APPROACHES TO MODELLING RACCOON RABIES

A Thesis Submitted to the Committee on Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the
Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Watershed Ecosystem Graduate Program

June 2008

Abstract

Approaches to Modelling Raccoon Rabies

Erin Elizabeth Rees

A modelling approach was used to increase understanding of the raccoon-rabies disease-host system. This was achieved through the development and application of the Ontario Rabies Model (ORM). The ORM is a spatially explicit individual-based model (IBM) that simulates raccoon demographics, rabies disease transmission and various rabies control strategies. An important first step before using this new tool for genetic analyses was to validate model processes by assessing parameter input values and their impact on simulated outcomes, and model structure and system processes. This resulted in the creation of a Raccoon Ecology Database (REDB), development of a novel approach to density estimation from capture-mark-recapture data and the incorporation of information theoretic methods into model sensitivity analysis (SA). The REDB confirmed ORM default parameter values fell within known variation, provided sources for citing their values and data for meta-analysis. The REDB also enables parameterisation of the ORM in geographic regions beyond southern Ontario and can be used as an example data model for creating other ecological databases. The method for estimating density from capture-mark-recapture data is applicable to systematic or non-systematic trapping arrays. Raccoon densities in the St. Lawrence region (44°N 75°W) were found to range from 5 to 6 raccoons / km^2 for forest and agricultural habitat, respectively. The SA further ensured that the ORM functions as intended and that the

major factors implicated to affect disease-host systems (e.g. density, transmission rate) are also critical factors in the simulated raccoon-rabies system.

Once validated, the ORM was extended to simulate and track maternal and bi-parentally inherited neutral genetic markers. Additional model validation became possible by comparing simulated and empirically derived genetic population structures. The revised model was used to quantify the effect of the Niagara River as a 50 % barrier on raccoon movement in the Niagara Region (43°N 79°W) to infer the effect of this landscape barrier on the spread of raccoon-rabies. This work provided further validation of the ORM simulation tool by comparing simulated and empirically derived genetic population structures. Model development and application has increased understanding of the raccoon-rabies system and demonstrated the value of a modelling approach for ecological explorations.

Keywords: individual based model (IBM), raccoon (*Procyon lotor*), raccoon rabies, infectious-disease modelling, ecological database, density estimation, sensitivity analysis, neutral genetic markers, landscape genetics, landscape barriers

Acknowledgements

Thank you to my advisors, Bruce and Rick, and committee members, Roly and Dennis. Your enthusiasm, support and advice were critical to my enjoyment and success and continued development as a scientist.

Thank you to the other members of the raccoon rabies project, Cathy, Brad, Chris and Gaby. Our weekly meetings were interesting, and mostly entertaining (ha ha!), and I definitely appreciated your suggestions and opinions.

I am a data parasite. To that, I am indebted to all the data collectors, data entry and processing personnel: Ontario Ministry of Natural Resources (OMNR) trappers, OMNR staff (Kirk Sobey, Tore Buchannan), Queen's GIS Lab, Trent summer students (Kate Pammett) and contract workers (Shannon Stephens, Denis Carr).

I could not be a modeller without a model. Thank you to the Ontario Rabies Model developers; and in particular, modelling discussions and support from David Ball and Roly.

Chapters 2, 3, 4, and 5 form the basis of papers intended for publication. I am very appreciative of the contributions from all of my co-authors: Chapter 2 (Bruce A. Pond, Julia R. Phillips, Dennis L. Murray); Chapter 3 (Bruce A. Pond, Rick C. Rosatte, Kirk G. Sobey); Chapter 4 (Bruce A. Pond, Rowland Tinline & David Ball); Chapter 5 (Bruce A. Pond, Catherine I. Cullingham, Rowland Tinline, David Ball, Christopher J. Kyle, Bradley N. White).

To my family and friends – thanks for the encouragement, reality checks and fun.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spatial epidemiology is the analysis of factors affecting the spatial-temporal distribution of disease incidence. This field of research has received increased attention as human globalisation and climate change make our communities more susceptible to exotic diseases, as was recently demonstrated with SARS (Severe Acute Respiratory Syndrome), and the current threat of avian influenza.

Raccoon rabies

Rabies is also a significant public health concern. The classic rabies virus still causes 1000's of human deaths annually in parts of Africa and India that lack effective public health and pet vaccination programs (Sterner and Smith 2006). More specifically, the raccoon rabies strain is estimated to cost the US economy over $USD 400 million per year due to loss of livestock, quarantining suspect animals, human post-exposure-prophylaxis, rabies control programs (Sterner and Smith 2006).

Raccoon rabies is a variant rabies virus specifically adapted to infect raccoons (*Procyon lotor*) (Winkler and Jenkins 1991). It was first detected in Florida in the 1940's, and a second major epizootic emerged in the late 1970's along the West Virginia/Virginia border (Winkler and Jenkins 1991), spreading northwards at a rate of about 30 - 47 km/year (Childs et al. 2000), (Rupprecht and Smith 1994). It was first detected in Canada near Brockville, Ontario, in 1999 (Wandeler and Salsberg 1999) and the disease continues to threaten becoming established in south-eastern Canada from infected animals crossing the US-Canada border. There is an epizootic in Quebec

(currently, as of August 2007), but the disease is under control in Ontario and New Brunswick (Rick Rosatte, pers. comm).

Infectious disease control

Raccoon rabies threatens the health of humans, domestic animals, and many other wild animals, including skunks and foxes. Public health and animal welfare agencies in Canada and the US exert tremendous effort to prevent, control and ultimately eliminate this disease (Rosatte et al. 2001). A variety of strategies are used such as public education, surveillance, depopulation and vaccination (Rosatte et al. 1997). The goal of disease eradication is to reduce the density of susceptible individuals below a threshold for which the average individual has less than one contact resulting in the successful transmission of the disease (Anderson and May 1991). This can be achieved through population reduction (e.g. culling, fertility control) or by increasing the proportion of immune animals (e.g. vaccination) (Ferguson et al. 2003).

The challenge for designing infectious disease programs is determining which type(s) of control to use and their timing, frequency, duration and location of implementation, while staying within available resources and public health, animal and environmental regulations. For example, the most effective time for raccoon rabies vaccination in Ontario is from early summer, when the young-of-year are old enough to respond to vaccination (at least 3 months) (Rosatte et al. 1990), to before periods of higher dispersal in the fall so the disease is not spread further afield (Rosatte et al. 2001). Appropriate control strategies require knowledge about the disease-host biology and ecology and an understanding of factors in the environment that affect the spatial-temporal spread. This information can be gained from the analysis of field data and

through simulation modelling. I use both of these approaches to accomplish the goal of my thesis: to further develop an infectious disease simulation model, the Ontario Rabies Model (ORM; Appendix), to utilise raccoon genetics as another perspective for understanding the spread of raccoon rabies.

Infectious disease modelling

The development and application of infectious disease models have contributed greatly to current understanding of disease-host systems (Anderson and May 1991). One example is by determining conditions under which a disease will become epidemic and the proportion of the population needing to be vaccinated to eradicate the disease (Anderson and May 1979) (Anderson and May 1982) (Coyne et al. 1989) (Ferguson et al. 1997) (McCallum et al. 2001). Advances in ecological and epidemiological theory and computing power have enabled the development of different types of models. For example, "classical" state variable models (SVMs) are well-known to ecologists (e.g. Lotka-Volterra predation model {Lotka 1925 290 /id}, {Volterra 1926 291 /id})) and epidemiologists (e.g. susceptible-infected-recovery (SIR) model {Kermack & McKendrick 1927 56 /id})). More recently, individual-based models (IBMs) have emerged as a promising modelling strategy (Sterner and Smith 2006).

IBMs are built from the "bottom-up", such that individuals are explicitly represented, and it is the sum of the individual behaviours that characterise the population(s) (Grimm 1999). Model processes are stochastically or deterministically determined from a set of rules. The simulations may progress continuously or in discrete time steps {Berec 2002 289 /id}. Model parameters tend to be more mechanistic than phenomenological because they often explicitly represent the processes defining the

system. The ORM is an individual-based spatially explicit model that simulates raccoon demographics, rabies disease transmission and rabies control strategies, and was recently created by the Ontario Ministry of Natural Resources (OMNR) and the Queen's University GIS Lab.

Why use an IBM?

The raccoon rabies system has been explored using SVMs {Coyne, Smith, et al. 1989 54 /id}, {Broadfoot, Rosatte, et al. 2001 20 /id}, and spatial-temporal statistical analysis {Rupprecht & Smith 1994 15 /id}, {Wilson, Bretsky, et al. 1997 12 /id}, {Moore 1999 13 /id}, {Childs, Curns, et al. 2001 132 /id} {Lucey, Russel, et al. 2002 148 /id} Tinline et al. 2002, {Guerra, Curns, et al. 2003 136 /id} {Jones, Curns, et al. 2003 146 /id} {Gordon, Curns, et al. 2004 137 /id}. Stochastic spatial simulation modelling has also been used as a predictive tool and for assessing the effects of rivers and mountains on the spread of the disease in Connecticut, New York and Ohio {Smith, Lucey, et al. 2002 110 /id}, {Russell, Smith, et al. 2004 280 /id} {Real, Russell, et al. 2005 293 /id}. An IBM approach has not been applied to the raccoon rabies system, though they have been used with success in other disease-host systems (e.g {Fa, Sharples, et al. 2001 296 /id} {Murray 2002 297 /id} Leung and Grenfell 2003, {Viet, Fourichon, et al. 2004 295 /id} {Bar-David, Lloyd-Smith, et al. 2006 294 /id}).

A primary motivation for modelling raccoon rabies with an IBM is to simulate disease spread as a spatial process. Many studies have reported an irregular wave of rabies spread across the landscape (Coyne et al. 1989) (Rupprecht and Smith 1994) (Wilson et al. 1997) (Moore 1999), (Childs et al. 2000) (Childs et al. 2001) (Lucey et al. 2002) (Smith et al. 2002) (Tinline et al. 2002) (Guerra et al. 2003), (Jones et al. 2003)

(Russel et al. 2003). More specifically, (Childs et al. 2001)) found significant differences in frequency, size and duration of epizootics among the mid-Atlantic States, such that the southern States had fewer, smaller, and shorter epizootics than the northern States. Moore (1999) used trend surface analysis to show the greatest rate of disease spread was from the south-central to northeast regions of Pennsylvania. It then slowed and progressed westwards towards Ohio.

Infectious disease transmission is a spatial process. Disease is transmitted along a network of individuals, and the spatial-temporal rate of spread is affected by heterogeneous environments and varying characteristics of host and disease populations (e.g. population density, disease susceptibility and virulence) and how these maybe affected by habitat types, landscape barriers, and disease control programs; (Voigt et al. 1985) (Fahse et al. 1998) {Murray 2002 297 /id}, {Bar-David, Lloyd-Smith, et al. 2006 294 /id}. IBMs are suitable for incorporating space because of their non-analytical framework, whereas, SVMs commonly define systems with differential equations and adding a spatial dimension often makes them unsolvable.

A second major motivation for using an IBM is the ability to define individual genetics and simulate their inheritance. Infectious disease models typically use empirical disease incidence data to construct and calibrate models or validate model outcomes. Unfortunately, the quality of disease incidence data is often poor, for reasons discussed in Chapter 5, increasing model outcome uncertainty. Hence, the ORM was further developed to simulate inheritance of genetic markers. Genetic measures from simulated genetic output (e.g. Φ_{ST} (Nei 1977)) can be compared with those derived from the

empirical raccoon genetic population structure. And in doing so, it is possible to estimate animal disease flow across the landscape for which genetic data are available.

Approaches to Modelling Raccoon Rabies

My thesis uses multiple approaches to modelling raccoon rabies. This was necessary to continue evaluating the ORM, if it is to be valued as a tool for understanding the raccoon-rabies system, since there were no published studies using the ORM when I began my research.

Parameterisation

My role in model development began with parameterisation. Model parameterisation defines the parameter input values. Ideally, the values should lie within the known ecological and biological variation. Furthermore, it is necessary to document the sources and justifications for parameter estimates used in the model to make it transparent for evaluation {Bart 1995 99 /id}, {Conroy, Cohen, et al. 1995 100 /id}. There are 20 parameters defining the fundamental disease-host dynamics in the ORM (Table 1.1). Their input values had been determined from OMNR, or OMNR-partnered (e.g. Canadian Food Inspection Agency) field and laboratory studies (e.g. (Rosatte et al. 1990) (Rosatte et al. 1992) (Wandeler and Salsberg 1999)).

Before commencing my research, initial parameter testing had been done in the context of variation of those data within the province. To further test the inherent theory in the ORM and to extend the spatial extent over which the model can be used I wanted to see how the initial parameter set compared with observed variation within North America. This was accomplished by developing and populating a Raccoon Ecology

Database (REDB) that contained peer-reviewed and unpublished ecological and biological data on the raccoon rabies system acquired from searching over 800 documents (Chapter 2). The REDB is available to researchers for meta-analyses and other ecological and biological explorations and this thesis discusses the role of this type of meta-data tool in other analyses.

Subsequent work with the ORM and the REDB confirmed the initial concerns of the ORM's builders that animal densities were a very important component in raccoon and raccoon-rabies ecology. Density is a critical factor in disease spread, hence is a fundamental component in infectious disease models (Anderson and May 1991). Furthermore, in the ORM, varying the target density of cells composing the virtual landscape is one means creating a landscape of different habitat.

Table 1.1 ORM parameters defining the fundamental disease-host dynamics. ORM age classes: young of year (0 – 52 weeks), juvenile (>52 - <75 weeks), adult (>75 weeks).

Parameter:
Male juvenile/adult dispersal distance
Female juvenile/adult dispersal distance
Male young of year dispersal distance
Female young of year dispersal distance
Average litter size
Mean percent mortality
Density dependent mortality control
Age of independence from mothers
Juvenile pregnancy rate
Adult pregnancy rate
Birth week
Litter size variance
Male juvenile/adult permissible movement period
Female juvenile/adult permissible movement period
Young of year permissible movement period
Target cell population density
Disease transmission rate
Mean incubation period
Contact rate
Time of infection

Consequently an important part of my work was to determine more accurate raccoon density estimates. Density data for the ORM has come from OMNR capture-mark-recapture studies. Density is calculated using a modified Lincoln-Petersen to estimate population size, *N* {Krebs 1989 235 /id}, and then dividing *N* by the entire area of the trapping cell. While this has been sufficient for the OMNR to monitor the raccoon rabies control program, other methods are available for calculating more accurate density estimates. Chapter 3 describes how distance sampling methods {Buckland, Anderson, et al. 2001 233 /id} were used to improve raccoon density estimates from OMNR capture-mark-recapture data. Program MARK {White & Burnham 1999 238 /id} was used to derive *N* using estimators that account for unequal capture rates relative to capture date, behaviour, unknown population heterogeneities and known factors (e.g. sex, age) {Otis, Burnham, et al. 1978 217 /id}, unlike the Lincoln-Petersen method which assumes equal capture rates {McCallum 2000 70 /id}. Density estimates were further improved by refining the area over which *N* is applied to account for the trap layout and raccoon movement characteristics. This was advantageous for reducing variation in habitat-density analyses, enabling the discrimination of raccoon densities relative to different types of rural landscapes in the St. Lawrence region of Ontario.

Sensitivity Analysis

The ORM was intentionally developed as an individual-based and complex model mirroring what was thought to be important biological processes underlying raccoon demography and rabies spread within raccoon populations. A more highly parameterised model enables detailed representations of the system being simulated, but also means there are more parameter values to estimate and contribute to outcome uncertainty. The

design philosophy of the ORM is that parameters/processes can be turned off or left as fixed values to simplify model input and experiments where possible (Tinline et al. 2007). Thus, some form of model sensitivity analysis (SA) is required to determine the importance of parameters by examining their impact on response variables chosen for the particular investigation.

I developed a novel SA approach using information theory methods (Chapter 4). Ideally a fully factorial design is used to define parameter input specifications for running models that test every combination of parameter values over the range of parameter space {Box, Hunter, et al. 1978 59 /id}. This is impractical for parameter-rich models because it would require m^n runs when checking m number of input values for n number of parameters {Voight, Tinline, et al. 1985 1 /id} {Blower & Dowlatabadi 1994 60 /id}. Instead, Latin hypercube sampling (LHS) was used as a more efficient method defining a set of parameter input specifications for multiple runs of the model that cover the full range of parameter space {McKay, Conover, et al. 1979 203 /id}; Chapter 4).

The next challenge was then to partial out the effects of parameters on model outcomes to determine if the parameters "sufficiently" contributed to characterising system behaviours. There are a variety of sensitivity testing procedures {Hamby 1994 68 /id},{Hamby 1995 69 /id}, though surprisingly none quantifiably assess whether the benefit of including a parameter to increase explanation of outcome variation outweighs the cost of increasing outcome uncertainty. Hence, this issue was explored using analytical methods founded on information theory, which also had the benefit of overcoming some inherent problems in frequentist statistical testing (Chapter 4).

In sum, the advantages of SA are to a) check for proper model functionality, b) identify parameters having the greatest impact on model outcomes to determine the parameters requiring the most accurate input values, c) increase understanding of the factors affecting the modelled system, and d) assess for model parsimony. SA was a particularly important step of ORM development because there has been a lack of empirical data to support or validate model results.

