pathogen 1 or 2, $S \to I_p$ where $p \in \{1,2\}$, is given by

$$P\left(I_{pnt'} = I_{pnt} + 1, S_{nt'} = S_{nt} - 1\right) = \beta S_{nt} I_{pnt}^{+}, \tag{4.7}$$

while coinfection, given a crossimmunity factor α, is given by

$$P\left(I_{12,nt'} = I_{12,nt} + 1, I_{pnt'} = I_{pnt} - 1\right) = \alpha \beta I_{nt} I_{pnt}^{+}. \tag{4.8}$$

The probability of migration from colony m (with degree k_m) to colony n, given a dispersal rate λ is given by

$$P\left(S_{nt'} = S_{nt} + 1, \, S_{mt'} = S_{mt} - 1\right) = \frac{\lambda S_{mt}}{k_m - 1} \tag{4.9}$$

$$P\left(I_{qnt'} = I_{qnt} + 1, \, I_{qmt'} = I_{qmt} - 1\right) = \frac{\lambda I_{qmt}}{k_m} \tag{4.10}$$

$$P\left(R_{nt'} = S_{nt} + 1, \, R_{mt'} = R_{mt} - 1\right) = \frac{\lambda R_{mt}}{k_m}. \tag{4.11}$$

Finally, recovery from any infectious class occurs at a rate γ

$$P\left(I_{qnt'} = I_{qnt} - 1, \, R_{nt'} = R_{nt} + 1\right) = \gamma I_{qnt}. \tag{4.12}$$

In each simulation the population is seeded with 200 infected individuals of disease 1 in each colony. Disease 1 is then allowed to spread and reach equilibrium. After 7×10^5 events, 10 individuals infected with disease 2 are added to one colony. After another 3×10^5 events the invasion of disease 2 is considered successful if any individuals with disease 2 still remain.

Dispersal

The values used for the independant variables are chosen to highlight the affects of these variables. Dispersal values are $\lambda = 0.1, 0.01$ and 0.001 dispersals per individual per year. $\lambda = 0.1$ relates to individuals moving between colonies on average twice per lifetime. Therefore exclusively juvenile dispersal would have dispersal rates similar to this. Otherwise it relates to dispersal being a rare event with animals often staying in a colony for many years. $\lambda = 0.01$ relates to 20% of individuals dispersing once in their lifetime. This value is therefore close to male-biased dispersal, with female philopatry. Finally, $\lambda = 0.001$ relates to 2% of individuals dispersing in their lifetime. This therefore relates to a population that does not habitually disperse.

Network structure

The network structure is synthetically created to be either fully or minimally connected (See Figure 2.2Network topologies used to compare network connectedness). 10 subpopulations was selected as a trade off between computation time and a network complicated enough that structure might have an effect. This value is artificially small compared to wildlife populations.

Parameter selection

The fixed parameters used are chosen to roughly reflect realistic wild bat populations. The death rate d is set as 0.05 per year giving a generation time of 20 years. The birth rate b is set to be equal to d so that the population size is stable. The recovery rate γ is set to 0.1 giving a average infection duration of 10 years. This is therefore a chronic infection. It is very difficult to directly estimate infection durations in wild populations. But it seems that these infections might be long lasting [].

Cross immunity is set to 0.1 so that an individual infected with one disease is 90% less likely to be infected with another. This is a rather arbitrary value. However, the model assumes complete cross immunity after infection. Furthermore, the rationale of the model is that the invading species might be a newly speciated strain of the endemic species. Therefore cross immunity is likely to be very strong.

The population size of each subpopulation is set to 3000. This is appropriate for many bat species [], especially the large, frugivorous Pteropodidae that have been particularly associated with recent zoonotic diseases.

Three values of the transmission rate β are used, 2, 5 and 10. All simulations are run under all three transmission rates as this is such a fundamental parameter. Given the recovery, birth and death rates we can calculate an approximation of R_0 that ignores spatial structure. That is, this is R_0 for the local, within-subpopulation dynamics. Furthermore, it is R_0 for the first pathogen; R_0 of the invading pathogen will be lower due to competition. We can calculate that $R_0 \approx \frac{\beta d}{d(d+\gamma)}$. For our three values of $\beta = 2, 5, 10$ we therefore get $R_0 \approx 13.3, 33.3, 66.6$. These values are very high in part to again find a reasonable trade off between the number of simulations and the reasonableness

of the parameters. $R_0 \approx 13.3$ is similar to a highly contagious disease such as measles or pertussis.

For the simulations where an invading pathogen is added to the populations the number of invading pathogens added is set to 10. This is a trade off between getting a reasonable proportion of invasions, while still retaining the stochastic nature of invasion.

For simulations studying extinction rates, half of the subpopulation were seeded with each pathogen. 2000 susceptibles and 1000 infected individuals were placed in each subpopulation in order to quickly reach equilibrium.

Results

Chapter 5

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

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 abundance 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.

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 [] and extinction risk []. Monitoring animal population changes in response to anthropogenic pressure is becoming increasingly important as humans rapidly modify habitats and change climates []. Sensor technology, such as camera traps [,] and acoustic detectors [,] 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 [,]. However, converting sampled count data into estimates of density is problematic as detectability of animals needs to be accounted for [].

Existing methods for estimating animal density often require additional information that is often unavailable. For example, capture-mark-recapture methods [,] require recognition of individuals, and distance methods [] require estimates of how far away individuals are from the sensor [,]. When individuals cannot be told apart, an extension of occupancy modelling can be used to estimate absolute abundance []. 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 [] which may be a poor assumption for nonrandomly dis-

tributed species. Furthermore repeat, independent surveys must be performed and the definition of a site can be difficult, especially for wide-ranging species [].

The REM method has been successfully applied to estimate animal densities from camera trap surveys []. 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 [].

