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

July 21, 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

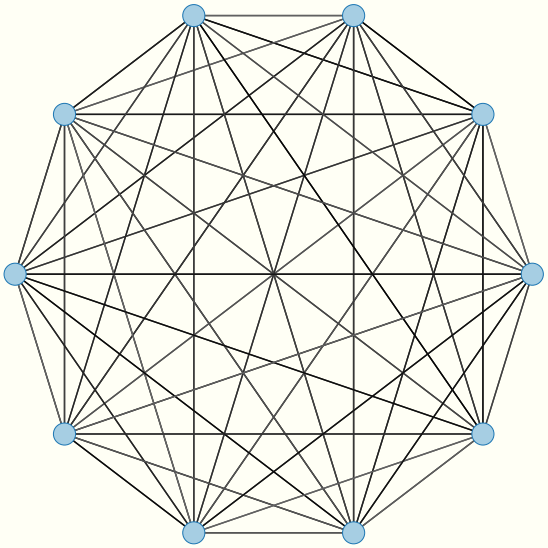
Introductory Material

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.

Chapter 2

Does network structure of bat populations promote viral diversity?



Abstract

One or two sentences providing a basic introduction to the field

An increasingly large fraction of emerging diseases come from animals[35,74] 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.

Two to three sentences of more detailed background

Host species traits such as population density, longevity, body size and population structure have been shown to correlate with pathogen diversity. However, our mechanistic understanding of how population structure (i.e. non random contacts across the population creating barriers to disease spread) affects pathogen diversity is poor. Greater mechanistic understanding is needed to clarify the exact causal role population structure has in controlling pathogen diversity. Mechanistic models are also likely to be more robust to transferring understanding between taxa and predicting changes.

Typically it is assumed that well-connected populations promote disease spread (high R_0) and therefore promote pathogen diversity. However, if competition is strong endemic pathogens will dominate and prevent new diseases from invading and spreading.[2,12,52,62] In a structured population, stochastic effects could create areas of low prevalence of the endemic disease, allowing new diseases to invade.

We consider bats as a case study as they have been implicated in a number of recent, high profile diseases such as Ebola, SARS, Hendra and Nipah. Bats have varied social structures and so the structure of populations could be one way to prioritise zoonotic disease surveillance in this group.

One sentence clearly stating the general problem (the gap)

It is unknown whether population structure allows escape from competition and therefore high diversity.

We hypothesise that low dispersal rates and a low number of connections in a metapopulation network will allow invading pathogens to establish more readily.

One sentence summarising the main result

I find that neither population connectedness nor dispersal rate affect the probability that a new pathogen will invade into a population.

Two or three sentences explaining what the main result reveals in direct comparison to what was thought to be the case previously

The common assumption that factors causing high R_0 allow new pathogens to invade and therefore increase pathogen diversity is not supported by our study. Instead we find that changes in population structure that would affect R_0 do not affect the probability of invasion of a new pathogen.

One or two sentences to put the results into a more general context.

This result means that large scale population structure does not seem to 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.

Two or three sentences to provide a broader perspective,

Introduction

General Intro

Why is pathogen diversity important?

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[35], 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.

We know some factors that correlate with pathogen diversity

A number of host traits have been shown to correlate with pathogen richness including body size[38,7], population density[56,7] and range size[11,38].

But we do not understand the mechanistic processes

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[38] for a meta-analysis). Furthermore, the correlation between many traits (e.g.[57]) 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[15], reduction of the rate of parasite extinction and an increased probability of pathogen invasion from other 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.

Specific Intro

Population structure could be important

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. However, the assumption that population structures that yield high R_0 will also give high pathogen diversity[56] is unfounded.

We cannot assume high R_0 gives high diversity

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[12,52,2,3,77]. 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 is seems that cross immunity can keep a pathogen at a low prevalence in the population (e.g. monkey pox[65] and monkey malaria[21]). 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.

Network structure has been studied

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 diversity[53,3] or competitive exclusion have both been predicted[2,12,52,62,5] When competitive exclusion occurs, population structure has sometimes been shown to allow coexistence[62,5,55].

Competing epidemics, or two pathogens spreading at the same time in a population, is a well studied area[60,59,40]. 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.