Genetic Simulation Modelling

Further developing the ORM through parameterisation and sensitivity analysis resulted in identifying and correcting coding errors (e.g. injecting rabies into a simulation when time of initial infection set to not occur but percentage of animals initially infected set to be greater zero), confirmed the accuracy of density input values, but also instilled more confidence in the ability of the ORM to model raccoon-rabies. Hence, in the belief of starting simply, and then validating before using the model or adding more complexity, I believed it was time to develop the ORM as a "genetic" simulation model.

Genetic data have been used in IBMs in two ways. One approach is to use functional genes, because they influence the fitness (or behaviours) of the individuals {Murray 2002 297 /id}, (Leung and Grenfell 2003)). For example, (Leung and Grenfell 2003) used an IBM to incorporate genetics in the form of disease resistant alleles, because this was found to be a critical factor for successfully modelling the epidemiological patterns of the coyote-scabies disease system. The second approach is to use neutral genes because they do not affect individual fitness, hence, experience no selective pressures. The ORM was further developed to define and implement mechanisms for maternal and bi-parental genetic inheritance of neutral markers. This

enables tracking of population-level movements because gene flow of these markers will only be indicative of dispersal and mating systems.

Modifications to the ORM required adapting the individual tracking processes and adding a mating system. This enabled using the ORM in a novel fashion to develop a method that has the potential to measure the impact of physiographic features on raccoon movement and potentially the spread of rabies. The specific application of this approach was to measure the effect of the Niagara River to raccoon movement, an area of great concern to raccoon rabies spread in Ontario and an area where appropriate genetic data were available (Chapter 5). This work had the additional benefit of providing further evidence to validate demographic behaviour within the ORM model and to develop further insight into the design of rabies control measures. Most importantly, this part of my research demonstrates the value of using an infectious disease model, which simulates genetic inheritance, to predict disease flow in an uninfected landscape.

Increasing Understanding through Model Development and Application

Thus, my infectious disease modelling approach has concentrated on testing the ORM, increasing its range of potential applications and, in doing so, examining how simulation modelling (and development) contributes to our understanding of disease-host systems.

Chapter 2

Raccoon Rabies Database: a meta-analysis tool for data storage, management and retrieval

Abstract

The value of scientific studies increases and is extended when their data are stored in a manageable and accessible format. This is demonstrated through development of a raccoon ecology database (REDB) to store, manage and disseminate available peer-reviewed and unpublished data on raccoon (*Procyon lotor*) biology, ecology and raccoon rabies, including citations for data sources. Over 800 documents were identified and citations for them entered into the database as literature references. Almost 1000 parameter values were entered from approximately 200 of these sources; these data included estimates of population density, survival rates, rabies incubation period, litter size, body weight, dispersal distance and home range size, often by age or sex class. Each datum is linked to a citation for its source, and to information about location and land use in the study area, time of year the study was undertaken, sample size, and variance. The relational database design enables querying and easy updating and manipulation of data. Hence, the REDB is well-suited for meta-analyses. The relational data model is presented and the application of the REDB to sensitivity testing of an individual-based, spatially explicit population model of raccoon rabies. Also given are example queries and interesting aspects of preliminary meta-analysis results. The REDB is a useful research tool that will increase in value with ongoing inclusion of data from

future raccoon and raccoon rabies studies and serves as a model for database design and research applications to other species.

Introduction

The synthesis of data from many independent studies into data warehouses encourages a more comprehensive analysis of ecological systems (Jones et al. 2006). This is useful for advancing ecological theory, for developing appropriate nature conservation strategies and extracting additional value from past research studies. Traditionally, ecological reviews were qualitative and used a standard "vote counting" approach. However, progress in the methodology of meta-analysis is increasing the ability of review studies to analyse and present their results more objectively and quantitatively (Arnqvist & Wooster 1995b). There are now well-established methodologies for properly undertaking such studies ((Gates 2002), (Roberts et al. 206)), and consequently the need for ecologists to systematically preserve and make their data available has become increasingly important and feasible (Michener et al. 1997, Michener 2006).

The multifaceted value of ecological databases is demonstrated through the creation of a raccoon (*Procyon lotor*) ecology database (REDB). The REDB was designed to store and manage parameters about raccoon biology, ecology and rabies collected from available peer-reviewed literature and unpublished "grey literature", and includes citations referencing the data sources. Parameters were collected primarily from ecological and biological field and laboratory studies undertaken in North America. Examples of the collected data include: mortality rates, litter size, home range size and raccoon rabies incubation period. Meta-data, such as sample size, a measure of the

variance for each demographic and disease parameter, sex and age class, time of year and geographic coordinates of the study area, also were collected.

All of the data were organised in a relational database. Relational databases store data in tables. A table represents an entity set, which is the generic structure of the objects being modelled in the database. Each row in the table contains data for an individual entry. Each column holds data for one attribute of the entity. A relational design enables tables to be linked to each other using a relational join. The function of a relational join is to link a table to another table by matching data values from a column or columns from one table to corresponding values in a column or columns in the other table. Column(s) in a table that uniquely identify each row are referred to as a *primary key*. The primary key is then used to establish a link with another table that has column(s) matching the primary key. These matching column(s) are referred to as a *foreign key*. Relational joins enable data to be queried across multiple tables.

There are many advantages to using a relational data model. Less data redundancy occurs than when data are stored in 2-dimensional data matrices (e.g. spreadsheets). Modifications to information are only necessary in tables that relate to that information. Relational databases are easy to use and implement compared to other data management systems (e.g. network and object-oriented systems). The relational approach is widely available through good proprietary software systems (e.g. Microsoft Access 2000®). Furthermore, extensive querying of the database is possible through a powerful and standardized query language facilities using Structured Query Language (SQL).

For these reasons, an important advantage of the REDB is that it is possible to query the database for numerous types of user-defined information. For example, queries can be run for i) raccoon densities grouped by rural, suburban and urban landscapes, ii) mortality rates for winter, spring, summer and fall, iii) sizes of male and female home ranges, or iv) litter size values reported at different geographic latitudes. The REDB is of value to wildlife biologists, ecologists and infectious disease researchers whose studies pertain to raccoons and raccoon rabies. Also, managers of raccoon populations and/or raccoon rabies disease control programs would benefit by having access to a wide range of data from many North American locations spanning decades of time to enable more informed decision-making. The REDB was used to determine parameter values and examine their variation within North America for sensitivity testing of a raccoon rabies individual-based spatial simulation model.

In this paper the REDB data model is presented, data are summarised and applied to a meta-analysis, and the utility of the REDB for sensitivity testing a simulation model is revealed. Similar studies have created databases that are also available for research (e.g.(Gachet *et al*. 2005), Onstad 2007). The REDB is a useful research tool that will increase in value with on-going inclusion of data from future raccoon and raccoon rabies studies.

Methods

An extensive literature search was performed using the ISI Web of Science® and Google Scholar ™ search engines. The aim was to obtain data from primary sources as opposed to review articles or books. Full citations of all raccoon and raccoon rabies literature sources were managed in ProCite version 5.0 (*www.procite.com*) and also

stored for the convenience of data users in the REDB. As many publications as were accessible from university and Ontario Ministry of Natural Resources (OMNR) provincial government libraries were reviewed. Statistics were collected for 56 different parameters (Table 2.1). Parameters were chosen based on their importance in defining raccoon biology, ecology and raccoon rabies disease dynamics. Parameters commonly reported in the reviewed studies were also included. For every parameter value entered into the database, ancillary data were collected pertaining to the value (e.g. sample size, variance, sex class) and study site (e.g. location coordinates, season) (See Figure 2.1. tables [T3_Parameter_Values] and [T2_Study_Area] for a complete list). A "comments" field in table [T3_Parameter_Values] was used to record information specific to each parameter value. For example, for the "movement_rate" parameter, the comments field was used to record if telemetry data were acquired during the hunting season, because hunting may affect raccoon movement. The "comments" field was also used to store verbal definitions of parameter values; for instance, providing the following detail about the "species range" parameter: "Assiniboine and Red River valleys limit the northern range of raccoons in Manitoba."

Table 2.1 REDB parameters and their description.

Parameter	Description
age_of_first_ movement_with_ mother	Age at which an offspring explores freely, but is still dependent on its mother
age_of_ indepedence_from_ mother	Age at which an offspring is no longer dependent on its mother for survival
age_of_weaning	Age at which an offspring is being weaned by its mother
age_ratio	Ratio of juveniles to adults
birth_period	Time of year when mothers are giving birth
body_condition	Body condition
body_weight	Body weight
body_weight_ change	Change in body weight over time
breeding_period	Time of year when animals are mating
capture_rate	Rate at which animals are captured in traps
chance_of_giving_ birth	Chance of a raccoon giving birth
consortship_ duration	Length of time spent courting and mating with another animal
consortship_ partners	Number of partners an individual has mated with during a breeding season
consortship_success	The percentage of consortships that resulted in the production of offspring
contact_rate	A measure of animals interacting among each other per unit time
core_home_range	The area of the most highly used portion of an animal's home range (e.g. daily activity space)
Dens	Number of dens used by a animal for a defined period of time (e.g. season, year)
Density	Density of animals: density = population size / area
density_ dependence	Documented factors regulating population levels of animals
disease_cycling_ period	Duration of a period in a cycling pattern of disease incidence in a population

Table 2.1. con't

Parameter	Description
disease_prevalence	Prevalence of disease among a group of animals (e.g. percentage, count)
dispersal_distance	Distance that an animal disperses
dispersal_period	Time of year an animal disperses from its mother
enzootic_duration	Length of time an established disease exists in an animal population
epizootic_period	Duration of time between cycling peaks of disease incidence
Fecundity	Offspring/individual productivity of an animal
first_epizootic_ duration	Length of time of the first epizootic (e.g. number of consecutive months when disease incidence is greater than median; Childs et al. 2000)
first_epizootic_ maximum	Number of rabid raccoons in the first epizootic (Childs et al. 2000)
first_epizootic_ period	Length of time of the first epizootic AND the interepizootic period before the start of the second epizootic (Childs et al. 2000)
Gestation	Length of the gestation period in days
habitat_selection	Proportion and/or type of habitat which the animal chooses to occupy
home_range	Area of the activity space of an animal
incubation period	Length of time an animal carries the disease from infection until they are symptomatic
infectious_period	Length of time an animal is symptomatic and can infect another individual
life_span	Length of time an animal lives
litter_size	Number of live offspring in a litter
mating_system	A numerical value indicating the type of mating system: 1 = polygynous (single male mates with multiple females), 2 = promiscuous (both males and females have multiple mates)
mortality_causes	Amount of mortality (e.g. count, probability) attributable to documented factors, as reported in the "comments" field
movement_rate	Rate of movement of animals

Table 2.1. con't

Parameter	Description
natural_immunity	Measure of animals natural immunity to a disease
neck_circumference	Circumference of an animal's neck
oestrus_duration	Length of time a female is in oestrus
oestrus_period	Time of year a female is in oestrus
overlapping_core_home_range	Area that animals' core home ranges overlap
overlapping_home_range	Area that animals' home ranges overlap
Parous	Percentage of females that have given birth more than once relative to all the sampled females
placental_scars	Number of placental scars
population_size	Number of animals within a population
positive_cases_time_to_peak	Length of time from initial outbreak to the time of maximum disease incidence
roadkill_index	Index of roadkills
sex_ratio	Ratio of males to females (or indicate if otherwise)
site_fidelity	Measure of site fidelity
species_range	Description of species range recorded in "comments" field
Survival	Survival of animals (e.g. proportion) for a defined period of time
tail_length	Tail length of an animal
total_length	Total length of an animal

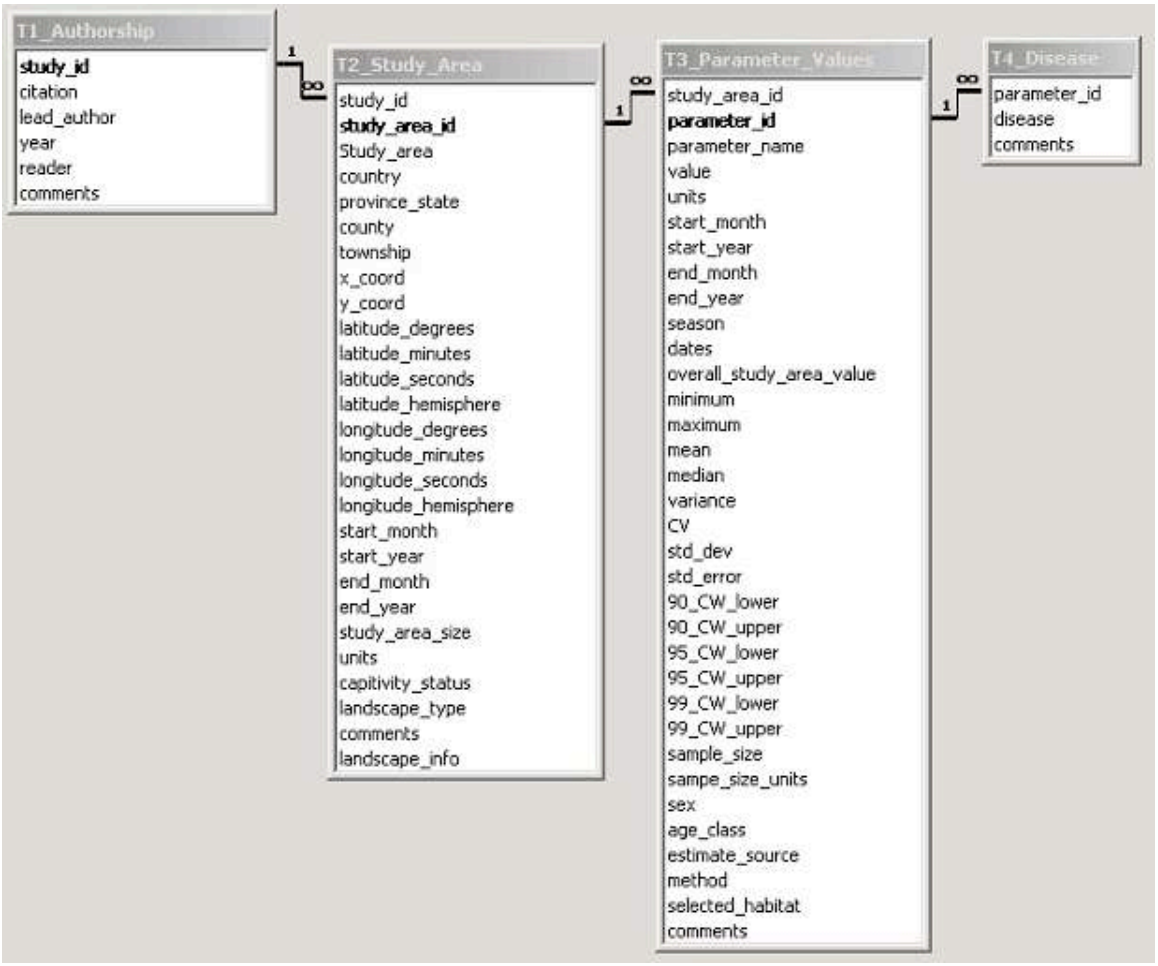


Figure 2.1 Relational database design between the four tables of the raccoon ecology database. Primary keys are in bold.

Multiple individuals reviewed studies to enter study sites and parameter information into the REDB. To check for bias and consistency in reader interpretations, all readers initially reviewed and entered data for the same 10 studies and the results were compared. Readers also entered study site and parameter information using data input forms built into the REDB to ensure consistency and minimise data entry errors. Commonly entered information was selected from predefined lists stored in the database in lookup tables. For example, "hectares", "square kilometres", or "square miles" could be selected from the units field.

Overview of the REDB

The data were organized and stored in a relational database using Microsoft Access 2000®. Four tables store the data. "T" prefixes the names of these tables. Lookup tables are prefixed by "LUP", and are linked to commonly entered fields in the data tables. One remaining table serves as a reference for users needing to convert calendar dates into Julian days [REFERENCE_Julian_dates], as was required for some of the temporal parameters (e.g. birth period, dispersal period) (Table 2.2).

Relational Database Design

The data tables are related to each other using one-to-many relationships (Figure 2.1). Starting with the [T1_Authorship] table, this means that every study reviewed is listed once in the table using the "study_id" field as the primary key. This field relates data from [T1_Authorship] to the [T2_Study_Area] table using one-to-many relationships, where the "study_id" field is the foreign key.

Table 2.2 Names and descriptions of the lookup and data tables found in the raccoon ecology database.