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, []). 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 [], 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 [], 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 []. Acoustic monitoring is being developed to study many aspects of ecology, including the interactions of animals and their environments [,], the presence and relative abundances of species [], biodiversity of an area [], and monitoring population trends [].

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 []. 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 []. In these situations distance-sampling methods can be applied []. 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 []. 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 [,]. 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 [], 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.

Methods

Analytical Model

The REM presented by [] 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 5.1Representation of sensor detection width and animal signal width). Additionally, we model the animal as having an associated signal width α between 0 and 2π (Figure 5.1Representation of sensor detection width and animal signal width, 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 [], 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
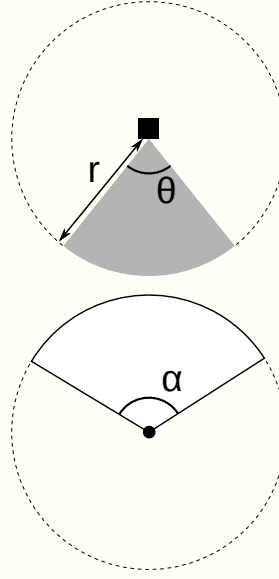
Figure 5.1: Representation of sensor detection width and animal signal width. The filled square and circle represent a sensor and an animal, respectively; θ, sensor detection width (radians); r, sensor detection distance; dark grey shaded area, sensor detection zone; α, animal signal width (radians). Dashed lines around the filled square and circle represents the maximum extent of θ and α, respectively.

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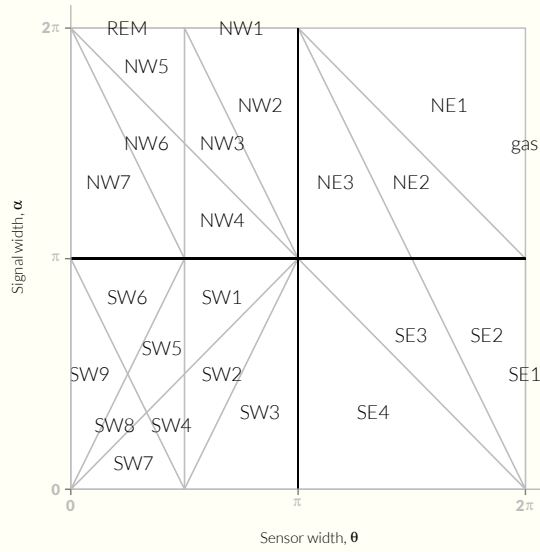
Figure 5.2: Locations where derivation of the average profile $\bar{p}$ is the same for differ-
ent combinations of sensor detection and animal signal widths. Symbols
within each polygon refer to each gREM submodel named after their com-
pass point, except for Gas and REM which highlight the position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

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 5.2Locations where derivation of the average profile $\bar{p}$ is
the same). For example, the gas model becomes the simplest gREM sub-model

(upper right in Figure 5.2Locations where derivation of the average profile $\bar{p}$ is the same) and the REM from [] is another gREM sub-model where $\theta < \pi/2$ and $\alpha = 2\pi$. We derive one gREM 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 [,]. Sensor width, θ, and detection distance, r would also need to be measured or obtained from manufacturer specifications [,].

Example derivation of SE2

In order to calculate $\bar{p}$, we have to integrate over the focal angle, x_1 (Figure 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2a). 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 sub-models (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 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2b). During this first interval, the size of α limits the width of the profile. When the animal reaches $x_1 = \theta/2 + \pi/2 - \alpha/2$ (Figure 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2c), 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 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2c). 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 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2d) 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 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2b–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

(a)

(b)

p = 2 r sin α/2
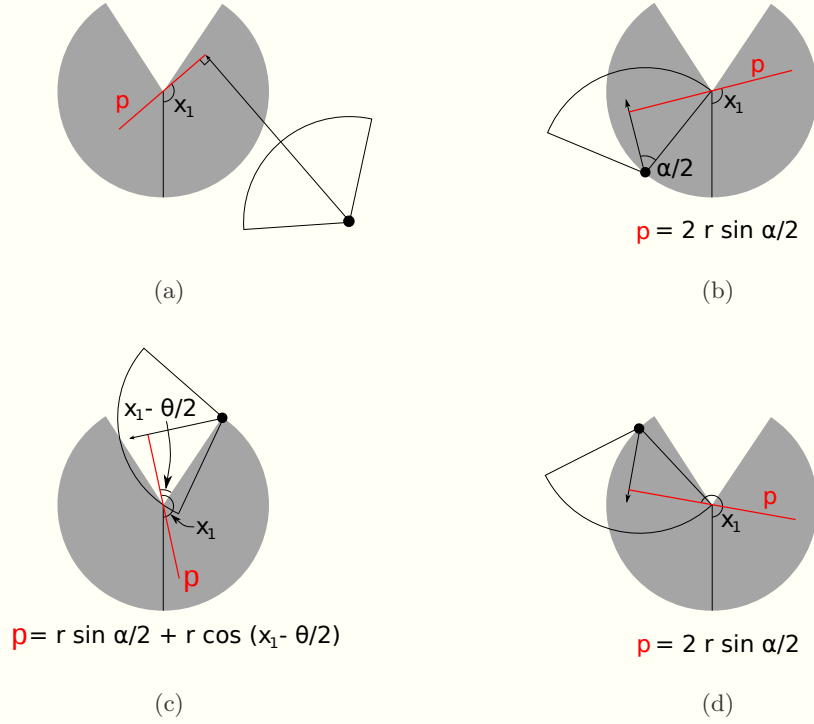
(c)

p= r sin α/2 + r cos (x₁- θ/2)

(d)

p = 2 r sin α/2

Figure 5.3: An overview of the derivation of the average profile $\bar{p}$ for the gREM sub-model 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.