Empirical evidence that structure might affect diversity

Few studies focus on bats, despite their role in recent zoonoses. Maganga *et al.* found that distribution fragmentation predicts viral richness[48], but[28] finds the opposite relationship. While the data set in[28] is larger, the analysis in[48] is much more focused on fragmentation.

Genetic correlates of population structure have also been used. Turmelle *et al.*[76], in a small analysis, find that high F_{st} (i.e. a structured population) correlates with high richness.

Types of population structure

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. We 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 subpopulations, 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.

Why bats

Bats (Order Chiroptera) have, over the last decade, become a focus for disease research[18,32]. Recently they have been implicated in a number of high profile diseases such as Ebola, SARS, Hendra and Nipah[18,45].

A number of traits have been suggested as predisposing bats towards being reservoirs of zoonotic diseases: high sympatry[46], flight[80] and longevity[80]

Bats have an unusual variety of social structures. Group living ranges from colonies 10–1 million[37]. Many bats also have interesting seasonal behaviour such as migration[64,26]

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 diversity.

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.

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. Note: I think I probably should run sims for two species together to also study the mechanism of pathogen extinction.

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.

Term	Definition	Synonyms
Metapopulation	A group of colonies with rare movement of animals between them. Closed to outside migration.	Network
Subpopulation	A group of animals. Social interactions within a colony is likely high.	Node, colony
Dispersal	Movement from one colony 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

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 2.1). 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 2.2) 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.

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

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

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

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

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

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

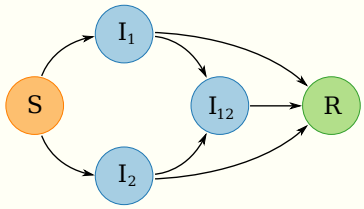


Figure 2.1: The SIR model used.

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

$$P(S_{nt'} = S_{nt} - 1) = dS_{nt} \tag{2.4}$$
$$P\left(I_{qnt'} = I_{qnt} - 1\right) = dI_{qnt} \tag{2.5}$$
$$P(R_{nt'} = R_{nt} - 1) = dR_{nt}. \tag{2.6}$$

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

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

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

$$P\left(I_{12,nt'} = I_{12,nt} + 1,\, I_{pnt'} = I_{pnt} - 1\right) = \alpha\beta I_{nt} I^{+}_{pnt}. \tag{2.8}$$

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

$$P(S_{nt'} = S_{nt} + 1,\, S_{mt'} = S_{mt} - 1) = \frac{\lambda S_{mt}}{k_m - 1} \tag{2.9}$$
$$P\left(I_{qnt'} = I_{qnt} + 1,\, I_{qmt'} = I_{qmt} - 1\right) = \frac{\lambda I_{qmt}}{k_m} \tag{2.10}$$
$$P(R_{nt'} = S_{nt} + 1,\, R_{mt'} = R_{mt} - 1) = \frac{\lambda R_{mt}}{k_m}. \tag{2.11}$$

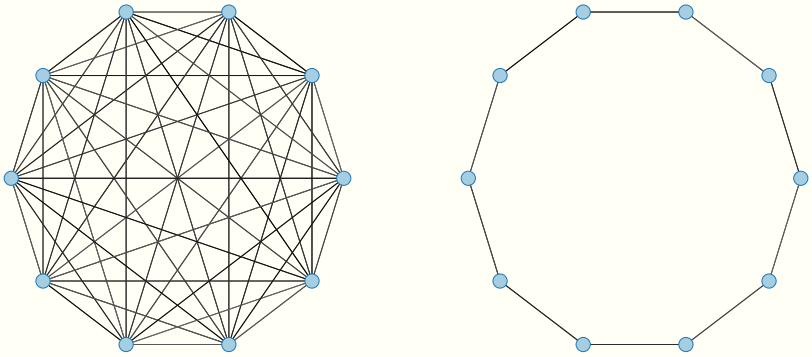
Finally, recovery from any infectious class occurs at a rate γ

$$P\left(I_{qnt'} = I_{qnt} - 1,\, R_{nt'} = R_{nt} + 1\right) = \gamma I_{qnt}. \tag{2.12}$$

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 40,000 events, 10 individuals infected with disease 2 are added to one colony. After another 10,000 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



(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. Dispersal is held constant between the two topologies.