Table Name	Description
LUP_age_class	Age class categories
LUP_capitivity_status	Captivity status categories
LUP_country	Countries where studies were undertaken
LUP_data_entry_person	Names of people reviewing and entering data into the database
LUP_disease	Types of diseases potentially affecting the animals
LUP_estimate_source	Type of study undertaken (e.g. field, lab)
LUP_landscape	Types of landscape where studies were undertaken
LUP_latitude	"North" and "south" hemispheres for study location
LUP_longitude	"East" and "west" hemispheres for study location
LUP_method	Methods used to collect data (e.g. radio collars, mark recapture)
LUP_organisation	Organisation to which the person entering data into the database is affiliated
LUP_parameter_name	Parameter names for data values (See Table 2.1 for list)
LUP_period	Periods of time as months and seasons
LUP_season	Periods of time as the seasons and an overall annual period
LUP_sex	Gender codes (e.g. male, female, both, unknown, N/A)
LUP_state	Provinces in Canada and states in the USA
LUP_study_area_name	Names of common study areas
LUP_units	Common units pertaining to measurement data (e.g. "hectares" for area of home range)
LUP_USA_counties	All counties by state in the USA
LUP_year	Years from 1930 to 2010 from which studies have been (and will be) undertaken
REFERENCE_Julian_dates	User reference table containing calendar days of the year and the Julian date equivalent
T1_Authorship	Data table containing full citation information for every study found during the literature search, and for those subsequent studies that have parameter values stored in the REDB. This table also stores the names of the “readers” who reviewed the studies and entered the data into the REDB
T2_Study_Area	Data table containing information about the study area (e.g. geographic, time at which the study was undertaken)
T3_Parameter_Values	The main data table of the database that contains the biological and/or ecological information published in the study
T4_Disease	Data table listing the types of diseases present in the studied animals

Consequently, the [T2_Study_Area] table can have more than one entry of a "study_id" if there are multiple study areas associated with one publication. The "study_area_id" is the primary key in the [T2_Study_Area] table and is used to form a one-to-many relational join with the [T3_Parameter_Values] table through its foreign key, "study_area_id". Hence, one study area may have information pertaining to more than one parameter. There is a unique record for each statistic reported in the results section of a study, as identified by the "parameter_id" primary key field. The "parameter_id" field in the [T3_Parameter_Values] table is used to form a one-to-many relational join with the "parameter_id" foreign key field in the [T4_Disease] table. Thus, animal(s) from which the parameter value is derived can be affected by more than one disease.

Meta-Analysis

A meta-analysis was performed on data for home range size, density, and litter size. Within the REDB, the data for these parameter values were converted to common units (e.g. raccoons/km^2 for density, and hectares for home range).

A meta-analysis, analogous to a one-way analysis of variance (ANOVA) {Lipsey & Wilson 2001 270 /id}, was used to assess the effect of sex on home range size. Q_B was used as the test statistic since it accounts for the portion of variation in the dependent variable, home range size, which is explained by the categorical variable, sex. A weighted least squares regression model was used to test for the relationship between latitude and density. Both analyses used a mixed-effects model. Mixed-effect models account for variation that results from known factors within studies (e.g. sex, latitude, season), while still accounting for the existence of factors causing variation among studies that are not being explicitly modelled (e.g. different sampling protocols, unique

study area characteristics, unknown random factors) {Lipsey & Wilson 2001 270 /id}. Mixed-effect models are also among the most "appropriate, powerful, and informative for the analysis of weighted meta-analysis data" {Gurevitch & Hedges 1999 276 /id}.

Effect size (*ES*) and inverse variance weight (*w*) variables were calculated to standardise the results among studies so that they could objectively be compared. These variables were calculated as outlined by Lipsey and Wilson (2001) for arithmetic means:

$$ES = \overline{X} = \frac{\sum x_i}{n} \qquad \text{(Eqn. 1)}$$

$$w = \frac{1}{v} \qquad \text{(Eqn. 2)}$$

where x_i is an individual observation within a study, from $i = 1$ to n, n is the sample size of the observations and v is the variance of the observations. The weighted least squares mixed-effects regression models were performed in SPSS for Win/Version 11.0 as specified from macros provided by Lipsey and Wilson (2001) and available at http://mason.gmu.edu/~dwilsonb/ma.html.

A formal meta-analysis was not possible with the litter size data because too few studies reported information on the variation of the litter size statistic; thus, *w* could not be calculated. Insufficient data is common in meta-analysis, but does not always preclude analysis from the raw data, since it still can be informative {Gurevitch & Hedges 1999 276 /id}. Therefore, correlation analyses were used to test the correlation between degrees of latitude and litter size.

Sensitivity Analysis

The initial motivation for creating the REDB was to enable sensitivity testing of a raccoon rabies spatial simulation model, the Ontario Rabies Model (ORM). The ORM is an individual-based spatially explicit simulation disease-host model currently configured for raccoons. It simulates raccoon population dynamics, raccoon rabies viral transmission, and rabies control strategies. A sensitivity analysis procedure was developed to test model behaviour for the 16 demographic and 5 disease parameters that characterise raccoon rabies dynamics. Parameter input values were defined using Latin hypercube sampling (LHS) because it is substantially more efficient than using a fully factorial design when assessing highly parameterised models (McKay et al. 1979). LHS draws parameter values without replacement and with equal probability over the entire parameter space of its probability distribution function (pdf). Thus, it was necessary to obtain as much information as possible for each parameter to define their pdf, and the REDB was an efficient means of storing, managing and querying these data for sensitivity testing. In this paper, example pdf's are given for home range size by sex and litter size. Parameter pdf's were fitted using Palisade Corporation's @RISK software, version 4.5.4 (*www.palisade.com*). Furthermore, the REDB was used to demonstrate that the ORM default values, as defined from Ontario field data (e.g. Rosatte 2000), fall within the range of known variation for rural raccoons across North America.

Results

The search of literature pertinent to raccoon biology, ecology and raccoon rabies produced a list of 864 documents. Approximately 200 of these articles and books were obtained, and of these sources, 114 contained relevant data (993 parameters) that were

entered into the REDB. The data entry readers were consistent and accurate in their input of data from the initial review of the same 10 studies. The majority of published data available for entry into the REDB pertained to density, litter size, survival, home range, dispersal period and incubation period (Table 2.3).

The query created to retrieve information for the meta-analysis of home range size yielded 116 parameter estimates for home range size from the database. The SQL query and the Microsoft Access Design View version of the query are shown in Table 2.4 and Figure 2.2.

Meta-analysis indicated that males had significantly larger home ranges (HR) than females (HR_{male} = 352.1 hectares ± SE 52.5, HR_{female} = 128.1 hectares ± SE 52.6; $Q_{B\,(\text{one tailed, }\alpha=0.05)}$ = 9.1, *P*-value = 0.003). There was a significant negative relationship of latitude and raccoon densities (density = 584.2 – 12.3latitude, R^2 = 0.11, p = 0.04). Meta-analysis of litter size reveals a significant positive correlation with latitude (r = 0.38, p = 0.032, n = 32). The correlation is strongest when studies are restricted to those occurring in rural landscapes and pertains to adult reproduction, omitting juvenile mothers (r = 0.74, p < 0.001, n = 18); a second order polynomial model fits most closely to the data than linear, power or exponential models (Figure 2.3; Table 2.5).

Table 2.3 Parameter names and the associated number of useable studies from which data was collected to contribute to the REDB.

Parameter Name	Count	Parameter Name	Count
age_of_first_movement_with_mother	1	first_epizootic_period	1
age_of_indepedence_from_mother	1	Gestation	1
age_of_weaning	1	habitat_selection	7
age_ratio	2	home_range	18
birth_period	5	incubation_period	10
body_condition	1	infectious_period	4
body_weight	6	life_span	2
body_weight_loss	1	litter_size	22
breeding_period	2	mating_system	2
capture_rate	1	mortality_causes	6
chance_of_giving_birth	5	movement_rate	3
consortship_duration	1	natural_immunity	1
consortship_partners	1	neck_circumference	1
consortship_success	4	oestrus_duration	1
contact_rate	1	oestrus_period	1
core_home_range	3	overlapping_core_home_range	1
Dens	2	overlapping_home_range	3
Density	34	Parous	2
density_dependence	1	placental_scars	1
disease_cycling_period	1	population_size	7
disease_prevalence	6	positive_cases_time_to_peak	1
dispersal_distance	13	roadkill_indice	1
dispersal_period	1	sex_ratio	2
enzootic_duration	1	site_fidelity	1
epizootic_cycling	3	species_range	1
epizootic_period	1	Survival	19
Fecundity	2	tail_length	1
first_epizootic_duration	1	total_length	1
first_epizootic_maximum	1		

Table 2.4 SQL query used to retrieve home range size information from the REDB.

SQL Query
SELECT DISTINCT T3_Parameter_Values.parameter_name, T1_Authorship.lead_author, T1_Authorship.year, T2_Study_Area.landscape_type, T3_Parameter_Values.season, T3_Parameter_Values.sex, T3_Parameter_Values.age_class, T3_Parameter_Values.value, T3_Parameter_Values.units FROM (T1_Authorship INNER JOIN T2_Study_Area ON T1_Authorship.study_id = T2_Study_Area.study_id) INNER JOIN T3_Parameter_Values ON T2_Study_Area.study_area_id = T3_Parameter_Values.study_area_id WHERE (((T3_Parameter_Values.parameter_name)="home range" Or (T3_Parameter_Values.parameter_name)="home_range" Or (T3_Parameter_Values.parameter_name)="core_home_range")) ORDER BY T1_Authorship.lead_author, T1_Authorship.year, T2_Study_Area.landscape_type, T3_Parameter_Values.season, T3_Parameter_Values.sex, T3_Parameter_Values.age_class, T3_Parameter_Values.value;

Design View

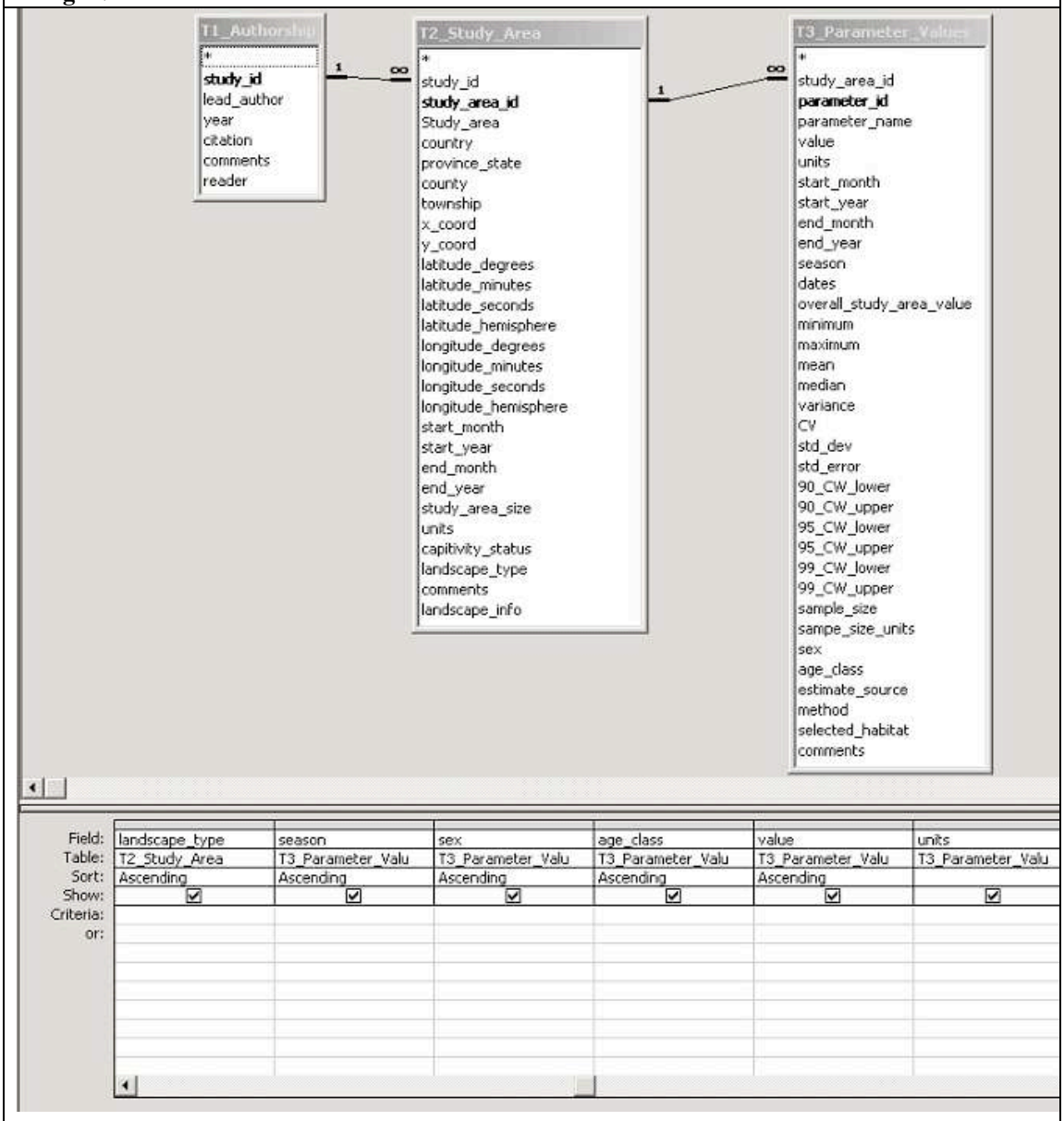


Figure 2.2 Design view of the query used to retrieve home range size information from the REDB.

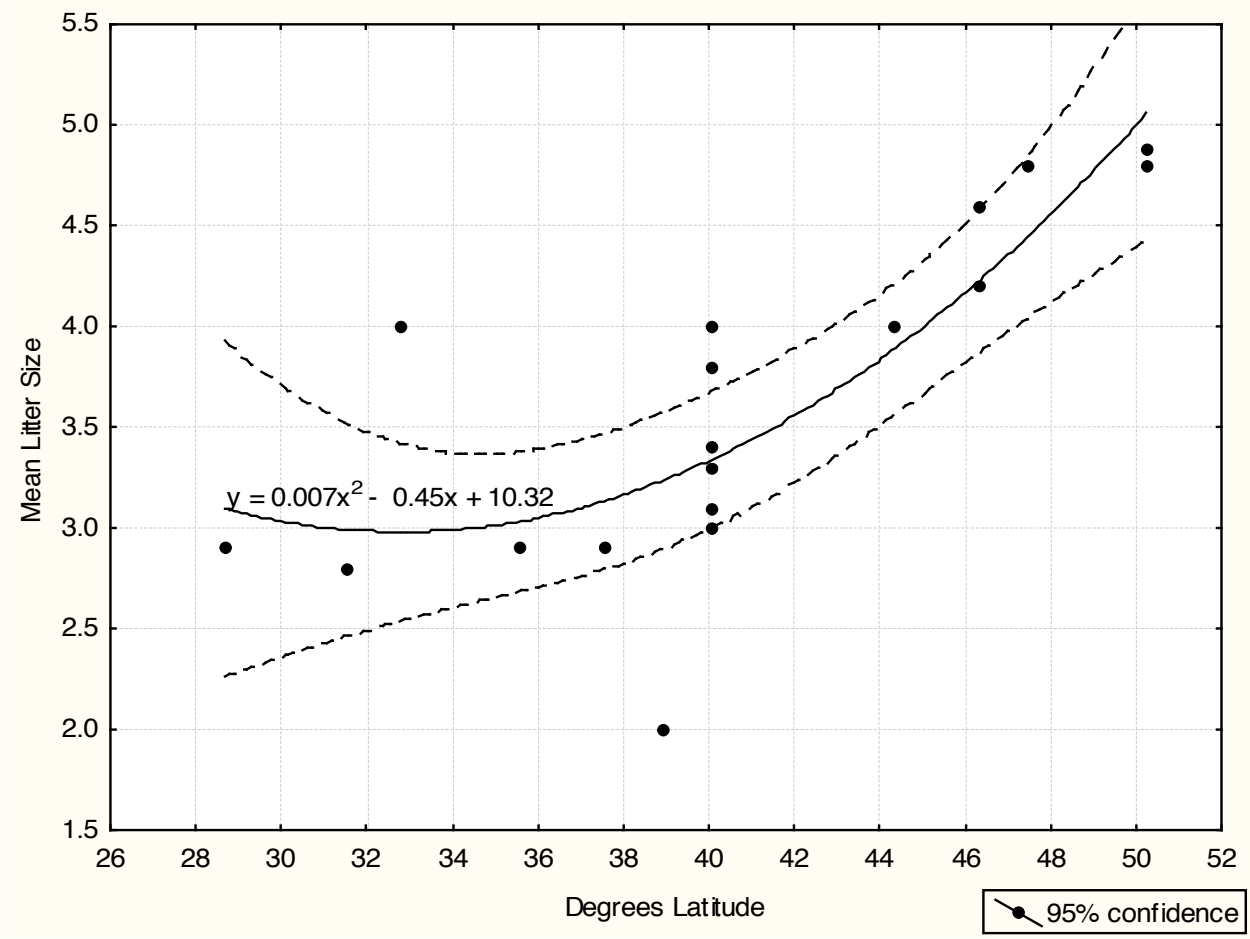


Figure 2.3 Scatterplot showing the correlation between the mean number of juveniles per litter and degrees latitude, and the associated 95% confidence bands.

Table 2.5 Parameters for which meta-analyses were conducted and the studies used for each meta-analysis.