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

(5.1)

$$= \frac{r}{\pi} \left(\theta\sin\frac{\alpha}{2} - \cos\frac{\alpha}{2} + \cos\left(\frac{\alpha}{2} + \theta\right) \right)$$

(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 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2c and 5.3An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2d 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 5.1gREM for estimating animal density would not give us this integral of $p = 0$.

gREM submodel specifications were done by hand, and the integration was done using SymPy [] 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 [] implementation of the gREM is given in Appendix S4.

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 $7.5\,\mathrm{km}$ by $7.5\,\mathrm{km}$ square with periodic boundaries. A stationary sensor of radius r, $10\,\mathrm{m}$, 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 $70\,\text{animals}\,\text{km}^{-2}$, calculated from the equation in [] as the expected density of mammals weighing $1\,\text{g}$. This density therefore represents a reasonable estimate of density of individuals, given that the smallest mammal is around $2\,\text{g}$ []. 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, $v = 40\,\text{km}\,\text{day}^{-1}$, was chosen based on the largest day range of terrestrial animals [], 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 5.2Locations where derivation of the average profile $\bar{p}$ is the same). 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%, ±5% and ±10%, for all gREM submodels.

Results

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 5.4Expressions for the average profile width shows the expression for $\bar{p}$ in each case. The general equation for density, 5.3gREM for estimating animal density, 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 5.4Expressions for the average profile width are equal when expressions for the boundaries between them are substituted in.
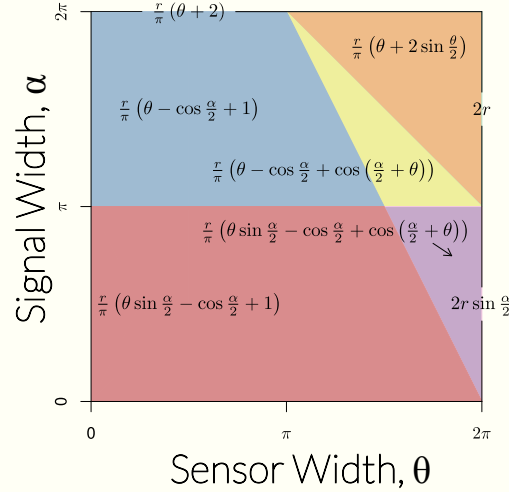
Simulation model

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 $\alpha, \theta = 2\pi$.

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 5.5Simulation model results of the accuracy and precision for gREM submodels). 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 5.5Simulation model results of the accuracy and precision for gREM submodels). 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 5.6Simulation model results of the accuracy and precision of four gREM submodels). However, the precision was dependent on
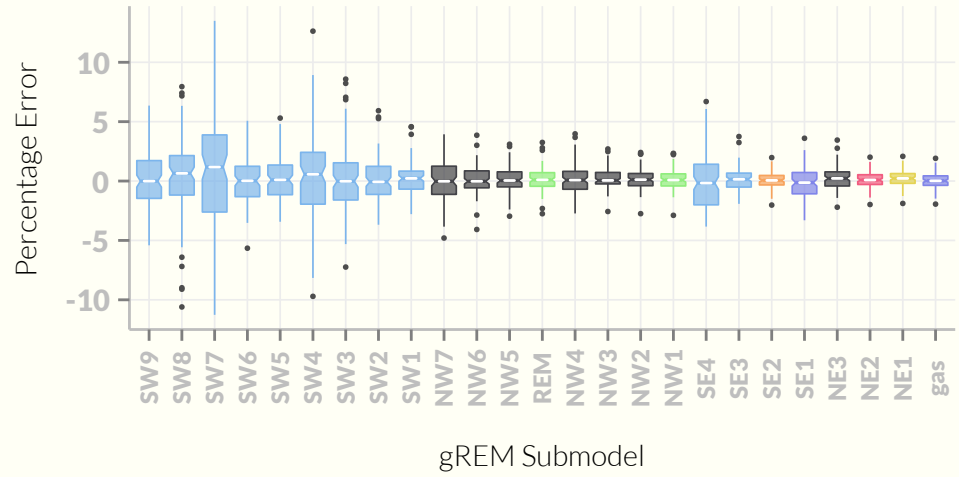
Figure 5.5: Simulation model results of the accuracy and precision for gREM sub-
models. The percentage error between estimated and true density for
each gREM sub model is shown within each box plot, where the white line
represents the median percentage error across all simulations, boxes rep-
resent 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 5.4Expressions for
the average profile width.

the number of captures across all four of the gREM submodels, where pre-
cision increases as number of captures increases, as would be expected for
any statistical estimate (Figure 5.6Simulation model results of the accuracy
and precision of four gREM submodels). For all gREM submodels, the the
coefficient of variation falls to 10% 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) (Fig-
ure 5.7AgREM for estimating animal density). Altering the maximum change
in direction in each step (0, $\pi/3$, $2\pi/3$, and π) did not affect the accuracy
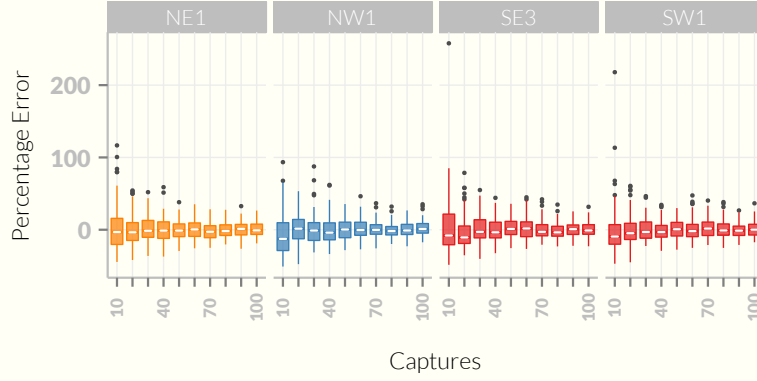or precision of the four gREM submodels (Figure 5.7BgREM for estimating
animal density).