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 2.2). 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.

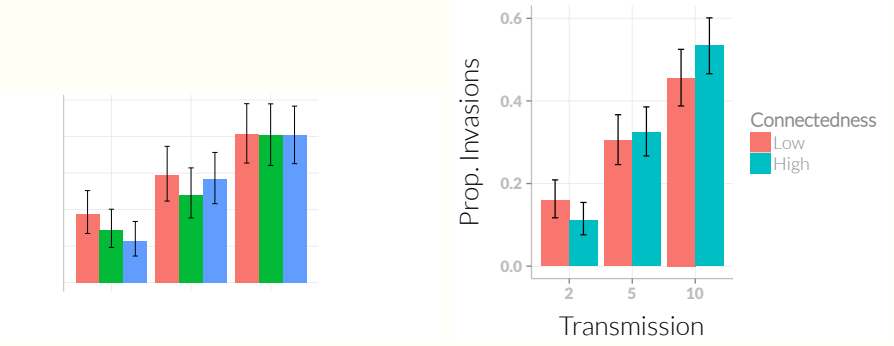
The population size of each subpopulation is set to 3000. This is appropriate for many bat species[37], especially the large, frugivorous *Pteropodidae* that have been particularly associated with recent zoonotic diseases.

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.



(a) Dispersal **(b)** Network topology

Figure 2.3: The probability of successful invasion. For three different transmission rates, the probability of invasion success does not change between different a) dispersal rates or b) network structures. Error bars are 95% confidence intervals. Other parameters are kept constant at: $N = 10$, $\bar{n} = 3000$, $b = d = 0.05$, $\gamma = 0.1$, $\alpha = 0.1$. When dispersal is varied, the population structure is fully connected. When population structure is varied, $\lambda = 0.01$.

Results

Invasion

The probability of the new pathogen invading ranged from 0.12 to 0.55.

Dispersal

The proportion of invasions was not different across dispersal rates across 1500 simulation run.. This was true at all transmission levels (χ^2 test. $\beta = 2$: $\chi^2 = 4.19$, df $= 2$, $p = 0.12$. $\beta = 5$: $\chi^2 = 1.27$, df $= 2$, $p = 0.53$. $\beta = 10$: $\chi^2 = 6.61 \times 10^{-3}$, df $= 2$, $p = 1.00$).

Network structure

I ran 2000 simulations over three transmission values ($\beta = 2, 5, 10$). The proportion of invasions was not different between highly connected and largely unconnected metapopulations (See Figure **??**). This was true at all transmission levels (χ^2 test. $\beta = 2$: $\chi^2 = 1.81$, df $= 1$, $p = 0.18$. $\beta = 5$: $\chi^2 = 0.62$, df $= 1$, $p = 0.43$. $\beta = 10$: $\chi^2 = 2.65$, df $= 1$, $p = 0.10$).

In a small proportion of simulations, both pathogens went extinct. In these cases we still counted the invading pathogen as failing to invade. However, the results are the same as above if these simulations are removed from the analysis (χ^2 test. $\beta = 2$: $\chi^2 = 2.29$, df $= 1$, $p = 0.13$. $\beta = 5$: $\chi^2 = 0.15$, df $= 1$, $p = 0.70$. $\beta = 10$: $\chi^2 = 2.36$, df $= 1$, $p = 0.12$).

Transmission

Inline with theory, increasing the transmission rate increased the probability of invasion (Figure **??**–**??**). This is true for all three dispersal values (χ^2 test. $\lambda = 0.001$: $\chi^2 = 19.35$, df $= 2$, $p = 10^{-4}$. $\lambda = 0.01$: $\chi^2 = 29.05$, df $= 2$, $p = 10^{-6}$. $\lambda = 0.1$: $\chi^2 = 38.9$, df $= 2$, $p = 10^{-8}$) and both network structures (χ^2 test. Fully connected: $\chi^2 = 66.4$, df $= 2$, $p = 10^{-14}$. Minimally connected: $\chi^2 = 66.4$, df $= 2$, $p = 10^{-14}$).