Meta-analyses	**Studies**
Density	Broadfoot et al. 2001; Caro 2000; Dorney 1954; Endres and Smith 1993; Fritzell 1978b; Gehrt 2002; Gehrt and Fox 2004; Gehrt and Fritzell 1998; Hable et al. 1992; Hartman et al. 1997; Hoffmann and Gottschang 1977; Kennedy et al. 1986; Kissell and Kennedy 1992; Leberg 1988; McCleery et al. 2005; Nottingham 1982; Perry et al. 1989; Prange et al. 2004; Ratnaswamy et al. 1997; Ratnayeke et al. 2002; Riley et al. 1998; Rivest and Bergeron 1981; Schubert et al. 1998; Seidensticker et al. 1988; Smith and Engeman 2002; Smith et al. 1994; Sonenshine 1972; Stevens et al. 1995; Twichell 1949; Urban 1970
Dispersal distance	Belant 1992; Cowan 1973; Fritzell 1978b; Gehrt and Fritzell 1998; Hartman and Eastman 1999; Hoffmann and Gottschang 1977; Lynch 1967; Mosillo et al. 1999; Rosatte et al. 1991; Rosatte et al. 1992; Seidensticker et al. 1988; Tabatabai and Kennedy 1989
Home range	Chamberlain et al. 2003; Chamberlain and Leopold 2002; Cowan 1973; Fritzell 1978b; Gehrt and Fox 2004; Gehrt and Fritzell 1997; Gehrt and Fritzell 1996; Hodges et al. 2000; Hoffmann and Gottschang 1977; Kamler and Gipson 2003; Prange et al. 2004; Ratnayeke et al. 2002; Rosatte 2000; Rosatte and MacInnes 1989; Rosatte et al. 1991; Roscoe et al. 1998; Totton et al. 2004; Urban 1970
Litter size	Asano et al. 2003; Bigler 1981; Broadfoot et al. 2001; Cagle 1949; Cowan 1973; Fiero and Verts 1986; Fritzell 1978a; Junge and Sanderson 1982; McKeever 1958; Mech 1966; Ritke 1990b; Ritke 1990a; Rosatte 2000; Rosatte et al. 1991; Sanderson 1987; Sanderson 1950; Scheffer 1950; Schneider et al. 1971; Smith 1985; Stuewer 1943; Wood 1955; Zeveloff 1981

The pdf's for male and female home range sizes illustrate that males have significantly larger home ranges than females, as determined by meta-analysis of raccoon home-range size (Figure 2.4). The pdf of litter size indicates that most studies report a litter size of 4 juveniles per litter, and that litter size ranges between 2 to 5 juveniles per litter for the North American studies reviewed (Figure 2.5). ORM default values are found to lie within the variation for rural raccoons across North America (Table 2.6).

Discussion

The relational design of the REDB follows the well-established relational data model for storing parameters from available literature, and enabled subsequent flexible management and manipulation of these data. The REDB was initially used to extract parameter values for sensitivity testing of the ORM, to define biologically appropriate values of the parameters. Biologically-defined values are necessary for determining processes that most significantly affect model outcomes (Ginot *et al*. 2006). This also means that the effect of the parameters on model outcomes can be interpreted in a more biologically meaningful context. This is more appropriate for a primarily mechanistic biological model than halving or doubling parameter inputs as described in Voigt et al. (1985), because even a parameter of minor importance could impact model results given a large enough input value. Doubling or halving input values has the potential to create an unrealistic scenario for the natural system being modelled, and might yield results that are less informative in terms of understanding the true biological effect of the parameter being examined.

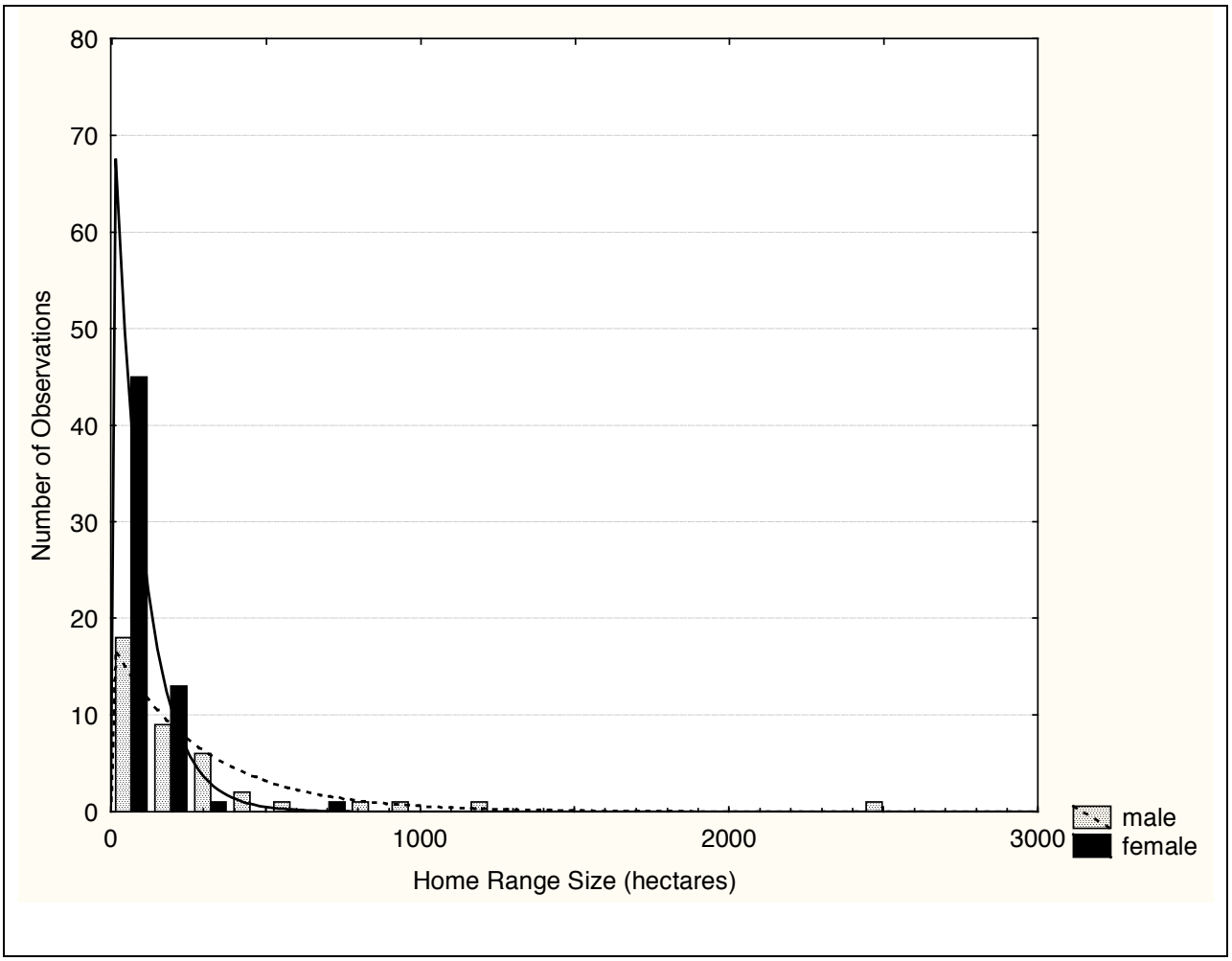


Figure 2.4 Probability distribution function of male and female home range sizes.

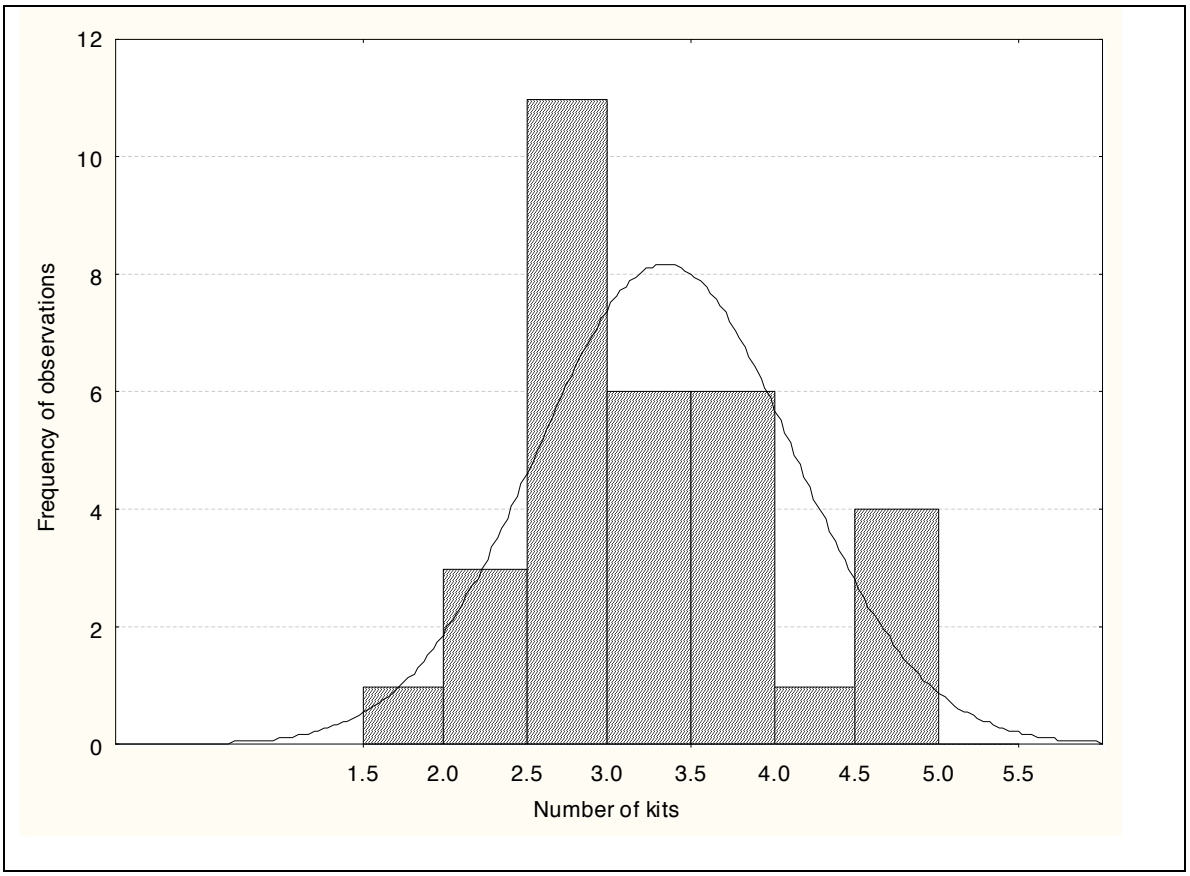


Figure 2.5 Probability distribution function of litter size reported in 32 studies.

Table 2.6 ORM parameter default values from Ontario field data and range of values found for rural raccoons in the REDB. *mean of values defined by a probability distribution function; unlisted REDB range values were not definable using the REDB.

Parameter	**ORM default**	**REDB range**
Annual rural raccoon density	5 raccoons / km^2	4.6 – 13.6 raccoons / km^2
Age of independence	20 weeks	20 – 36 weeks
Adult age	75 weeks	-
Year 0 male (female) mortality rate	0.6	sub-adult: 0.4 – 0.51
Year 1 male (female) mortality rate	0.4	
Year 2 male (female) mortality rate	0.3	
Year 3 male (female) mortality rate	0.3	
Year 4 male (female) mortality rate	0.3	
Year 5 male (female) mortality rate	0.6	
Year 6 male (female) mortality rate	0.6	
Year 7 male (female) mortality rate	0.6	adult: 0.3 – 0.9
Mortality adjuster	0.2	-
Birth week	calendar week 18 (end of April / beginning of May)	mid-end April to end of May
Juvenile birth rate	60%	29 – 66%
Adult birth rate	95%	73 – 100%
Average litter size	4	2 – 5
Litter size variance	1	1
Sex ratio at birth	50:50 male:female	50:50 male:female
Male juvenile/adult movement weeks	calendar weeks: 8-43 (spring, summer, autumn)	no movement during severe winters
Female juvenile/adult movement weeks	calendar weeks: 12-17, 38-43 (spring, autumn)	no movement during severe winters
Male young-of-year movement weeks	calendar weeks: 38-43 (autumn)	no movement during severe winters; disperse July to November
Female young-of-year movement weeks	calendar weeks: 38-43 (autumn)	no movement during severe winters; disperse July to November

Table 2.6. con't

Parameter	**ORM default**	**REDB range**
Juvenile/adult male dispersal distance*	1.36 km	1.2 – 10.2 km
Juvenile/adult female dispersal distance*	0.68 km	0.3 – 8.5 km
Young-of-year male dispersal distance*	2.12 km	0.2 – 16.1 km
Young-of-year female dispersal distance*	1.02 km	0.4 – 6.3 km
Incubation period*	6.47 weeks	1 – 6 weeks
Infectious period*	1.00 week	1 week
Chance of spread	1.5%	-
Contact within cell	77.8%	-

Furthermore, the REDB confirms that default ORM values fall within the range of known raccoon variation. The ORM was built using the expertise of field biologists who have studied Ontario raccoons. To increase credibility in model design and results beyond the builders it is necessary to document sources of model design as extensively as possible {Bart 1995 99 /id}, {Conroy, Cohen, et al. 1995 100 /id}.

Another important and more common benefit of databases like the REDB is the contribution that they make through meta-analyses. Science is advanced by comparing new ideas to old ideas ((Arnqvist & Wooster 1995b), (Gates 2002)), and in the case of the REDB, the analysis of raccoon biology, ecology and raccoon rabies is advanced by enabling comparisons among different landscapes, seasons, age classes, sexes and geographic locations. Results from a single study may seem unimportant, but may appear more meaningful when observed in the context of other studies. For instance, meta-analysis of litter size versus latitude clearly demonstrates a trend of larger litter size at higher latitudes (Figure 3.2). Positive relationships between litter size and latitude has been reported in many species (e.g. Conaway et al. 1974, Innes 1978, Cockburn et al. 1983, Bilenca et al. 1994). A possible casual mechanism causing this relationship is attributed to higher latitudes having more extreme environments, causing increased mortality risks due to seasonal variation in food, hence, a shorter breeding period. To increase fitness, animals must invest more per reproductive event, such as, by having 1 large litter. This is opposed to have 2 or 3 smaller litters at lower latitudes, spread over a longer period of time coinciding with more favourable weather (Zeveloff 2002, Rademaker and Cerqueira 2006).

Latitude also had a significant negative correlation with density. This result may be consequence of the geographic sampling distribution of the studies. Species are expected to have higher densities at the core of their range (Hengeveld and Haeck 1982, Williams et al 2003), because of habitat limitation at range peripheries (Erb and Boyce 1999). The raccoon species range extends as far south as Central America (Zeveloff 2002). Meta-analysis for this study did not include studies further south than the United States. Thus, it is conceivable that the negative correlation between density and latitude is caused by sampling raccoons at the core of the range in the United States, and then northwards to studies sampled at the periphery of their range in Canada.

With regards to home range size, males were found to have significantly larger home ranges than females. This is also expected since raccoons are mostly solitary carnivores ((Kaufmann 1982), (Sanderson 1987)) that demonstrate sexual differences in behavioural strategies for habitat use and breeding success ((Fritzell 1978b), (Gehrt and Fritzell 1997), (Kamler and Gipson 2003)). Furthermore, male raccoons tend to be larger than female raccoons (Zeveloff 2002), and the positive correlation between larger body size and home range size has been reported in other species (Grigione et al. 2002, Anderson et al. 2006), as is likely due to larger energetic requirements needed to support a greater mass (White et al. 2007).

An initial review of the same 10 studies by the data entry readers was important to ensure quality data collection and enhance the database design. This exercise demonstrated that readers were similarly thorough and accurate in their data input. Also, it adjusted the design of the database to be more efficient and helped formulate more useful lookup tables. Despite this, there are a number of cautions for using data compiled

from multiple studies (Arnqvist & Wooster 1995b), (Kotiaho 2002)): 1) They can present a biased estimate of a true effect. This has largely been attributed to the "file drawer problem" {Bauchau 1997 266 /id}. Since most published studies report significant effects, the data used in the meta-analysis can be unrepresentative of the true system (Kotiaho 2002). This bias can be countered by collecting data from grey literature, unpublished Master's or PhD theses and government reports (Roberts *et al*. 206). An attempt was made to include as much grey literature as possible; unfortunately, searching for grey literature requires far more effort than acquiring data from published literature. However, because grey literature is not peer-reviewed, and is therefore potentially more erroneous, it could be viewed as a benefit to meta-analyses to limit data from these sources or account for an effect of grey literature on the variation of reported results. 2) The research question or interpretation of data can bias selection of data for entry into the database. The REDB is largely free from this bias because all available data were collected without filtering to predetermined ecological questions. Furthermore, several observers collected and entered data, hence reducing the potential bias of reader-misinterpretation. 3) Poor quality studies will degrade the value of the meta-analysis. These studies need to be identified through objective measures and eliminated from analysis. Roberts et al. (2006) recommends "assessing the experimental design, implementation and analysis" of each study being included in the meta-analysis. It was difficult to rigorously assess every study, given the high number included in the database; however, a field was included in the database to "comment" on any cautions or peculiarities of a study.

Health sciences have made the greatest contribution, of all the scientific disciplines, to advancing the quality of meta-analysis. The Cochrane Collaboration details 27 criteria for undertaking and reporting meta-analyses (Higgins and Green 2005). Gates (2002) and Roberts et al. (2006) advocate a similar standard to improve the quality of ecological meta-analyses. These standards should be applied in the creation of ecological databases and in subsequent meta-analyses.

Conclusions

The availability of the REDB and other similar databases contribute greatly to the exhaustive effort required to locate, store, manage and query data needed for meta-analyses. Results from meta-analyses are frequently used by conservation and wildlife managers, biologists and ecologists ((Pullin *et al*. 2004)). Ecology benefits from databases like REDB by increasing the amount of data accessible for comparative and broad-scale analyses and thus, increasing the value of science's investment in past studies.