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 5.4Expressions for the average profile width.

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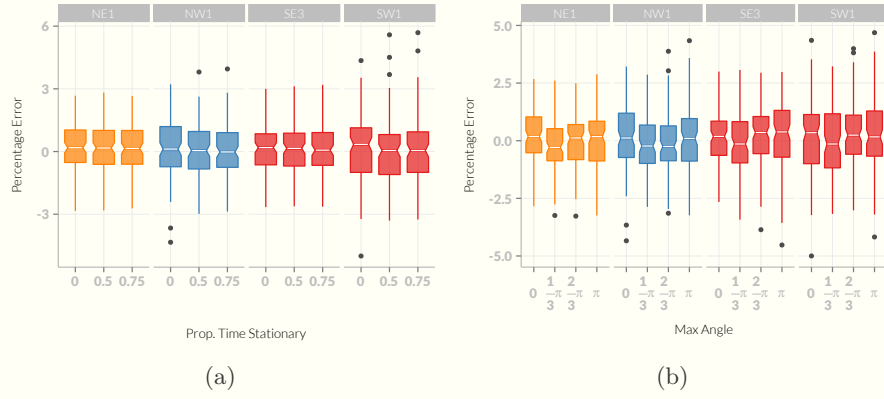
Discussion

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.

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 [] 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 both animals and sensor are moving []. 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 animal could occur, e.g. an acoustic detector towed from a boat when studying porpoises [] or surveying echolocating bats from a moving car [].

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) [], and secondly, sensors which repeatedly turn on an off during sampling []. 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.

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 [,]. 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 [], while the sensor angle is often easily measured [] 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 [], more work should be done on obtaining estimates for a range of acoustically surveyed species.

Limitations

Although the REM has been found to be effective in field tests [,], 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 []. 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 []. 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 [], it will affect precision and further simulations should be used to quantify this effect. Additionally, we allowed the sensor to be stationary and continuously detecting, 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 []. 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 [,   ]. 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.

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 [], songbirds [], whales [] and forest primates []. 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 []. 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 [].

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 []. The gREM is also suitable for species that are sensitive to human contact or are difficult or dangerous to catch []. 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.

Data Accessibility

The code used in this chapter is available on Github at `https://github.com/timcdlucas/lucasMoorcroftManuscript/tree/postPeerReview`.

Chapter 6

General Conclusions

My conclusions. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Appendix A

gREM Appendix

Table of symbols

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

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

Supplementary Methods

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 use in the gREM is included in the Python script S3.

Gas model

Following [], 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 $D = z/2rvt$.

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
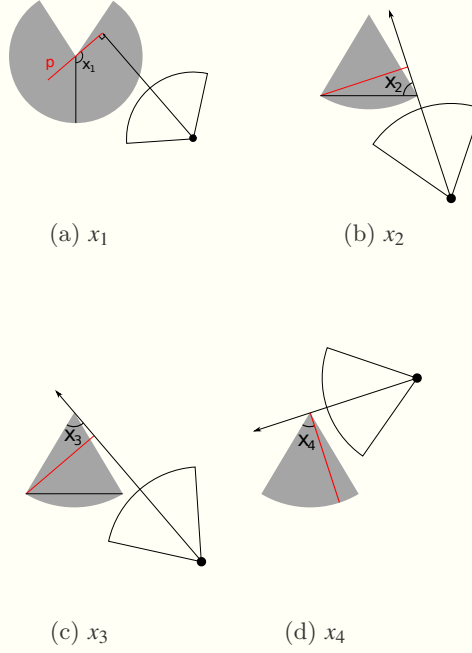
(a) x_1 (b) x_2

(c) x_3 (d) x_4

Figure A.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.

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 combinations 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) (Fig. ??). For example, the gas model becomes the simplest
gREM sub-model (upper right in Fig. ??) and the REM from [] 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.

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.

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

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

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 (Fig. A.1The location of the focal angles $x_{i\in[1,4]}$a).

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 (Fig. ??). 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 (Fig. A.2AgREM Appendix).

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

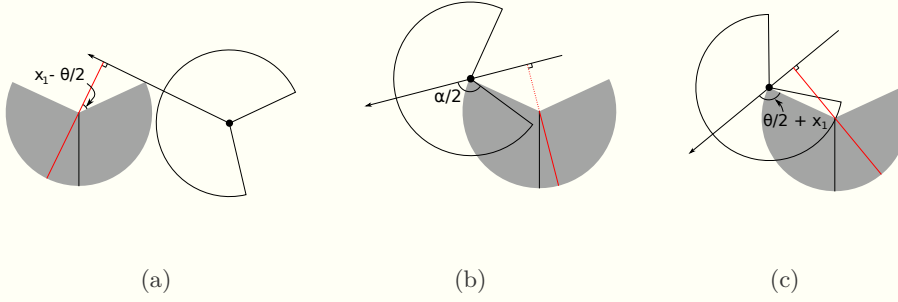(a)   (b)   (c)

Figure A.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
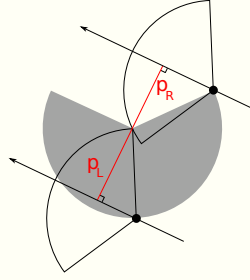shown and animals to the right of the detector can again be detected.

Figure A.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.

whole right side of the profile is invisible to the sensor (Fig. A.2BgREM
Appendix). 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 $r + r\cos(x_1 + \theta/2)$ (Fig. A.2CgREM
   Appendix). From $x_1 = \pi$, 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.