Extinction

I ran 1000 simulations varying dispersal rate and 1000 varying network topology. However, in these simulations no pathogens went extinct. Therefore it seems that

once a pathogen is established in the these populations, it is very rare for them to stochastically go extinct. Dispersal nor network topology altered the proportion of extinctions, but clearly this comparison does tell us much.

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[19]. However, the model in[19] 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[59] 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[59] 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[54,54] or infectious period[poletto2013hostinfec] . 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.	0.1
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		
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 ecological and epidemiological seasonality promote viral diversity?

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.

Chapter 4

Does social structure affect viral diversity in wild bat populations?

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.

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[81] and extinction risk[61]. Monitoring animal population changes

in response to anthropogenic pressure is becoming increasingly important as humans rapidly modify habitats and change climates[25]. Sensor technology, such as camera traps[39,68] and acoustic detectors[1,78] 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[68,42,79]. However, converting sampled count data into estimates of density is problematic as detectability of animals needs to be accounted for[6].

Existing methods for estimating animal density often require additional information that is often unavailable. For example, capture-mark-recapture methods[39,10] require recognition of individuals, and distance methods[29] require estimates of how far away individuals are from the sensor[8,51]. When individuals cannot be told apart, an extension of occupancy modelling can be used to estimate absolute abundance[70]. 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[70] which may be a poor assumption for nonrandomly distributed species. Furthermore repeat, independent surveys must be performed and the definition of a site can be difficult, especially for wide-ranging species[47].

The REM method has been successfully applied to estimate animal densities from camera trap surveys[83]. 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[69].

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,[14]). 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[42], 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[9], 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[42]. Acoustic monitoring is being developed to study many aspects of ecology, including the interactions of animals and their environments[9,66], the presence and relative abundances of species[49], biodiversity of an area[24], and monitoring population trends[79].

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[50]. 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[44]. In these situations distance-sampling methods can be applied[17]. 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[17]. 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[71,79]. 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

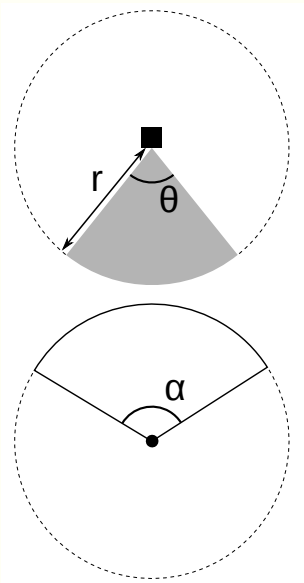


Figure 5.1: Representation of sensor detection width and animal signal width. The filled square and circle represent a sensor and an animal, respectively; θ, sensor detection width (radians); r, sensor detection distance; dark grey shaded area, sensor detection zone; α, animal signal width (radians).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

trap model of[69], 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[69] adapts the gas model to count data collected from camera trap surveys. The REM is derived assuming a stationary sensor with a detection width less than $\pi/2$ radians. However, in order to apply this approach more generally, and in particular to stationary acoustic detectors, we need both to relax the constraint on sensor detection width, and allow for animals with directional signals. Consequently, we derive the gREM for any detection width, θ, between 0 and 2π with a detection distance r giving a circular sector within which animals can be captured (the detection zone) (Figure 5.1). Additionally, we model the animal as having an associated signal width α between 0 and 2π (Figure 5.1, see Appendix S1 for a list of symbols). We start deriving the gREM with the simplest situation, the gas model where $\theta = 2\pi$ and $\alpha = 2\pi$.

Gas Model

Following[82], 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

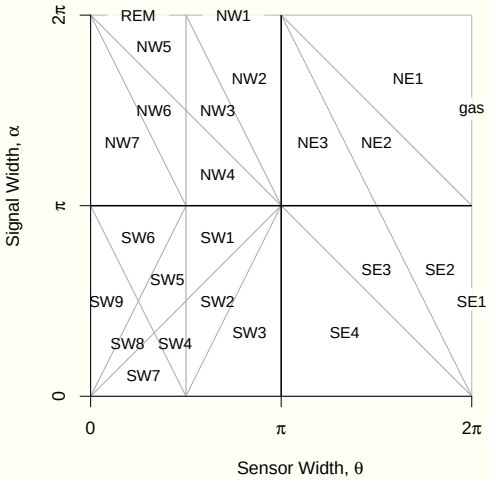


Figure 5.2: Locations where derivation of the average profile $\bar{p}$ is the same for different combinations of sensor detection and animal signal widths. Symbols within each polygon refer to each gREM submodel named after their compass point, except for Gas and REM which highlight the position of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