The REDB itself offers a robust and simple data model for compilation, maintenance and analysis of ecological parameters from published studies. The database, and an empty database for used with other species, composed of empty data tables, lookup tables and their relationships, are available from the author (erin.rees@nrdpfc.ca).

Chapter 3

A new approach to density estimation: exploring the density-area relationship

Abstract

Knowledge of animal densities is invaluable for understanding, managing and conserving wildlife and ecological systems. Density, D, is commonly calculated from capture-mark-recapture (CMR) data using estimates of population size, N, and effective trap area, A_E, as $D = N/A_E$. Since a major difficulty of this approach is determining A_E, its estimation procedure is refined to be more representative of the true area corresponding to N. A distribution of distances characterising consecutive captures of raccoons (*Procyon lotor*) from a trapping program in Ontario, Canada, are used to develop the buffered trap area (BTA) approach. The density estimates were within the range for North America and those derived using an established direct density estimation procedure. Negative density-area relationships are characteristics of study designs that compromise the accuracy of density estimates. This trend is detected for "crude" density estimates that used the entire trapping cell area for A_E, but not for the BTA estimates. Further advantages of BTA were an explicit estimation of N, A_E and its geographic boundary, trap configuration, overall capture probability, CMR distances and density. These estimates enable assessment of factors such as habitat, season or trap configuration on the methodological parameters and permit more precise exploration of ecological relationships.

Introduction

Animal population density estimates are essential for wildlife biologists and ecologists. Knowledge of densities is used to understand animal population dynamics and how they may be affected by other species, disease, habitat quality and anthropogenic factors. This information is fundamental to furthering understanding of principles of population dynamics and for creating effective conservation and management strategies.

Animal density *(D)* is typically calculated indirectly as the ratio of population size *(N)* to area over which the population estimate applies *(A)*: $D = N/A$. The relationship between *D* and *A* is of particular interest in studies of ecological factors affecting animal densities and population extents. Gaston and Matter (2002) categorise two types of *D-A* relationships: "patch" individuals-area relationships (PIARs) and "generalized" individuals-area relationships (GIARs). In PIARs, *A* is defined as an isolated habitat patch, island, or experimentally created habitat where the population exists. PIARs can be used to explore whether species density differs with the size of a patch. *D-A* relationships have been hypothesized to be non-existent, positive or negative by the equilibrium theory of island biogeography (MacArthur and Wilson 1967), density compensation (MacArthur and Wilson 1972) and resource concentration (Root 1973), respectively. PIARs have been used to understand the scale of habitat use by a species, and this information is used to determine whether a species is better protected using fewer large habitats instead of many small habitats (see the single-large or several-small (SLOSS) debate, Simberloff 1988).

GIARs differ because *A* is defined simply as the area over which density is calculated, irrespective of habitat. GIARs can reveal how various estimates of *A* affect density estimates ((Blackburn and Gaston 1996)(Gaston and Matter 2002), (Mayor and Schaefer 2005)). Capture-mark-recapture (CMR) data are used from a raccoon rabies control program to calculate D-A_E relationships, where A_E is the effective trap area. A negative GIAR is indicative of underestimated density estimates as a consequence of methodological problems; 1) inclusion of non-habitat or low density areas; 2) not accounting for the edge effect (trapping animals whose home ranges are entirely outside of the grid, but are attracted to the traps, or partially enclosed within the boundary of the grid; (Dice 1938)) and the fact that smaller areas are more affected than larger areas because of having a larger edge area to core area ratio; 3) the tendency to establish larger A_E's when animals occur at low densities or are rare and 4) the decline in sampling efficiency with increasing A_E (Conner et al. 2000, Gaston et al. 1999, Gaston and Matter 2002, Matter 2000).

The first issue is most obvious when *N* is applied to an administrative unit such as a park boundary rather than to the habitat occupied by the population of interest or the specific area subject to trapping. There is a greater likelihood of including areas of non-habitat, meta-population patches currently uninhabited or untrapped areas (Bender et al. 1998, Conner et al. 2000, Gaston et al. 1999). Distance sampling techniques (Buckland et al. 2001, Efford 2004, Parmenter et al. 2003) can be used to minimise this problem because they estimate *N* for the area over which the organism of interest could be detected.

The simplest approach for accounting for the edge effect is to use only data from the interior traps of a trapping grid (Hansson 1969). This provides a relatively unbiased estimate but causes a significant loss of data. A more widely used strategy is to increase A_E by adding a boundary strip around the trapped area, but then the difficulty lies in choosing an appropriate strip width (*W)* (Wilson and Anderson 1985). The value chosen for *W* is one-half the home range radius (Dice 1938); however, this strategy simply shifts the challenge to defining a suitable radius for home range. More recently Efford (2004) proposed a distance–based strategy where density is estimated directly from CMR data, without calculating A_E, by using a spatial simulation model and inverse prediction (SSIP). A model is run repetitively for random permutations of input: *D*, number of home range centres in the model landscape, g_0, maximum probability of capture, and σ, maximum distance over which the probability of capture varies from a trap. The aim is to minimise the deviation between simulated and calculated values of *N*, $\hat{p}$, estimated capture probability, and $\overline{d}$, mean successive capture-recapture distance per animal. In addition to accounting for edge effect, SSIP reduces the issues of rare species and sampling inefficiency respectively, by avoiding the requirement to explicitly define A_E in deriving density.

In this study GIARs are used to explore the accuracy of density estimates, using data from a raccoon (*Procyon lotor*) CMR program in south-eastern Ontario, where raccoons were trapped in multiple trapping cells (Rosatte et al. 2001). A "crude" density was estimated with A_E defined by the area of the entire trapping cell and was compared to a novel density estimator that refines estimation of A_E to produce a more representative estimate of density. This strategy used a detection function curve fitting facility of the

distance sampling program DISTANCE (Thomas et al. 2004) to estimate A_E based the frequency distribution of CMR distances and on the overall probability of capture (*OPC*) of all traps, as calculated in Program MARK (White and Burnham 1999). Program MARK was also used to estimate N, so density could be calculated as: $D = N/A_E$. The method presented in this study for estimating density is referred to as the buffered trap area (BTA) approach.

Crude density estimates are expected to have a negative GIAR because of the discussed methodological problems. BTA was further assessed by comparing its density estimates to those produced by SSIP, using DENSITY 3.3 software (Efford et al. 2004). SSIP is an elegant direct approach to density estimation, but does not explicitly estimate A_E nor delineate the A_E spatially. In contrast, the BTA method is explicit in the definition of N and A_E and its spatial expression. This enabled exploring effects of trap layout, habitat, season, trapping cell and year on CMR distances and *OPC* and account for significant effects that could improve the estimate of A_E. Furthermore, an explicit empirically derived definition of A_E that is more ecologically related to raccoon activity space enables a clearer exploration of habitat effects, and addresses the contentious issue of estimating an appropriate A_E for density calculation.

Methods

Four years of CMR data, from 2000 to 2003, were acquired from the Ontario Ministry of Natural Resources (OMNR) St. Lawrence trap-vaccinate-release and point-infection-control programs located in eastern Ontario (45°N 75°W). These programs are designed to prevent spread of raccoon rabies from New York State into Ontario (Rosatte et al. 2001). Trapping occurred in contiguous trapping cells generally 10 km^2 that were

roughly rectangular in shape over an area approximately 760 km^2 (excluding water bodies; Figure 3.1). Standard protocol required trapping each cell once a year for two consecutive weeks. The trapping season occurred during May-November, when raccoons were most active. Live traps were placed randomly though out the area, with approx. 100 traps per trapping cell. Trap configurations were often irregular because of trap placement constraints, such as private land and habitat heterogeneity (Figure 3.1). For each animal trap location, age class, sex, and weight were recorded, and uniquely-coded eartags were applied.

Density estimates were calculated by trapping cell to produce a sufficiently fine spatial grain. At this scale, Program MARK (White and Burnham 1999) was used to estimate *N* for each trapping cell, per year, using the CMR encounter history data. *N* was calculated from multiple closed-capture population estimators (M_0, M_t, M_b, M_h, M_{th}, M_{bh}, M_{tb}, M_{tbh} (Otis et al. 1978), to account for factors potentially influencing density: temporal, *t*, (trap night), behavioural, *b*, (trap happy / shy) and unknown, *h*, and known individual heterogeneities (sex, age).

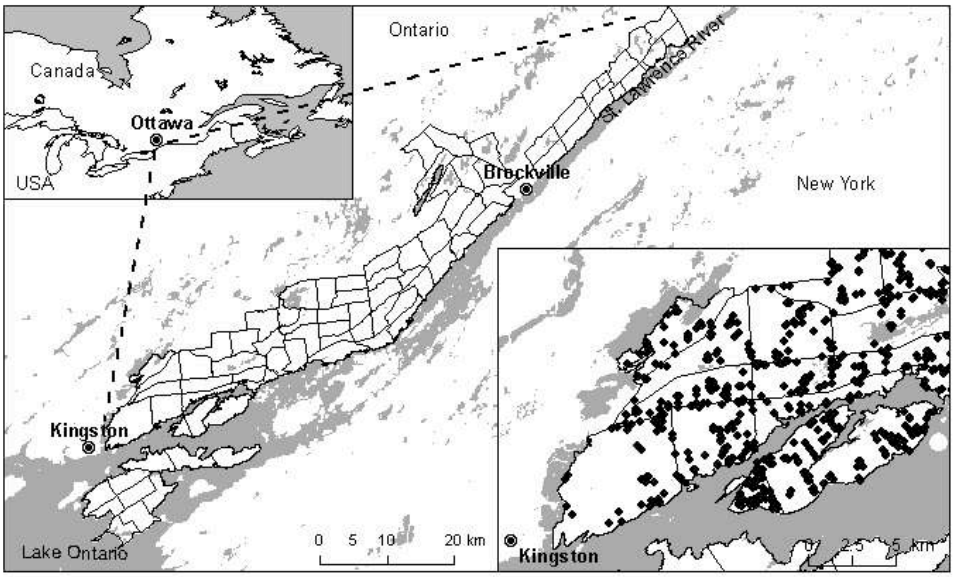


Figure 3.1 Trapping cell configuration for 2000, and an example of trap locations in cells near Kingston, Ontario.

Models were ranked using Akaike's Information Criterion (AIC) (Burnham and Anderson 2002). A final estimate of N was achieved by using model-averaged coefficient estimates (Burnham and Anderson 2002). Dividing the trapping cell N by trapping cell area produced crude density estimates for each trapping cell.

BTA also used Program MARK estimates of N, but calculated A_E using CMR distances of raccoons and their OPC, the latter as estimated by Program MARK. The first step in deriving A_E was to produce frequency distributions of CMR distances compiled for any animal caught two or more times (Figure 3.2), at the scale of the entire study area and on a yearly basis. To assess whether the trap configuration (density and spatial layout) influenced the CMR distances of raccoons, because high trap densities or clustering of traps might result in smaller CMR distances and higher OPC. If this were true, the data would need to be normalized to account for trap configuration. To test for this, the spatial configuration of neighbouring traps around each trap from which an animal was released were quantified using a spatial autocovariate (Weiner 1982),(Weiner 1984); Eqn. 1):

$$SA_i = \sum_{\substack{j=1 \\ (i \neq j)}}^{T} \frac{1}{d_{ij}^{\alpha}} \qquad \text{Equation 1}$$

where SA_i is the spatial autocovariate of a trap i relative to all the other traps j, (j = 1 to T, $i \neq j$, T is the number of traps), that might recapture the animal in the trapping cell, d_{ij} is the distance between traps i and j, ($i \neq j$), and α is a weighting factor of α = 0.5, 1 and 2, where higher values decrease the effect of distance on SA_i (Weiner 1982, 1984).

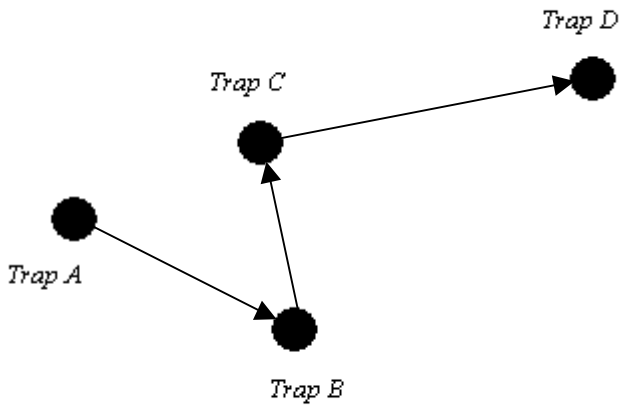


Figure 3.2 An example of successive CMR distances (AB, BC, CD) of one animal captured at traps A, B, C then D.

To explore the effect of neighbouring traps, quantified by the spatial autocovariate in relation to CMR distances, the correlation between the distribution of mean CMR distance raccoons travel from each trap and the spatial autocovariates were tested, for each level of α. Both variables were log transformed to be normally distributed, and linear correlations were performed using simple Mantel's tests (Mantel 1967).

The Mantel's test is appropriate for assessing correlations between distance measures because it takes into account a spatial effect of nearer samples being more similar than distant samples. Furthermore, this test uses a random permutation procedure to define the null hypothesis distribution; thus, does not require the assumptions of parametric tests. To remove the effect of trap configuration, CMR distances were adjusted by using residuals of the CMR distance predicted by the best fitting model of the spatial autocovariate.

$$\text{CMR}_{\text{adjusted}} = 10^{\left(mean(\log(\overline{CMR})+residual\right)} \qquad \text{Equation 2}$$

where $\text{CMR}_{\text{adjusted}}$ is the adjusted CMR distances; $\overline{CMR}$ is mean CMR distance at a trap and residual is:

$$\text{residual} = \log(\text{CMR}) - (a + b(\log(\text{spatial autocovariate})) \qquad \text{Equation 3}$$

where a is y-intercept and b is the slope of the relationship between mean CMR distance by trap and the spatial autocovariate. After the adjustment was made to the CMR data, these data were averaged per animal, per trapping cell, so that each animal contributed equally to the subsequent analysis occurring at the level of the trapping cell.

The next step was to determine whether it was justifiable to calculate A_E at a coarser spatial grain than the individual trapping cell to increase sample size of CMR data per spatial unit. Analysis of variance (ANOVA) was used to test for effects of year,

habitat and season on CMR distances and *OPC*, calculated as a mean overall measure for each trapping cell. To determine the habitat factor levels for the ANOVA, satellite landcover classes were aggregated into water (wetlands and open water), urban, agriculture and forest. The proportion of each habitat type was calculated for each trapping cell. Cluster analysis was then used to group the cells into composite landcover classes, which maximized their distinctiveness. This analysis resulted in two levels for the ANOVA: predominantly forest (>50%) and predominantly agriculture (>60%). Trapping seasons for the ANOVA were partitioned into "nursing" (May, June), "weaning" (July, August) and "dispersal" (September, October, November) based on seasonal variation in raccoon movements (Sanderson 1987). If the ANOVA results indicated significant yearly, habitat and seasonal effects then multiple A_E's were calculated for each of the factor levels. This would also require calculation of densities specific to the factor levels and enable ANOVA testing for the effects of these factors on the density estimates. Conversely, in the absence of significant effects, one A_E would be calculated from all trapping cell data across all years, seasons and habitats.

A_E was estimated using Program DISTANCE version 4.1 release 2 (Thomas et al. 2004) to fit a curve to the frequency distribution of CMR distances, averaged per animal. The program default uses a maximum probability capture, $g_0 = 1$, when distance from the trap is zero. The fitted function defined the decline in capture probability as distance from the trap increased. With the known function, it was analytically possible for the software to calculate the area under the curve and derive an effective trap radius, r. Since g_0 is likely below 1, because a 100% capture success is not expected directly at the trap, g_0 was reduced to *OPC*, while holding the area under the curve constant to determine an

adjusted r. Then by assuming the trap-specific effective trap area, A_{trap}, was circular, r was used as the radius to calculate the A_{trap} (Figure 3.3).

All traps are used in the calculation of A_E, including traps that did not capture raccoons, because the number of available traps influences the frequency of captures. Some traps were discovered to have coordinates which positioned them outside their known trapping cell. These traps should neither contribute a whole A_{trap} to calculation of A_E , since they would likely overlap the A_{trap} of other traps, nor should they make no contribution to A_E. Random locations of a trap were defined within a trapping cell for each trap which was mapped outside its identified trapping cell. The A_E for a trap cell was calculated as the area enclosed by the boundary of the spatial union of A_{trap}, for all traps in the trap cell. The resulting overall A_E for each trap cell was then used to estimate raccoon densities: $D = N / A_E$, and produce the BTA estimates of density.