Model NE1

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

$$\bar{p}_{\text{NE1}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r \, dx_1 + \int_{\frac{\theta}{2}}^{\pi} r\cos\left(\frac{\theta}{2} - x_1\right) + r \, dx_1 \right.$$

$$\left. + \int_{\pi}^{2\pi - \frac{\theta}{2}} r\cos\left(\frac{\theta}{2} + x_1\right) + r \, dx_1 + \int_{2\pi - \frac{\theta}{2}}^{\frac{3\pi}{2}} 2r \, dx_1 \right) \tag{A.3}$$

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

Model NE2

Model NE2 is bounded by $\alpha \leq 3\pi - \theta$, $\alpha \geq 4\pi - 2\theta$ and $\alpha \geq \pi$ (Fig. ??). 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.

$$\bar{p}_{\text{NE2}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r \, dx_1 + \int_{\frac{\theta}{2}}^{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}} r\cos\left(\frac{\theta}{2} - x_1\right) + r \, dx_1 \right.$$

$$\left. + \int_{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}}^{2\pi - \frac{\theta}{2}} r\cos\left(\frac{\theta}{2} + x_1\right) + r \, dx_1 + \int_{2\pi - \frac{\theta}{2}}^{\frac{3\pi}{2}} 2r \, dx_1 \right) \tag{A.5}$$

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

Model NE3

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

$$
\bar{p}_{\text{NE3}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\theta}{2}} 2r \, dx_1 + \int_{\frac{\theta}{2}}^{\frac{\theta}{2}+\frac{\pi}{2}} r \cos\left(\frac{\theta}{2} - x_1 \right) + r \, dx_1 \right.
$$

$$
\left. + \int_{\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}} r \, dx_1 + \int_{\frac{5\pi}{2}-\frac{\theta}{2}-\frac{\alpha}{2}}^{2\pi-\frac{\theta}{2}} r \cos\left(\frac{\theta}{2} + x_1 \right) + r \, dx_1 + \int_{2\pi-\frac{\theta}{2}}^{\frac{3\pi}{2}} 2r \, dx_1 \right) \quad \text{(A.7)}
$$

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

Models SE2–4

Quadrant SE contains three submodels excluding SE1 (Fig. ??). 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 $r\sin(\alpha/2) + r\cos(x_1 - \theta/2)$ (Fig. A.3The second integral in SE).

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 $\alpha \leq 4\pi - 2\theta$.

Model SE2

SE2 is bounded by $\alpha \geq 4\pi - 2\theta$, $\alpha \leq \pi$ and $\theta \leq 2\pi$ (Fig. ??). 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 AgREM Appendix.

$$\bar{p}_{\text{SE2}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}} 2r \sin\left(\frac{\alpha}{2}\right) dx_1 + \int_{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}} r \sin\left(\frac{\alpha}{2}\right) + r \cos\left(\frac{\theta}{2} - x_1\right) dx_1 + \int_{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{3\pi}{2}} 2r \sin\left(\frac{\alpha}{2}\right) d. \right.$$

$$\tag{A.9}$$

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

Model SE3

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

$$\bar{p}_{\text{SE3}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}} 2r \sin\left(\frac{\alpha}{2}\right) dx_1 + \int_{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{\theta}{2} + \frac{\pi}{2}} r \sin\left(\frac{\alpha}{2}\right) + r \cos\left(\frac{\theta}{2} - x_1\right) dx_1 \right.$$

$$\left. + \int_{\frac{\theta}{2} + \frac{\pi}{2}}^{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}} r \sin\left(\frac{\alpha}{2}\right) dx_1 + \int_{\frac{5\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{3\pi}{2}} 2r \sin\left(\frac{\alpha}{2}\right) dx_1 \right) \tag{A.11}$$

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

Model SE4

Finally SE4 is bounded by $\alpha \le 4\pi - 2\theta$, $\alpha \le \pi$ and $\theta \le \pi$ (Fig. ??). 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.

$$\bar{p}_{\text{SE4}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}}^{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}} 2r \sin\left(\frac{\alpha}{2}\right) dx_1 + \int_{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{\theta}{2} + \frac{\pi}{2}} r \sin\left(\frac{\alpha}{2}\right) + r \cos\left(\frac{\theta}{2} - x_1\right) dx_1 + \int_{\frac{\theta}{2} + \frac{\pi}{2}}^{\frac{\alpha}{2} + \frac{\theta}{2} + \frac{\pi}{2}} r \sin\left(\frac{\alpha}{2}\right) dx_1 \right.$$

$$\tag{A.13}$$

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

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
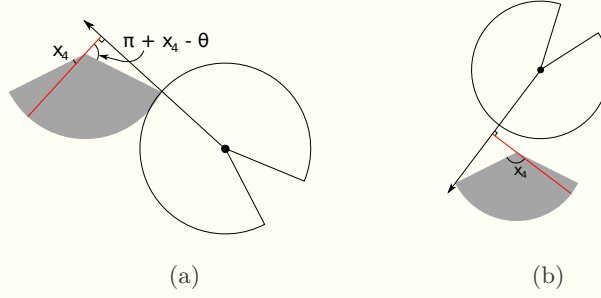
(a)   (b)

Figure A.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.

and γ_2 in [] while x_4 is new. They are described in Fig. A.1The location of the
focal angles $x_{i\in[1,4]}$b–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 (Fig. A.1BgREM Appendix). 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 $-r\cos(x_4 - \theta)$ (Fig. A.4AgREM Appendix).

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
   (Fig. A.4BgREM Appendix). Therefore the width of the profile is
   $r - r\cos(x_4)$.