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[33].

gREM derivations for different detection and signal widths

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

We have identified the parameter space for the combinations of θ and α for which the derivation of the equations are the same (defined as sub-models in the gREM) (Figure 5.2). For example, the gas model becomes the simplest gREM sub-model (upper right in Figure 5.2) and the REM from[69] is another gREM sub-model

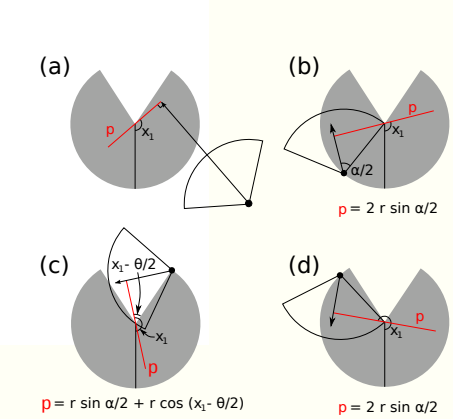


Figure 5.3: An overview of the derivation of the average profile $\bar{p}$ for the gREM submodel SE2, where (a) shows the location of the profile p (the line an animal must pass through in order to be captured) in red and the focal angle, x_1, for an animal (filled circle), its signal (unfilled sector),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

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[13,20]. Sensor width, θ, and detection distance, r would also need to be measured or obtained from manufacturer specifications[31,4].

Example derivation of SE2

In order to calculate $\bar{p}$, we have to integrate over the focal angle, x_1 (Figure 5.3a). This is the angle taken from the centre line of the sensor. Other focal angles are possible (x_2, x_3, x_4) and are used in other gREM 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 5.3b). 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 5.3c), the size of the profile is $r\sin(\alpha/2)+r\cos(x_1-\theta/2)$ and the size of θ and α both limit the width of the profile (Figure 5.3c). Finally, at $x_1 = 5\pi/2-\theta/2-\alpha/2$ until $x_1 = 3\pi/2$, the width of the profile is again $2r\sin\alpha/2$ (Figure 5.3d) and the size of α again limits the width of the profile.

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

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

$$= \frac{r}{\pi}\left(\theta\sin\frac{\alpha}{2} - \cos\frac{\alpha}{2} + \cos\left(\frac{\alpha}{2}+\theta\right)\right) \tag{5.2}$$

We then use this expression to calculate density

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

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

gREM submodel specifications were done by hand, and the integration was done using SymPy[73] 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[63] 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 km by 7.5 km square with periodic boundaries. A stationary sensor of radius r, 1.0×10^1 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 7.0×10^1 animals km^{-2}, calculated from the equation in[23] as the expected density of mammals weighing 1 g. This density therefore represents a reasonable estimate of density of individuals, given that the smallest mammal is around 2 g[37]. 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, $v = 4.0 \times 10^1$ km day^{-1}, was chosen based on the largest day range of terrestrial animals[20], and represents the upper limit of realistic speeds. At the end of each step, individuals were allowed to either remain stationary for a time step (with a given probability, S), or change direction where the change in direction has a uniform distribution in the interval $[-A, A]$. This resulted in seven different movement models where: (1) simple movement, where S and $A = 0$; (2) stop-start movement, where (i) $S = 0.25$, $A = 0$, (ii) $S = 0.5$, $A = 0$, (iii) $S = 0.75$, $A = 0$; (3) correlated random walk movement, where (i) $S = 0, A = \pi/3$, (ii) $S = 0, A = 2\pi/3$, iii) $S = 0, A = \pi$. Individuals were counted as they moved into the detection zone of the sensor per simulation.

We calculated the estimated animal density from the gREM by summing the number of captures per simulation and inputting these values into the correct gREM submodel. The accuracy of the gREM was determined by comparing the true simulation density with the estimated density. Precision of the gREM was determined by the standard deviation of estimated densities. We used this method to compare the accuracy and precision of all the gREM submodels. As these submodels are derived for different combinations of α and θ, the accuracy and precision of the submodels was used to determine the impact of different values of α and θ.