BTA density and A_E estimates at the trapping cell level defined a GIAR, for each year of data. At the same spatial and temporal scale, GIARs were estimated for the crude density estimates using A_E, defined as the entire trapping cell area. While the logic may seem circular to use the correlation of density with A_E when A_E is a component of the dependent variable, Prairie and Bird (1989) and Gaston et al. (1999) indicate that this analysis is permissible when the variables for correlation do not violate the assumptions of correlation analysis, "are meaningful, that is, they represent the concepts of interest and not just a component of them" and "do not share a large measurement error term"

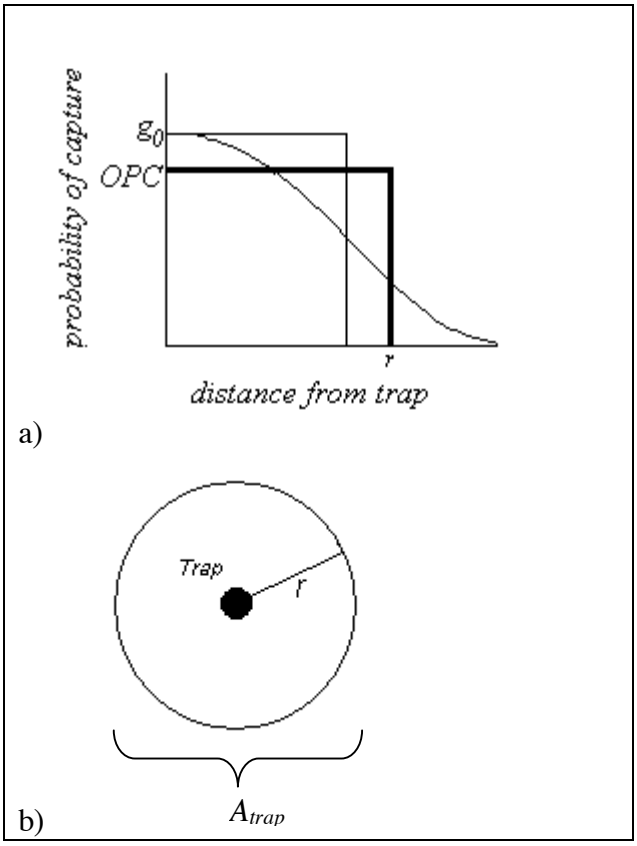


Figure 3.3 a) The area under the curve of the function modelling the frequency distribution of CMR distances is used to determine the effective trap radius, r, b) which defines A_{trap}, assuming it is circular.

(Prairie and Bird 1989); all of which applied to the present study.

Comparing BTA density estimates to SSIP estimates generated using Program DENSITY 3.3 (Efford et al. 2004) were further BTA values. This program required trap location and trap encounter history data, to estimate density at the level of trapping cells. Several population estimators were considered. To choose a final estimator, AIC model comparison facility in DENSITY 3.3 ranked estimators, and the top ranked estimator was used to calculate population size and, subsequently, density. SSIP estimates of A_E were produced by rearranging $D = N / A_E$. Values of D, N and A_E were calculated for 27 trapping cells from the year 2000 data. BTA and SSIP estimates of D, N, and A_E were compared for these trapping cells using basic statistical descriptors (mean, minimum, maximum and standard error) and correlation analyses between the two methodologies.

Results

The top N estimators indicated by AIC rankings differed among trapping cells. In general, the models providing strongest inference were mixture models, M_h, that account for unknown individual heterogeneity (Pledger 2000), temporal models, M_t, that account for variation attributable to trap night (e.g. weather), and behavioural models, M_b, that differentiate between initial and recapture behaviours (e.g. trap happy / shy). Models accounting for individual covariates, sex and age, were less significant (Table 3.1).

Table 3.1 Percentage of models selected as AIC best model for estimating population size. Model factors were time (Mt), behavioural difference (Mb) between capture (p) and recapture rates (c), unknown heterogeneity between two groups (A and B), in the sense of Pledger's (2000) mixture models (Mh), individual covariates (sex, age), and any combination of all of these factors.

Dominant Model	**Percentage**
(Mbh p_A(.) p_B(.) c_A(.) c_B(.)}	9.7
(Mth p_A(time) p_B(time)}	9.7
(Mt p(time)}	9.2
(Mth p_A(time sex age) p_B(time sex age)}	6.0
(Mth p_A(time sex) p_B(time sex)}	6.0
(Mtbh p_A(time) p_B(time) c_A(time) c_B(time)}	5.5
(M0 p(.)}	5.1
(Mb p(.) c(.)}	4.1
(Mt p(time age)}	4.1
(Mt p(time sex)}	4.1
(Mtbh p_A(time age) p_B(time age) c_A(time age) c_B(time age)}	3.7
(Mbh p_A(age) p_B(age) c_A(age) c_B(age)}	2.8
(Mbh p_A(sex age) p_B(sex age) c_A(sex age) c_B(sex age)}	2.8
(Mh p_A(.) p_B(.)}	2.8
(Mt p(time sex age)}	2.8
(Mb p(age) c(age)}	2.3
(Mh p_A(age) p_B(age)}	2.3
(Mth p_A(time age) p_B(time age)}	2.3
(Mtb p(time) c(time)}	1.8
(Mtbh p_A(time sex) p_B(time sex) c_A(time sex) c_B(time sex)}	1.8
(M0 p(age)}	1.4
(Mb p(sex) c(sex)}	1.4
(Mbh p_A(sex) p_B(sex) c_A(sex) c_B(sex)}	1.4
(Mh p_A(sex) p_B(sex)}	1.4
(Mtbh p_A(time sex age) p_B(time sex age) c_A(time sex age) c_B(time sex age)}	1.4
(Mb p(sex age) c(sex age)}	0.9
(Mh p_A(sex age) p_B(sex age)}	0.9
(Mtb p(time age) c(time age)}	0.9
(M0 p(sex)}	0.5
(Mtb p(time sex age) c(time sex age)}	0.5
(Mtb p(time sex) c(time sex)}	0.5

Trap configuration significantly influenced the CMR distances (Table 3.2). The linear relationship regressing the log transformed spatial autocovariates, calculated with $\alpha = 2$, and the log transformed mean CMR distances was used to adjust the successive CMR distances to account for trap configuration, since this relationship had the strongest significant correlation ($r = -0.35$, $P < 0.0001$).

Year, habitat and season influenced mean successive CMR distances, whereas only season affected *OPC* (Table 3.3). The significant effects of year, habitat and season on mean trapping cell CMR distances, and of season on *OPC*, necessitated calculating effective trap radii and trap-specific effective trap area (A_{trap}), for each permutation of factor levels (Table 3.4). The resulting effective trap radii and A_{trap} ranged from 613.8 m and 1.18 km^2 (year: 2000, habitat: forest, season: dispersal) to 1615.1 m and 8.19 km^2 (year: 2000, habitat: agriculture, season: nursing) with a mean of 899.4 m and 2.74 km^2. Buffering traps by the effective trap radius delineates the A_E boundary to extend several hundred metres beyond the border of trapping cells, and results in a few small areas within the trapping cell that are not covered by A_E (Figure 3.4). Seasonal and habitat specific BTA density estimates fluctuated over the four years (Figure 3.5). Over all years and seasons, forest and agriculture estimates of density were significantly different ($F=4.58$, $P=0.03$), and over all years and habitat types, trapping season significantly affected density estimates ($F=3.16$, $P=0.04$; Table 3.5).

Table 3.2 Simple Mantel test results correlating distributions of spatial autocovariates, indexing trap configuration, and the mean CMR distance travelled from each trap, for three spatial autocovariate a levels.

correlation coefficient	***P*-value**	**α level**
-0.16	< 0.001	0.5
-0.33	< 0.0001	1
-0.35	< 0.0001	2

Table 3.3 Significant three-way ANOVA results from testing the effects of year, season and habitat on mean trapping cell CMR distance, in metres, and *OPC*. The mean, standard error (SE) and sample size (n) of trapping cells for factor levels are also given.

Data	Factor	F-ratio	*P*-value	Level	Mean	SE	n
mean trapping cell CMR distance	Year	75.45	0.00	2000	700.10	37.17	63
				2001	1865.98	102.49	67
				2002	1655.97	76.88	69
				2003	731.75	45.46	45
	Season	7.09	0.001	nursing	1332.47	222.08	14
				weaning	1303.81	75.31	101
				dispersal	1286.65	71.66	129
	Habitat	6.00	0.01	forest	1131.08	57.81	137
				agriculture	1508.03	84.12	107
OPC	Season	3.88	0.022	nursing	0.53	0.02	13
				weaning	0.66	0.003	98
				dispersal	0.58	0.004	126

Table 3.4 Effective trap radii (ETR), individual effective trap area (A_{trap}), mean density ($\overline{D}$) and standard deviation (stdev) relative to the levels of year, season and habitat.

Season	Habitat	Year	ETR (m)	A_{trap} (km^2)	$\overline{D}$	stdev
		2000	1141.40	4.09	8.46	5.79
		2001	1269.98	5.07	5.45	1.48
		2002	-	-	-	-
	forest	2003	-	-	-	-
		2000	1615.09	8.19	6.24	3.97
		2001	1023.57	3.29	4.69	1.41
		2002	-	-	-	-
nursing	agriculture	2003	-	-	-	-
		2000	750.95	1.77	7.29	2.39
		2001	1166.23	4.27	4.57	2.84
		2002	661.24	1.37	7.47	3.15
	forest	2003	762.41	1.83	5.86	2.11
		2000	766.62	1.85	8.11	4.41
		2001	701.77	1.55	10.08	3.73
		2002	814.89	2.09	5.96	3.36
weaning	agriculture	2003	698.32	1.53	10.62	5.24
		2000	613.76	1.18	10.03	5.73
		2001	725.43	1.65	6.08	3.36
		2002	620.09	1.21	9.12	6.89
	forest	2003	791.60	1.97	3.93	2.09
		2000	991.78	3.09	5.65	3.62
		2001	1030.36	3.34	2.75	2.24
		2002	1015.81	3.24	4.82	2.60
dispersal	agriculture	2003	827.22	2.15	4.32	2.60

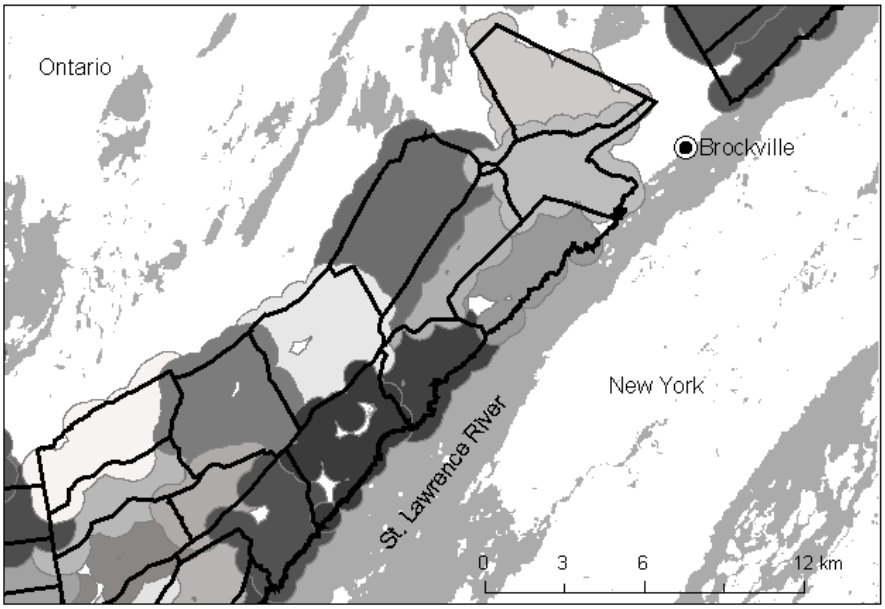


Figure 3.4 An example of the refined effective trapping cell areas produced by the BTA method for data from 2002, near Brockville, Ontario. The trapping cell boundaries for the 2002 trapping program are also shown.

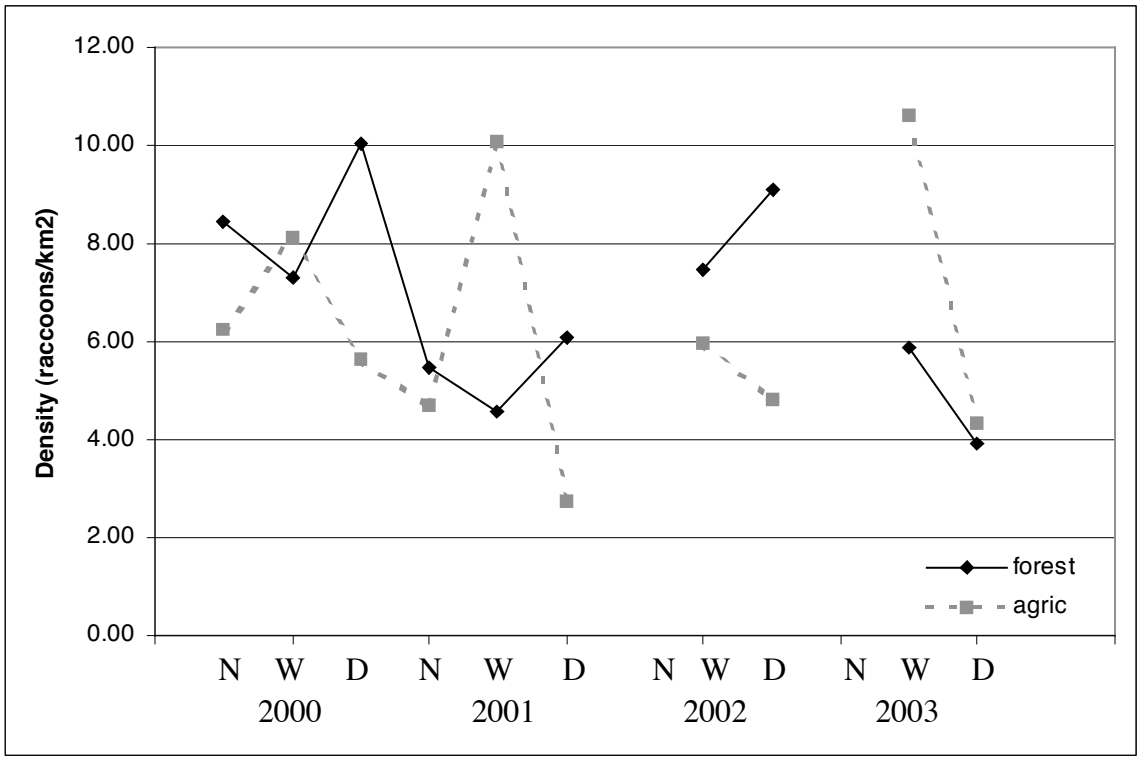


Figure 3.5 Trend in BTA density estimates over time for estimates specific to habitat (forest, agriculture) and season (nursing: N, weaning: W, dispersal: D).

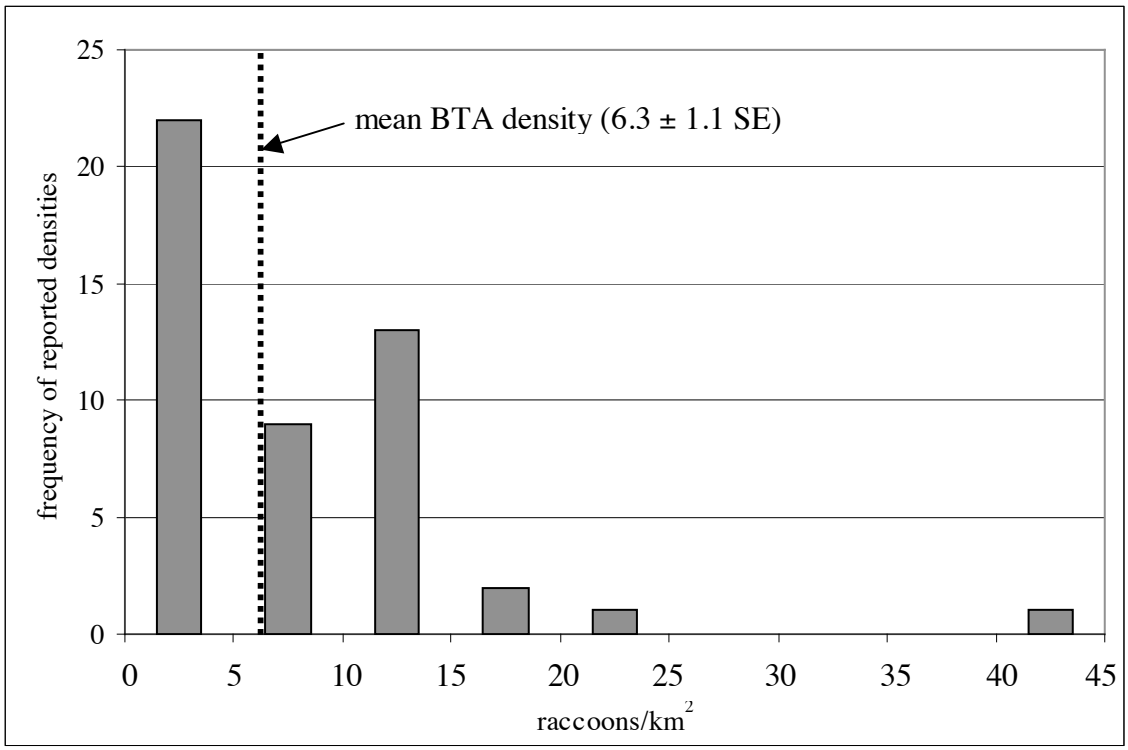


Figure 3.6 Frequency of density values (raccoons/km^2) reported for rural areas across North America in comparison with the mean BTA density.

Table 3.5 Habitat and season specific BTA estimates of density, D, (raccoons/km^2).