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

$$\bar{p}_{\text{NW1}} = \frac{1}{\pi} \left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2) \, dx_2 + \int_{0}^{\theta - \frac{\pi}{2}} r - r \cos(-x_4 + \theta) \, dx_4 \right.$$

$$\left. + \int_{\theta - \frac{\pi}{2}}^{\frac{\pi}{2}} r \, dx_4 + \int_{\frac{\pi}{2}}^{\theta} r - r \cos(x_4) \, dx_4 + \int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2) \, dx_2 \right) \quad \text{(A.15)}$$

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

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 $\alpha \le 3\pi - 2\theta$. This inequality is found by noting that 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 $\alpha \le 2\pi$ 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 $\alpha/2 \le \pi - \theta/2$ (Fig. A.5Profile sizes when an animal approaches from behind in models NW2–4). This inequality (simplified as $\alpha \le 2\pi - \theta$) defines the boundary between NW3 and NW4.
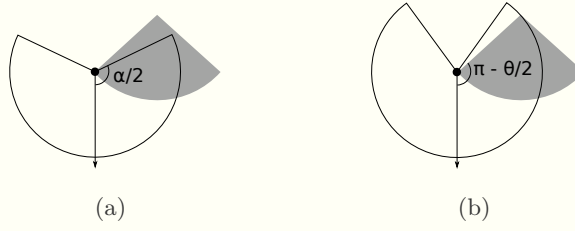
(a)   (b)

Figure A.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 $2r\sin(\theta/2)\sin(x_2)$.

Model NW2

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

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 $x_4 = \theta$.

$$\bar{p}_{\text{NW2}} = \frac{1}{\pi} \left(\int\limits_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right) \sin(x_2)\, dx_2 + \int\limits_{0}^{\theta - \frac{\pi}{2}} r - r\cos(-x_4 + \theta)\, dx_4 \right.$$

$$\left. + \int\limits_{\theta - \frac{\pi}{2}}^{\frac{3\pi}{2} - \frac{\alpha}{2}} r\, dx_4 + \int\limits_{\frac{3\pi}{2} - \frac{\alpha}{2}}^{\theta} r - r\cos(x_4)\, dx_4 + \int\limits_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right) \sin(x_2)\, dx_2 \right)$$

$$(A.17)$$

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

Model NW3

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

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 $2r\sin(x_2)\sin(\theta/2)$.

$$
\bar{p}_{\text{NW3}} = \frac{1}{\pi} \left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\, dx_2 + \int_0^{\theta - \frac{\pi}{2}} r - r\cos(-x_4 + \theta)\, dx_4 \right.
$$
$$
\left. + \int_{\theta - \frac{\pi}{2}}^{\theta} r\, dx_4 + \int_{\frac{\theta}{2}}^{\frac{3\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}} r\cos\left(\frac{\theta}{2} - x_2\right)\, dx_2 + \int_{\frac{3\pi}{2} - \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\, dx_2 \right)
$$

$$(A.19)$$

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

Model NW4

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

$$
\bar{p}_{\text{NW4}} = \frac{1}{\pi}\left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\, dx_2 + \int_0^{\theta - \frac{\pi}{2}} r - r\cos(-x_4 + \theta)\, dx_4 \right.
$$
$$
\left. + \int_{\theta - \frac{\pi}{2}}^{\theta} r\, dx_4 + \int_{\frac{\theta}{2}}^{\frac{\alpha}{2} + \frac{\theta}{2} - \frac{\pi}{2}} r\cos\left(\frac{\theta}{2} - x_2\right)\, dx_2 \right) \tag{A.21}
$$

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

Model REM

REM is the model from []. It has $\alpha = 2\pi$ and $\theta \leq \pi/2$ (Fig. ??). 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.

$$\bar{p}_{\text{REM}} = \frac{1}{\pi}\left(\int\limits_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 + \int\limits_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 \right.$$

$$\left. + \int\limits_{0}^{\theta} r\,\mathrm{d}x_4 + \int\limits_{\theta}^{\frac{\pi}{2}} r\sin(x_3)\,\mathrm{d}x_3 + \int\limits_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\,\mathrm{d}x_2 \right) \qquad \text{(A.23)}$$

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

Models NW5–7

In the models NW5–7, the sensor has $\theta \le \pi/2$ as in the REM. As $\alpha \ge \pi$ 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 $\alpha \ge \pi$ 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 $\alpha \le 2\pi + 2\theta$, 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 $2r\sin(x_2)\sin(\theta/2)$.

Model NW5

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

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.

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

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

Model NW6

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

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.

$$\bar{p}_{\text{NW6}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2} - \frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2)\, dx_2 + \int_{\theta}^{\frac{\pi}{2}} r \sin(x_3)\, dx_3 \right.$$

$$\left. + \int_{0}^{\theta} r\, dx_4 + \int_{\theta}^{\frac{\pi}{2}} r \sin(x_3)\, dx_3 + \int_{\frac{\pi}{2} - \frac{\theta}{2}}^{\frac{\alpha}{2} + \frac{\theta}{2} - \frac{\pi}{2}} r \cos\left(\frac{\theta}{2} - x_2\right) dx_2 \right) \qquad \text{(A.27)}$$

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

Model NW7

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

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.

$$\bar{p}_{\text{NW7}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2} - \frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2)\, dx_2 + \int_{\theta}^{\frac{\pi}{2}} r \sin(x_3)\, dx_3 \right.$$

$$\left. + \int_{0}^{\theta} r\, dx_4 + \int_{\pi - \frac{\alpha}{2}}^{\frac{\pi}{2}} r \sin(x_3)\, dx_3 \right) \qquad \text{(A.29)}$$

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

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 $2r \sin(\alpha/2)$ (Fig. A.6The first profile in SW models)
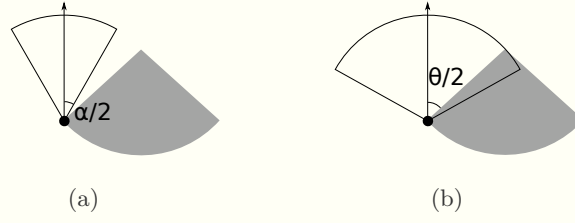
Figure A.6: The first profile in SW models is limited by either α or β depending on whether $\alpha < \beta$. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. (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 $2r\sin(\theta/2)\sin(x_2)$.