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

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 5.4 shows the expression for $\bar{p}$ in each case. The general equation for density, 5.3, is used with the correct value of $\bar{p}$ substituted. Although more thorough checks are performed in Appendix S3, it can be seen that all adjacent expressions in Figure 5.4 are equal when expressions for the boundaries between them are substituted in.

Simulation model

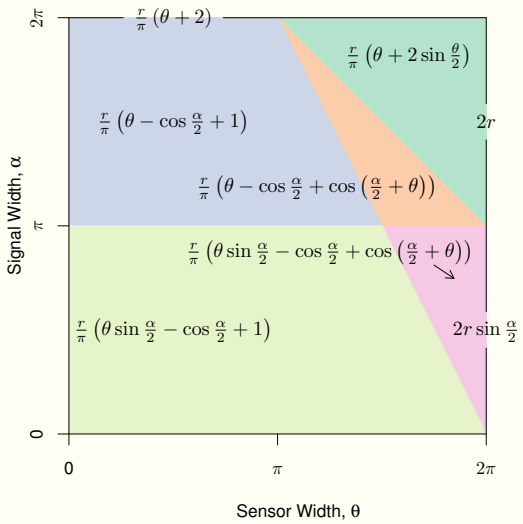


Figure 5.4: Expressions for the average profile width, $\bar{p}$, given a range of sensor and signal widths. Despite independent derivation within each block, many models result in the same expression. These are collected together and presented as one block of colour. Expressions on the edge of the plot 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 5.5). However, the precision of the submodels do vary, where the gas model is the most precise and the SW7 sub model the least precise, having the smallest and the largest interquartile range, respectively (Figure 5.5). The standard deviation of the error between the estimated and true densities is strongly related to both the sensor and signal widths (Appendix S5), such that larger widths have lower standard deviations (greater precision) due to the increased capture rate of these models.

Number of captures

Within the four gREM submodels tested (NW1, SW1, SE3, NE1), the accuracy was not strongly affected by the number of captures. The median difference between the estimated and true values was less than 15% across all capture rates (Figure 5.6). However, the precision was dependent on the number of captures across all four of the gREM submodels, where precision increases as number of captures increases, as would be expected for any statistical estimate (Figure 5.6). For all gREM submodels, the the coefficient of variation falls to 10% at 100 captures.

Movement models

Within the four gREM submodels tested (NW1, SW1, SE3, NE1), neither the accuracy or precision was affected by the average amount of time spent stationary. The median difference between the estimated and true values was less than 2% for each category of stationary time (0, 0.25, 0.5 and 0.75) (Figure 5.7a). 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 5.7b).

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

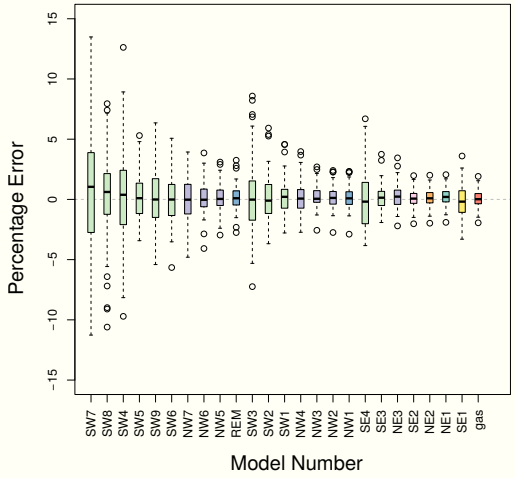


Figure 5.5: Simulation model results of the accuracy and precision for gREM submodels. The percentage error between estimated and true density for each gREM sub model is shown within each box plot, where the black line represents the median percentage error across all simulations, boxes represent the middle 50% of the data, whiskers represent variability outside the upper and lower quartiles with outliers plotted as individual points. Box colours correspond to the expressions for average profile width $\bar{p}$ given in Figure 4.

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

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[69] 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[33]. 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[43] or surveying echolocating bats from a moving car[36].