Factor	Level	Mean D	Standard error	Minimum	Maximum	Sample size
Habitat	forest	6.38	0.44	0.83	16.74	54
	agriculture	5.00	0.46	1.13	12.56	41
Trapping Season	nursing	6.96	1.19	3.34	18.17	14
	weaning	7.07	0.35	1.38	19.9	100
	dispersal	5.70	0.41	0.70	30.0	127

BTA density estimates were lower than crude estimates, with yearly means ranging from 5 to 8 raccoons/km^2, while the crude estimates ranged from 10 to 15 raccoons/km^2 (Table 3.6). Correlation analysis of the GIARs for BTA showed no significant relationship, whereas, crude density GIARs demonstrated a significant negative relationship for all years except for 2003 (Table 3.7). There was a significant association between BTA and SSIP estimates of D and N, and to a lesser extent with A_E (Tables 3.8 and 3.9).

Discussion

Negative GIARs produced by correlating crude density estimates and the entire trapping cell areas are attributed to the: 1) inclusion of non-habitat or low density areas; 2) not accounting for the edge effect; 3) tendency to establish larger A_E's when animals occur at low densities or are rare and 4) decline in sampling efficiency with increasing A_E. BTA density estimates are more accurate because its GIARs do not exhibit a significant negative correlation between density and A_E, and density estimates agree more highly with those produced by SSIP. BTA estimates of 5 to 8 raccoons/km^2 (6.3 ± 1.1 SE) are also within the range of densities reported for rural areas throughout North America (7.5 ± 1.1 SE; Figure 3.6; Dorney 1954, Sonenshine and Elton L. Winslow 1972, Hoffmann and Gottschang 1977, Fritzell 1978, Rivest and Bergeron 1981, Nottingham et al. 1982, Kennedy et al. 1986, Leberg 1988, Seidensticker et al. 1988, Perry et al. 1989, Hable et al. 1992, Kissell and Kennedy 1992, Smith et al. 1994, Smith et al. 1994, Stevens et al. 1995, Endres and Smith 1993, Hartman et al. 1997,

Table 3.6 Summary statistics for the BTA and crude densities (raccoons/km^2) for all four years of data.

Year	Method	Mean density	Standard error	Minimum	Maximum
2000	**BTA**	7.6	2.3	2.4	23.0
	Crude	15.2	11.9	3.7	54.3
2001	**BTA**	4.7	1.4	0.7	16.1
	Crude	9.6	5.6	1.1	26.2
2002	**BTA**	7.1	2.7	1.4	30.0
	Crude	12.2	10.3	2.1	55.6
2003	**BTA**	5.8	1.9	0.8	19.9
	Crude	9.6	4.1	1.4	21.6

Table 3.7 Correlation coefficient and *P*-values from the GIAR correlation analysis for each year of data. *Significant correlation

Year	**Dataset**	**Correlation (*P*-value; sample size, n)**
2000	BTA	-0.17 (*P* = 0.176; n = 63)
	Crude	-0.35 (*P*=0.005; n=63)*
2001	BTA	-0.20 (*P* = 0.114; n = 66)
	Crude	-0.26 (*P*=0.035; n=65)*
2002	BTA	-0.15 (*P* = 0.21; n = 68)
	Crude	-0.27 (*P*=0.025; n=68)*
2003	BTA	-0.14 (*P* = 0.36; n = 45)
	Crude	-0.22 (*P*=0.129; n=45)

Table 3.8 Correlation coefficients and *P*-values for the correlation between BTA and SSIP, and Crude and SSIP estimates of a) population size and b) density values and c) area, for 27 trapping cells in year 2000. *Significant correlation

	a) Population Size	**b) Density**	**c) Effective Trap Area**
BTA vs SSIP	0.72 (*P*<0.001)*	0.68 (*P*<0.001)*	0.45 (*P*=0.02)*
Crude vs. SSIP	0.72 (*P*<0.001)*	0.37 (*P*=0.06)	0.11 (*P*=0.60)

Table 3.9 Basic statistical descriptors of the estimates of a) density, b) population size and c) effective trap area for 27 trapping cells in year 2000 for BTA and SSIP.

Parameter	**Estimator**	**Mean**	**Minimum**	**Maximum**	**Standard error**
a) Effective trap area (km^2)	SSIP	18.3	2.60	41.7	1.9
	BTA	21.8	10.0	37.6	1.3
b) Population size	SSIP	150.5	49.00	444.0	17.0
	BTA	186.7	45.56	666.3	26.1
c) Density (raccoons/km^2)	SSIP	11.1	2.32	37.2	1.8
	BTA	8.8	2.4	23.0	1.0

Ratnaswamy et al. 1997, Gehrt and Fritzell 1998, Caro et al. 2000, Rosatte et al. 2001, Gehrt 2002, Ratnayeke et al. 2002, Gehrt 2004, Prange et al. 2004, McCleery et al. 2005, Rosatte et al. 2007a, 2007b, 2007c).

Density estimates are strongly influenced by study design and in particular by the definition of the study area. BTA estimates of density differed from the crude approach by using an estimate of area that was more reflective of the true area where raccoons are susceptible to trapping. BTA is a trap-centric approach. The frequency distribution of mean CMR distances can be modelled as a three-dimensional continuous surface defining the probability of a trap to capture an animal. Maximum capture likelihoods exist as peaks in the surface, whose locations correspond with trap locations and have a magnitude of 1, defined as g_0 in the DISTANCE software (Thomas et al. 1994). The surface extends infinitely outwards because there is always a possibility of capturing animals, though, the likelihood becomes increasingly negligible with distance. The BTA approach generalises this continuous surface into a discrete definition of a geographically bounded A_E over which to apply the estimated N to derive density. The generalization reduces the maximum probability of capture to *OPC* (area "A", Figure 3.7) and defines a distance over which it applies. This does not preclude that animals outside A_E are not being captured or do not exist; rather, the tail of the distribution extending beyond the effective trap radius represents their probability of capture (area "C", Figure 3.7), and their presence is evidenced by the CMR data that creates the tail and is used in formulating the A_E generalization.

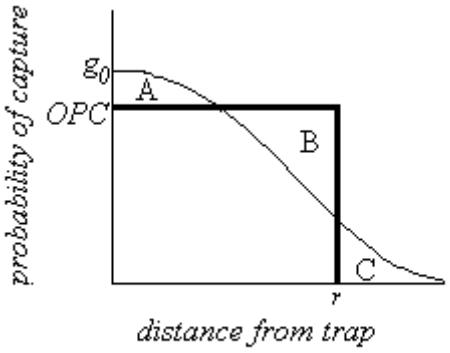


Figure 3.7 The continuous probability surface modelling the frequency distribution of CMR distances, with a maximum probability of capture, g_0, and its generalisation to a discrete function, bounded by the overall probability of capture (*OPC*) and the effective trap radius (*r*). The area within the rectangle is equal to that under the curve; the sum of the area of regions "A" and "C" is equivalent to the area of region "B".

BTA's ability to define an A_E reflective of population presence accounts for the edge effect, and this is the major benefit of BTA for the trap configuration used in this study. There were few areas within the trapping cell that were not covered by the A_E, making it very similar to using the entire trapping cell area for estimating crude densities (Figure 3.4). However, the BTA approach produced larger A_E's per trapping cell than the crude method because BTA A_E's extended beyond trapping cell boundaries (Figure 3.4). This addresses the edge effect by recognizing that traps close to the trapping cell boundary are likely capturing raccoons from outside of the trapping cell. Furthermore, the "reach" of a trap to capture these animals is defined by movement data of the animals in the study. A negative GIAR relationship can develop by not accounting for the edge effect because this effect has a greater impact on smaller trapping areas since the edge:area ratio is larger. There is a higher contribution of animals attracted to the trapping area relative to the trapping area then would occur for larger trapping areas; thus, smaller trapping areas would experience a more inflated estimate of N than larger trapping areas.

Problems that might arise from surveying scarce or rare animals and issues of sampling efficiency were not expected to affect density estimates for this study. Raccoons are not scarce animals in the St. Lawrence region and the size of the trapping cells was developed considering the need for trappers to maximize capture rates.

Thus in light of the GIAR analysis, the BTA density estimates are more accurate representations of the area on which the population estimates are based than the crude density estimates, as further supported by the agreement with the SSIP density estimates. Efford's (2004) SSIP procedure is an effective means of calculating density; however, it

is the differences in estimation flexibility and the parameters estimated that make BTA a more attractive method than SSIP when there is an interest in further ecological analysis. BTA explicitly delineates the geographic extents of A_E (Figure 3.4), and this is more easily utilised by ecologically focused management programs and habitat analyses. The habitat composition within this refined A_E serves as a more accurate representation of the habitat of the animals on which the population estimate is based than one based on the entire trapping cell because areas that do not contribute to the population size estimate would be excluded. This would decrease the noise when investigating habitat to density relationships.

Another advantage of BTA is that Program MARK is used as a more flexible means of estimating *N* than that which can be accomplished using the DENSITY 3.3 software (Efford et al. 2004). Program MARK offers a wider array of population estimators that are able to account for variation in capture probabilities attributable to time, behaviour and unknown individual heterogeneities (Otis et al. 1978). Program MARK also enables covariates to be included in the population estimators that can account for heterogeneity caused by known biological differences in individuals (e.g. sex, age class). Models that accounted for variation in capture rates were most often the best predictive model (Table 3.1). It is important to account for a wide array of factors causing variation in capture rates that affect *N* estimation to prevent spurious estimates of *N* (Boulanger et al. 2004), (Conn et al. 2006). Program MARK can also produce an overall model-weighted estimate of *N*. In this approach, AIC was used to estimate the likelihood of each model being the best representation of variation given the data and

other candidate models. These likelihoods were used to derive weightings for calculating a composite model averaged value of *N*.

A further benefit of BTA is enabling more comprehensive ecological analyses than SSIP data because data from BTA calculations are transparent. BTA produces explicit estimates of *N*, trap configuration, *OPC*, A_E and density, and frequency distributions for mean CMR distances. It is possible to test for factors believed to have a significant effect on these variables. Year, trapping season and habitat are hypothesised to influence *OPC* and mean CMR distances, affecting the derivation of A_E and subsequent density estimates. These hypotheses were not rejected except for the effects of year and season on *OPC*, resulting in multiple A_E's and densities being defined for the different levels of the effects. However, if effects were not found to be significant, data could be pooled for the derivation of A_E, which beneficially serves to increase sample size for estimation of A_E. Pooling data advantageously supplements areas where capture rates are low or trapping is restricted to small temporal windows, thus, increasing the precision of A_E and reducing its variation.

Ecological analyses using BTA data found raccoons to have higher mean CMR distances in 2001 and 2002 (Table 3.3), possibly a consequence of intensive raccoon depopulation efforts in the study area during 1999 and 2000 in response a raccoon rabies outbreak (Rosatte et al. 2001). There may have been a time-delayed increase in raccoon movements, such that by 2001 and 2002 the raccoons discovered the depopulated voids and moved greater distances to inhabit them, which was reflected in greater mean CMR distances in those years (Rosatte et al. 2007b). Other possible explanations for yearly

effects may be due to annual weather variation or differences in agriculture crops and their ability to attract foraging raccoons (Schneider et al. 1971).

Season significantly affected mean CMR distances. Distance decreased from springtime nursing to summertime weaning to a low in autumn when raccoons dispersed. This result is in accordance with known raccoon ecology, where for the same region, Totton et al (2004) found summer raccoon home ranges to be larger than fall home ranges. It is believed that raccoons decrease their movements in the fall to conserve their physiological energy stores in preparation for winter denning when food resource availability is lowest (Rosatte 2000). A significant effect of trapping season was also found for density (Table 3.5). Densities are expected to decrease over the trapping season as mortalities reduce the population from the spring birthing season. The trend was reproduced from data in this study, except "nursing" densities were slightly lower than "weaning" densities. This is attributed to a small sample size of cells trapped during the nursing season, and also that raccoons may have been harder to capture during this season because newly born young are within dens.

Density is a critical factor influencing whether an infectious disease becomes established or burns itself out (Anderson and May 1991). More accurate density estimates will help determine the threshold density of susceptible animals (Rosatte et al 2007b) below which the disease will not spread ((Kermack and McKendrick 1927), (Anderson and May 1985)). In areas where raccoons exceed this density, rabies control measures (e.g. oral vaccination baiting, trap-vaccinate-release, depopulation) can decrease the susceptible population below the threshold. Furthermore, season and habitat specific estimates of effective trap radii and A_{trap} give an indication of the size of raccoon

activity space relative to these factors in south-eastern Ontario. The larger these values, the greater the activity space. Consequently, it is more likely a raccoon will come in contact with other raccoons because it is travelling greater distances (Totton et al 2002), which also means it can spread the disease further afield. Therefore, disease control would need to be applied to a sufficiently large area to at least encompass raccoon activity space. Also, it would be possible to investigate directional biases in capture locations and to assess this in relation to landscape features (e.g. water barriers, topography, habitat quality). Thus, coupling an understanding of factors influencing movement patterns with the knowledge of season and habitat specific densities would aid the development of rabies control programs targeting areas of high movement and densities, as well as contribute to spatial ecological studies that explore dispersal and foraging behaviours.

The inevitable issue arising from using BTA is whether CMR data sufficiently represents movement patterns of animals to derive A_E. CMR data only give "snapshots" of the spatio-temporal activity of animals. The actual travel path of raccoons between capture locations is unknown. This means it is possible that captured animals use areas outside of the refined A_E but within the trapping cell. The degree to which this occurs can be explored by tracking animals at finer spatial and temporal resolutions using Global Positioning System collars. An assessment can then be made whether finer grained location data improve the estimate of A_E beyond that calculated using CMR data.

Conclusions

CMR studies are commonly employed for investigating wildlife population dynamics. Unfortunately competing research questions and/or limited resources can

hamper the ability to design a trapping grid suitable for density estimation. A common circumstance is contending with a trapping configuration that inadequately measures population size for a defined area. The uneven spatial distribution in traps of this study may have left gaps in the landscape where raccoons were present, but had no chance of being trapped. Thus, it was inappropriate to apply *N* to the entire trapping cell to calculate density; however, as the entire area raccoons travel before capture is not known, the true effect this had on density estimation is also unknown. BTA addressed the edge issue by using movement data to define areas beyond the trapping cell potentially contributing animals to *N* to be included in the estimation of A_E. BTA makes feasible the exploitation of data sets not specifically designed for estimating *N* or density, such as from disease monitoring programs (e.g(Henning et al. 2006)). Changes caused by improving density estimates could have important consequences for operational and policy decisions (Rosatte et al. 2007a, 2000b, 2000c) and for use in wildlife models such as the Ontario Rabies Model (Tinline et al. 2007).

BTA and SSIP are applicable to any trapping design (e.g. grid, web, line, or irregular trap placement). Investment in time and the software requirements of BTA are greater than SSIP because Program MARK is used to estimate *N*, and a Geographic Information System is needed to calculate the A_E. However, the more segmented approach of estimating density with BTA exposes data intermediary to density calculations that are useful for addressing interesting ecological questions. This is valuable to researchers exploring seasonal, habitat or landscape effects on animal movements and capture probabilities, and their implications for conservation, disease spread, or climate change induced range shifts.

Chapter 4

Sensitivity analysis of a raccoon rabies disease simulation model using Information Criteria

Abstract

Sensitivity analysis is a fundamental step in model development. Akaike's and Bayesian information criteria are used in combination with regression analysis to examine the relative contribution of 17 demographic and 3 disease parameters that characterise raccoon rabies behaviour in the Ontario Rabies Model (ORM). Two different types of information criteria were used to more comprehensively evaluate whether the benefit of including a parameter to increase the explanation of outcome variation outweighed the cost of increasing outcome uncertainty. Univariate results were most useful because the sample size of the multivariate analysis was too small to discern parameter effects. Overall, an assessment of parsimony was achieved by identifying parameters *a priori* defined to be superfluous and parameters representing system components not found to affect system dynamics (e.g. separating birth rates of juveniles and adults). The ORM was found to be functioning as intended (e.g. population density and transmission rate are critical factors of disease dynamics), and ecological insight was gained from the sensitivity analysis (e.g. female mortality has a greater affect on system dynamics than male mortality). Hence, Information Theory Sensitivity Analysis (ITSA) approach enabled a more in-depth analysis of model parsimony while still yielding the traditional benefits of sensitivity analysis.

Introduction

Computer models are used in epidemiology to understand the dynamics of disease transmission and spread, and to assess efficiency of alternate control strategies (Coyne et al. 1989, Barlow et al. 1996, Smith and Cheeseman 2002, Sterner and Smith 2006). Modelling disease-host systems necessitates making assumptions about underlying mechanisms so they can be simplified for representation. Advantages of simple models include having fewer processes to parameterise, validate, and contribute to outcome uncertainty. Simple models can effectively capture general trends of disease-host systems (Anderson and May 1991), e.g. simulating population dynamics of average individuals in a homogenous environment of contact rates and disease susceptibility. More complex models are required to explore heterogeneous rates of disease spread over spatially varying populations and landscapes. Detail is added by modelling the population as individuals or structuring the population into classes (e.g. age, sex), contact networks, or patches in which each individual, class, network node, or patch are defined to reflect the heterogeneous nature of the population and disease (Ferguson et al. 2003). Increasing model realism in these ways requires more parameters. An issue that inevitably arises is the degree of complexity required to fulfil modelling objectives. Application of the principle of parsimony (in the sense of Burnham and Anderson 2002, pp.31-35) is advocated to choose among models of varying complexity. Adhering to this principle requires balancing the benefits of few parameters, which will minimise outcome uncertainty arising from parameter estimation error, against the benefit of having a sufficient number of parameters to adequately represent the modelled system (Snowling and Kramer 2001, Burnham and Anderson 2002).