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 $r\sin(\alpha/2) - r\cos(x_4 - \theta)$.

Model SW1

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

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 $r\sin(\alpha/2)$.

$$\bar{p}_{\text{SW1}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin(x_2)\, dx_2 + \int_{\frac{\theta}{2}}^{\frac{\pi}{2} + \frac{\theta}{2} - \frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2} + x_2\right) dx_2 \right.$$
$$\left. + \int_{0}^{\theta - \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos(\theta - x_4)\, dx_4 + \int_{\theta - \frac{\pi}{2}}^{\frac{\alpha}{2} + \theta - \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_4 \right) \tag{A.31}$$

$$\bar{p}_{\text{SW1}} = \frac{r}{\pi} \left(\theta\sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1 \right) \tag{A.32}$$

Model SW2

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

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 $x_2 = \pi/2 - \alpha/2 + \theta/2$.

$$\bar{p}_{\text{SW2}} = \frac{1}{\pi} \left(\int_{\frac{\alpha}{2} - \frac{\theta}{2} + \frac{\pi}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) dx_2 + \int_{\frac{\theta}{2}}^{\frac{\alpha}{2} - \frac{\theta}{2} + \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2} + x_2\right) dx_2 \right.$$
$$\left. + \int_{0}^{\theta - \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos(\theta - x_4)\, dx_4 + \int_{\theta - \frac{\pi}{2}}^{\frac{\alpha}{2} + \theta - \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_4 \right) \tag{A.33}$$

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

Model SW3

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

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.

$$\bar{p}_{\text{SW3}} = \frac{1}{\pi} \left(\int_{\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \, dx_2 + \int_{0}^{-\frac{\pi}{2}+\theta-\frac{\alpha}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \, dx_4 \right.$$

$$\left. + \int_{-\frac{\pi}{2}+\theta-\frac{\alpha}{2}}^{\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos(\theta-x_4) \, dx_4 + \int_{\theta-\frac{\pi}{2}}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \, dx_4 \right) \quad \text{(A.35)}$$

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

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 Fig. A.7AgREM Appendix 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 (Fig. A.6The first profile in SW models), 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 (Fig. A.7BgREM Appendix). This gives a width of $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$. This profile does not occur in models with $\alpha \geq 2\theta$.

Model SW4

SW4 is bounded by $\alpha \leq \theta$, $\alpha \geq \pi - 2\theta$ and $\theta \leq \pi/2$ (Fig. ??). 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.
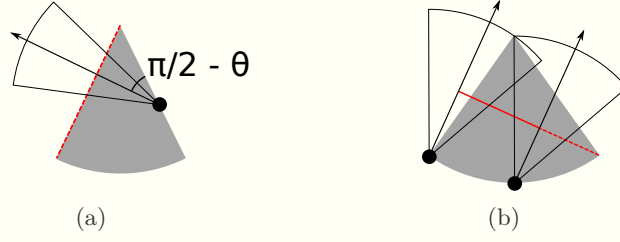
Figure A.7: Description of two profiles in SW models. The sector shaped detection region is shown in grey. Animals are filled black circles and the animal signal is an unfilled sector. The animals direction of movement is indicated with an arrow. The profile p is shown with a red line. Dashed red lines 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 $r\sin(\alpha/2) - r\cos(x_2 + \theta/2)$.

$$\bar{p}_{\text{SW4}} = \frac{1}{\pi} \left(\int_{\frac{\alpha}{2} - \frac{\theta}{2} + \frac{\pi}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) \, dx_2 + \int_{\frac{\pi}{2} - \frac{\theta}{2}}^{\frac{\alpha}{2} - \frac{\theta}{2} + \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2} + x_2\right) \, dx_2 \right.$$

$$\left. + \int_{\theta}^{\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \, dx_3 + \int_{0}^{\frac{\alpha}{2} + \theta - \frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) \, dx_4 \right) \tag{A.37}$$

$$\bar{p}_{\text{SW4}} = \frac{r}{\pi} \left(\theta \sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1 \right) \tag{A.38}$$

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 $\theta \leq \pi/2$ (Fig. ??). 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.

$$\bar{p}_{\text{SW5}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right) dx_2 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right) dx_2 \right.$$

$$\left. + \int_{\theta}^{\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_3 + \int_{0}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_4 \right) \tag{A.39}$$

$$\bar{p}_{\text{SW5}} = \frac{r}{\pi} \left(\theta \sin\left(\frac{\alpha}{2}\right) - \cos\left(\frac{\alpha}{2}\right) + 1 \right) \tag{A.40}$$

Model SW6

SW6 is bounded by $\alpha \geq \pi - 2\theta$, $\alpha \geq 2\theta$ and $\alpha \leq \pi$ (Fig. ??). 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.

$$\bar{p}_{\text{SW6}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\theta}{2}\right)\sin\left(x_2\right) dx_2 + \int_{\theta}^{\frac{\alpha}{2}} r\sin\left(x_3\right) dx_3 \right.$$

$$\left. + \int_{\frac{\alpha}{2}}^{\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_3 + \int_{0}^{\frac{\alpha}{2}+\theta-\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) dx_4 \right) \tag{A.41}$$

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

Model SW7

SW7 is bounded by $\alpha \leq \pi - 2\theta$, $\alpha \leq \theta$ and $\alpha < 0$ (Fig. ??). 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.