Interesting but unstudied problems impacting the gREM are firstly, edge effects caused by sensor trigger delays (the delay between sensing an animal and attempting

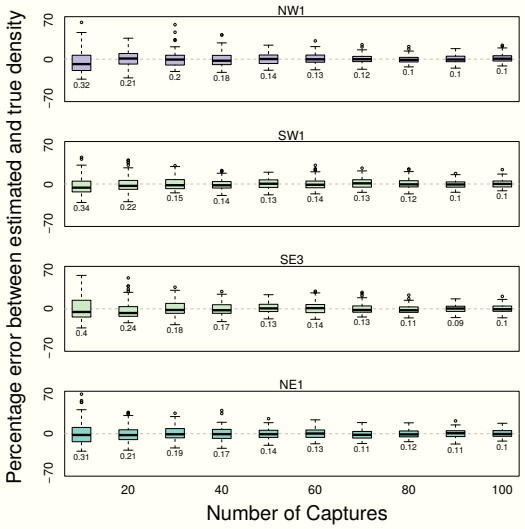


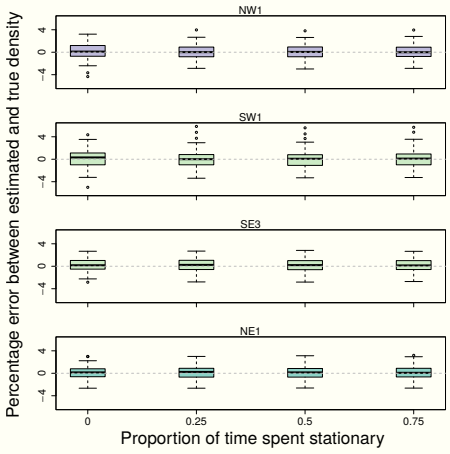
Figure 5.6: Simulation model results of the accuracy and precision of four gREM submodels (NW1, SW1, SE3 and NE1) given different numbers of captures. The percentage error between estimated and true density within each gREM sub model for capture rate is shown within each box plot, where the black line 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. 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 4.

to record the encounter)[67], and secondly, sensors which repeatedly turn on an off during sampling[36]. 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

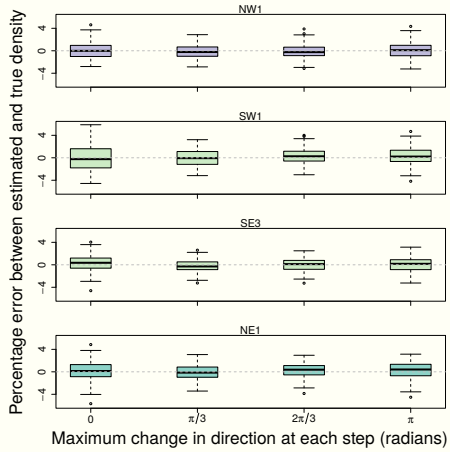
Based on our simulations, we believe that the gREM has the potential to produce accurate estimates for many different species, using either camera traps or acoustic detectors. However, the precision of the gREM differed between submodels. For example, when the sensor and signal width were small, the 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



[t]60mm

Figure 5.7



[t]60mm

Figure 5.8

Figure 5.9: 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 black 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. 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.

of variation, considered to be a very good level of precision and better than many previous studies[75,58,27]. 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[31], while the sensor angle is often easily measured[4] 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[13], 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[69,83], 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[69]. 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[72]. 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[33], 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[20]. 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[70,8,51]. 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[78], songbirds[16], whales[51] and forest primates[30]. 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[22]. 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[34].

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[41]. The gREM is also suitable for species that are sensitive to human contact or are difficult or dangerous to catch[75]. 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.

Acknowledgments

We thank Hilde Wilkinson-Herbot, Chris Carbone, Francois Balloux, Andrew Cunningham, Steve Hailes, Richard Glennie and an anonymous referee for comments on previous versions of the manuscript. This study was funded through CoMPLEX PhD studentships at University College London supported by BBSRC and EPSRC (EAM and TCDL) and The Darwin Initiative (Awards 15003, 161333, EIDPR075), NERC (NE/H525003/1), and The Leverhulme Trust (Philip Leverhulme Prize) for KEJ.

Data Accessibility

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

Chapter 6

General 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[82], 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 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. 5.2). For example, the gas model becomes the simplest gREM sub-model (upper right in Fig. 5.2) and the REM from[69] 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.