$$\bar{p}_{\text{SW7}} = \frac{1}{\pi} \left(\int_{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}}^{\frac{\pi}{2}} 2r\sin\left(\frac{\alpha}{2}\right) dx_2 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\alpha}{2}-\frac{\theta}{2}+\frac{\pi}{2}} r\sin\left(\frac{\alpha}{2}\right) - r\cos\left(\frac{\theta}{2}+x_2\right) dx_2 + \int_{\theta}^{\frac{\alpha}{2}+\theta} r\sin\left(\frac{\alpha}{2}\right) dx_3 \right)$$

$$\tag{A.43}$$

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

Model SW8

SW8 is bounded by $\alpha \leq \pi - 2\theta$, $\alpha \geq \theta$ and $\alpha \leq 2\theta$ (Fig. ??). 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.

$$\bar{p}_{\text{SW8}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2) \, dx_2 + \int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}+\frac{\theta}{2}-\frac{\alpha}{2}} r \sin\left(\frac{\alpha}{2}\right) - r \cos\left(\frac{\theta}{2}+x_2\right) dx_2 + \int_{\theta}^{\frac{\alpha}{2}+\theta} r \sin\left(\frac{\alpha}{2}\right) dx_3 \right)$$

$$\tag{A.45}$$

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

Model SW9

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

$$\bar{p}_{\text{SW9}} = \frac{1}{\pi} \left(\int_{\frac{\pi}{2}-\frac{\theta}{2}}^{\frac{\pi}{2}} 2r \sin\left(\frac{\theta}{2}\right) \sin(x_2) \, dx_2 + \int_{\theta}^{\frac{\alpha}{2}} r \sin(x_3) \, dx_3 + \int_{\frac{\alpha}{2}}^{\frac{\alpha}{2}+\theta} r \sin\left(\frac{\alpha}{2}\right) dx_3 \right)$$

$$\tag{A.47}$$

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

Supplementary Information: Simulation model results of the gREM precision
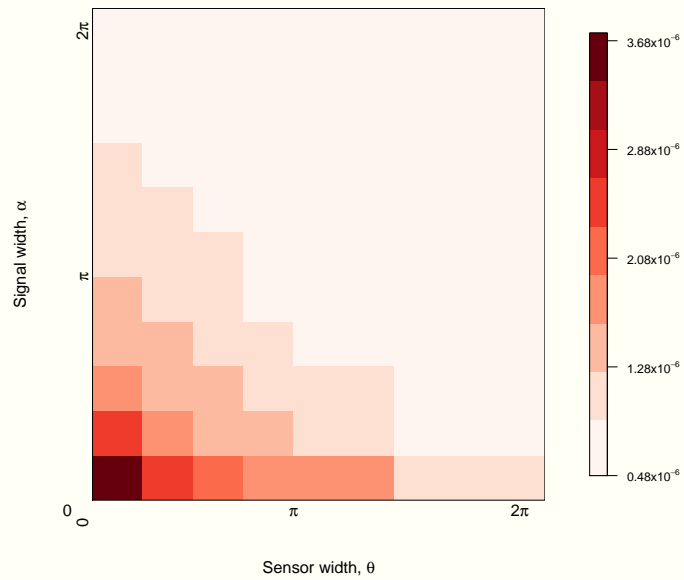

Figure A.1: Simulation model results of the gREM precision given a range of sensor and signal widths, shown by the standard deviation of the error between the estimated and true densities. Standard deviations are shown from deep red to pink, representing high to low values between 0.483×10^{-6} to 3.74×10^{-6}.

Supplementary Information: Impact of parameter error

(a) labelf:signal

(b) labelf:sensor
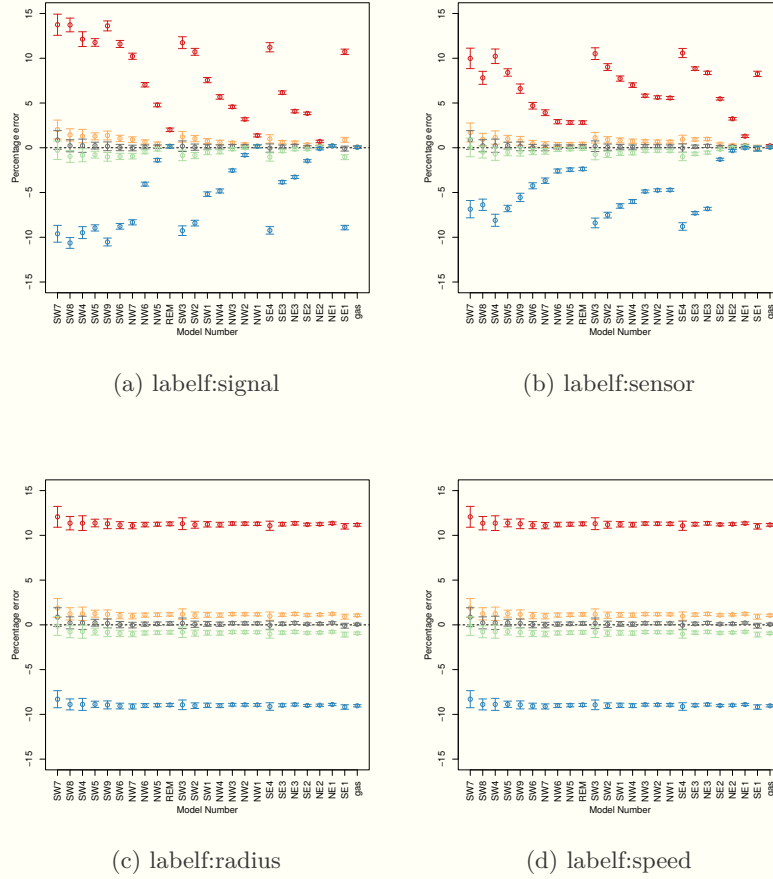
(c) labelf:radius

(d) labelf:speed

Figure A.1: 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.

Appendix B

Colophon

This document was set using LaTeX, XeLaTeX and BibLaTeX. The TeX Gyre Pagella typeface is used throughout. Plots were created with a combination of [,] and Inkscape.