Sensitivity analysis can assess whether model dynamics appropriately represent system behaviours. A variety of sensitivity analysis techniques are used to develop disease simulation models (Voigt et al. 1985, Blower and Dowlatabadi 1994, Sanchez and Blower 1997). Hamby (1994) provides a thorough description of numerous sensitivity analysis techniques. A general framework i) identifies parameters for testing, ii) defines a probability distribution function (pdf) of values for each parameter, iii) applies a sampling strategy to determine input parameter value and iv) analyses the effect of the parameters on one or more response variables characterising model outcomes. The final step can be used to identify parameters needed to model system behaviours. If a model process is defined by more than one parameter, and independent variation of their values does not have an effect on model outcomes, then the process can be modelled more succinctly. In this instance, a more parsimonious model can be defined by combining the parameters of the modelled process.

One concern for using current sensitivity analysis techniques to evaluate model complexity is that null hypothesis significance testing (NHST) is used to determine whether parameters have an effect on model outcomes. NHST is a poor approach for model selection for several reasons. NHST uses *P*-values to define the probability of obtaining a result, or one more extreme, given the null hypothesis is true, and sampling is random. One problem with *P*-values is that cases identified as being significant may be beyond the true value in order for the case to be sufficiently different from the sample mean to reject the null hypothesis (Whittingham et al. 2006). *P*-values are also criticised because conclusions concerning significance are not based entirely on the observed data when a predefined distribution of the test statistic is used to derive *P*-values (Johnson

1999, Anderson et al. 2000). Predefined distributions are not reflective of the true data distribution. Fortunately, randomisation tests that derive a test statistic's distribution under a null hypothesis from numerous random sampling of the data can be used to overcome this concern.

Other shortcomings of NHST arise from its relationship with sample size. As sample size increases the probability of making an incorrect interpretation should decrease. However, in hypothesis testing, the probability of committing a Type I error (rejecting the null when it is true) remains the same regardless of the sample size. Conversely, the probability of committing a Type II error (accepting the null when it is false) decreases with increasing sample size. As sample size increases, the variance of the sample group decreases, and the sample mean becomes closer to the true mean. However, as the power to detect a difference increases, a significant difference may be detected when it is not biologically meaningful, because the sample mean is so close to the population mean (Johnson 1999, Anderson et al. 2000, Robinson and Wainer 2002).

NHST is also criticised for the arbitrary choice of significance levels (α). A level of significance allows for a degree of deviation of data from the null hypothesis; however, this level is arbitrarily chosen, and can lead to differences in interpretation. For example, with a *P*-value of 0.03 one fails to reject the null hypothesis at a significance level $\alpha = 0.01$, but the null would be rejected at a significance level $\alpha = 0.05$. The classification of results into significant or non-significant categories is uninformative when no estimate of effect size or its certainty is given (Anderson et al. 2000). Though, Robinson and Wainer (2002) state that even if this information is given it provides little context for choosing among models.

Two common sensitivity analysis procedures from which conclusions are based on *P*-values are analysis of variance (ANOVA) and regression. ANOVA can be used to test for significant main and interaction effects of the parameters on the model outcomes. This technique has the advantages of being a simple and well-known statistical procedure that can easily assess the effect of parameter interactions (Ginot et al. 2006). The main disadvantage is that input values must be discrete in order to conform to "levels" of ANOVA factors. Furthermore, it is too computationally expensive to test more than a small subset of parameters that have a limited number of levels. Regression analysis is a more comprehensive model selection technique (Hamby 1994). It is feasible to analyse models with a high number of parameters and test for a limited number of interactions, if desired. The coefficient of determination, R^2, quantifies model predictive power by defining the amount of variation in a dependent variable explained by the linear combination of independent variables. Similarly, residual sum of squares (RSS) in an ANOVA quantifies the deviance of the predicted model from the data. However, if candidate models have similar R^2 or RSS values but differ in their number of parameters, there is no accounting for the cost of increasing model complexity on outcome uncertainty. Thus, the second reason current sensitivity analysis techniques are less suitable for model selection is that they fail to evaluate model parsimony. Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) provide a means of selecting the most appropriate model. Furthermore, these information criteria avoid the problems of null hypothesis testing described above because they do not use *P*-values. In this study, Information Theory Sensitivity Analysis (ITSA) technique is presented as a means of overcoming these problems by using AIC and BIC.

AIC and BIC can choose among models by measuring whether the benefit of allowing a parameter to vary and potentially improve model prediction outweighs the cost of increasing uncertainty in model outcome through parameter estimation error. Given outcome uncertainty inherent with any analysis, using two different types of information criteria enables a more comprehensive assessment of model parsimony. AIC unites theory from Kullback-Leibler information and maximum likelihood (Anderson et al. 2000). Kullback-Leibler information measures the difference between truth and a modelled approximation. Since the truth about a phenomenon is often unobtainable, Akaike's information theory derives an estimate of Kullback-Leibler truth from the maximized log-likelihood function. The estimated maximized log-likelihood function forms the backbone of AIC,

$$AIC = -2log_e(l(\theta|data)) + 2K \qquad \text{Equation 1}$$

where $log_e(l(\theta|data))$ is maximized log-likelihood over the unknown parameters (θ), and K is the number of estimated parameters in the model.

BIC is derived from Bayesian statistics. For Bayesian analysis, the probability of a model being the "top model", given the data and other candidate models, is referred to as the posterior probability. A prior probability is used to derive the posterior probability by "*a priori*" defining the likelihood of the candidate model being the top model. A central quantity to Bayesian approach of model comparison is the Bayes factor. The Bayes factor is the measure of evidence in favour of one model over another (Kuha 2004). BIC is an approximation of the Bayes factor (Kuha 2004). Model specific prior

probabilities are not specified in the calculation of BIC because they are assumed to be equal for each candidate model (Burnham and Anderson 2002):

$$BIC = -2\log_e(l(\theta|data)) + K \bullet \log_e(n) \qquad \text{Equation 2}$$

The main difference between AIC and BIC is that for BIC the true model is assumed to be within the list of candidate models. The goal is to identify the true model, and as sample size increases, the probability of determining the true model approaches 1, though there is no requirement that the true model must be within the list of candidate models (Burnham and Anderson 2002, Reineking and Schroder 2006). Whereas for AIC, the aim is to identify the model that gives the best approximation of truth. For both approaches, the best model, given the data and set of models, has the smallest information criterion value. The trade-off between fitting the data and outcome uncertainty is inherent in equations 1 and 2. The more parameters used, the better the model approximation and the lower the value of $-2\log_e(l(\theta|data))$ and the resultant AIC or BIC. However, as the number of parameters increase, so too does the uncertainty of the model in approximating the phenomenon. To penalize this effect, AIC and BIC values become larger through the $2K$ and $K \bullet \log_e(n)$ terms, respectively. The BIC formulation more heavily penalises models with more parameters at larger sample sizes than AIC (Reineking and Schroder 2006).

AIC and BIC analyses avoid the problems of null hypothesis because they do not use *P*-values, null hypotheses, predefined probability distributions of test statistics, or significance values for deriving conclusions. Confidence in the results increases with

increasing sample size. Furthermore, the results are more informative because AIC and BIC values can be used as the strength of evidence in favour of a model being the most parsimonious (Burnham and Anderson 2002). Unlike NHST, which yields the probability of a result given the null hypothesis being true from following an assumed distribution, AIC and BIC give the probability of a result derived entirely from the data.

ITSA was applied to the Ontario Rabies Model (ORM) to assess its complexity. Parameter richness is a consequence of being an individual-based spatially explicit disease-host simulation model. ORM simulates raccoon (*Procyon lotor*) population dynamics, raccoon rabies viral transmission, and rabies control strategies. ITSA is used to test model behaviour for 17 demographic and 3 disease parameters that characterise raccoon rabies dynamics as estimated from field and laboratory studies. A description of the model is reported elsewhere (Tinline et al. 2007). Parameter impacts are evaluated in terms of parsimony and determining what additional fieldwork might be required to define parameters most representative of system behaviours.

Methods

Ontario Rabies Model

The ORM operates on a lattice of hexagonal cells, each with a radius of 2 km, the approximate activity range of raccoons in Ontario (Rosatte 2000). The resulting cell area is 10.39 km^2. The shape and size of the model landscape is user-definable. Experiments used a 15x20 lattice of cells simulating an area of 3000 km^2, approximately equal to the area infected by the 1999 raccoon rabies outbreak in eastern Ontario (Rosatte et al. 2001, 2006). Model processes operate at one-week intervals and are stochastically determined. Parameters for several demographic behaviours and one disease behaviour (incubation

period; Tinline et al. 2002) are determined from probability distributions, while the remaining behaviours are defined using a single value (Table 4.1). Each raccoon is in one of the following health states: i) healthy and susceptible to rabies, ii) infected and incubating rabies, or iii) infectious, i.e. capable of infecting susceptible raccoons. Raccoons die after a 1 week infectious period (Winkler and Jenkins 1991). Each year, juvenile (52 – 74 weeks old) and adult females (≥ 75 weeks old) have a chance of producing offspring as determined by a probability distribution. Mortality rates are defined for year class, and can be sex specific if desired. Model experiments typically use an initial raccoon population, which has been grown to a stable population in the whole study area starting with a pair of breeding raccoons. Each cell in the model landscape is assigned a target carrying capacity (k), a value measuring the population density at a specific time period (currently set by default as week 30 in the ORM). In the ORM, k is used for a given cell as a target population about which the model adjusts mortalities dynamically to ensure that the population oscillates around the target population. Mortality adjustments are made weekly for each animal in each cell according to equation 1 below:

$$\text{Probability of dying} = AGM\ [(N/k) + A] \qquad \text{Equation 3}$$

where the AGM is the user defined age/gender determined probability of dying on a given (weekly) interval, N is the current total population of the cell, k is the user assigned carrying capacity and A is a user defined adjustment factor between 0 and 1. For example, if the AGM was 0.006 with N=30, K=70 and A=0.002 then the adjusted

probability of dying is calculated as: 0.006[(30/70)+0.002] or 0.00457. As the population builds up, N approaches k. Thus the adjusted mortality will exceed the pre-defined AGM effectively damping the increase in population. The adjustment factor A in Equation 3, also known as the mortality adjuster parameter (P8), acts as an intercept of a linear equation which ensures that there is an ambient level of mortality insensitive to changes in the cell population relative to k. P8 controls the amount of increase or decrease in the mean mortality rates. Its default value is derived through a model fitting process minimising the deviation between observed and modelled raccoon demographics.

Every raccoon has a defined probability of contact (P20) with other raccoons. They can interact with raccoons within their home cell and up to six neighbouring cells, depending on the cellular landscape configuration. The amount of interaction depends on the size of the cells and the home range. Dispersal occurs once a year. Raccoons may move one or more cells, as determined from a pdf of dispersal distances (P1, P2, P3, P4). The annual timing of dispersal is defined by the permissible movement period parameters (P14, P15, P16). Rabies can be introduced into any cell in the landscape, for any given week of the simulation, to an assigned percentage of animals within the infected cell(s). The reflective disease spread parameter (P22) is optionally used to increase chance of infection to adjacent cells when there are fewer than six contiguous neighbouring cells (e.g. along the boundary of the landscape) by increasing the contact rate (P20) with available adjacent cells.

Experimental Design and Model Parameterisation

Seventeen demographic (P1 – P17, P20) and 3 disease parameters (P18, P19, P21) were selected for sensitivity analysis (Table 4.1). It is possible to test additional parameters,

such as the effect of different winter severities or various rabies control strategies on raccoon rabies dynamics; however, this study focuses on assessing fundamental parameters considered to simulate raccoon rabies dynamics. Three additional parameters were included (P22, P23, P25), that were not expected to affect model behaviours, to test the ability of the methodology to identify superfluous parameters. P25 was included to set a quantitative reference for the lack of effect of a truly superfluous parameter on model outcomes because its values were derived randomly and not used as model input. A final parameter, "rabies run" (P24), also provided no input for the model, but existed for the sensitivity analysis to define whether a simulation was infected or not with rabies.

Table 4.1 Fundamental parameters in the Ontario Rabies Model assessed by ITSA. *Parameters whose values are defined by a probability distribution function; all other parameter use single input values. $^{+}$Non-ORM input parameters.

Demographic parameters	Male adult/juvenile dispersal distance; P1 *
	Female adult/juvenile dispersal distance; P2 *
	Male young of year dispersal distance; P3 *
	Female young of year dispersal distance; P4 *
	Average litter size; P5
	Male mean percent mortality; P6 *
	Female mean percent mortality; P7 *
	Mortality Adjuster; P8
	Age of independence from mothers; P9
	Chance of juveniles giving birth; P10
	Chance of adults giving birth; P11
	Birth week; P12
	Male juvenile/adult permissible movement period; P14
	Female juvenile/adult permissible movement period; P15
	Young of year permissible movement period; P16
	Target cell population density; P17
Disease parameters	Disease transmission rate; P18
	Incubation period; P19 *
	Contact rate; P20
	Time of infection; P21
Superfluous parameters	Reflective disease spread; P22
	Sex specific mortality rates; P23
	Random variable; P25 $^{+}$
Flagging parameter	Rabies run; P24 $^{+}$

Biologically appropriate data for defining ORM parameter inputs were queried from the Raccoon Ecology Database (REDB; Chapter 2). The REDB includes estimates of population density, survival rates, rabies incubation period, litter size, body weight, dispersal distance and home range size, often classified by age or sex class for 100's of peer-reviewed and unpublished data regarding raccoon biology, ecology and raccoon rabies. The data were used to define pdf's for each parameter using Palisade Corporation's @RISK software, version 4.5.4 (www.palisade.com). Latin hypercube sampling (LHS) from the pdf's specified parameter inputs for model simulations. LHS is a substantially more efficient means of defining input values that cover the entire parameter space for a highly parameterised model than a fully factorial design (McKay et al. 1979). LHS draws parameter values without replacement and with equal probability over the entire parameter space. McKay et al (1979) recommend that the number of samples (n) drawn should be $>4K/3$, where K is the number of parameters. For analysis, $K = 24$, hence, n is set to 50 to satisfy $n > 4K/3$, and to sufficiently sample parameter space. This resulted in 50 different sets of input parameter specifications, of which each are referred to as a model simulation.

More specifically, there were two strategies for defining parameter inputs depending on whether the parameter was defined by a single value or by a pdf within the model. In the first case, data from reviewed literature are used to create a pdf from which a value is drawn during LHS and used as input in the model (e.g. contact rate, P20). For parameters defined within the model using a pdf, reviewed data enable definition of three different pdf's in terms of shape (e.g. uniform, leptokurtic, platykurtic for incubation period, P19) or its mean (e.g. low, medium or high mean pdf values for dispersal

distances, P1, P2, P3, P4, mortality rates, P6, P7). LHS sampling from a uniform distribution is then used to determine with equal likelihood whether a low, medium or high pdf is used as input for each of the 50 simulations. Similarly, LHS sampling of only two values from a uniform distribution is used to determine whether a simulation is infected with rabies (P24), uses reflective disease spread (P22), or has sex specific mortality rates (P23), where either case of each parameter can occur with equal probability.

Simulation Runs

A raccoon population was grown from a single pair of raccoons into a stable population over 100 years. Rabies was not introduced into this population. This population is referred to as the *seed population*, because it is the starting population from which sensitivity analysis simulations are run. The model was run for another 200 years, for a total of 300 years, for each of the 50 uniquely parameterised model simulations. Thus, all simulations start with the same conditions, but then model dynamics are expected to change once the model runs using the unique parameter specification. Simulations with rabies were infected at year 200, at a week defined by P21. Infecting "rabid" simulations at year 200 ensured sufficient time had passed for parameter specifications to effect model dynamics. Running the model for a further 100 years allowed sufficient time for rabies to have an effect on model dynamics. Only half of the simulations were infected with rabies to enable adequate assessment of model dynamics with and without disease. Rabies was seeded in five cells of the leftmost column of the 20x15 hexagonal cell lattice. These cells were evenly spaced along the column. Furthermore, each model simulation was run 10 times to capture the outcome variation

caused by model stochasticity, to yield a sample size of 10 for each of the 50 unique model simulations.

Response Variables

Effects of parameter inputs on modelled demographic and disease characteristics were quantified temporally and spatially using four demographic and five disease response variables (Table 4.2). Response variables were chosen to examine overall characteristics of model output; however, as further discussed, the choice of response variables was expected to influence results because parameters likely differ in the scale of their spatial and temporal scale effects. Individual parameter impacts on selected response variables are illustrated using scatterplots: density versus TotalPop, and contact rate versus InfectY1.

Correlations among response variables were expected and they could potentially confound the results. Hence, principal components analysis (PCA) of standardised response variables was performed for i) all simulations, ii) infected simulations and iii) non-infected simulations using STATISTICA 7.0 (*www.statsoft.com*). Component scores were calculated from eigenvectors for the most explanatory components (Manly 1994) and used for subsequent sensitivity analysis.

Table 4.2 Response variables used in the sensitivity analysis.

Demographic Response Variables	**Description; (Value for sensitivity analysis)**
Total population (TotalPop)	Total number of raccoons in week 30 averaged over the final 20 years of a simulation trial; (size of population)
Temporal variance:mean ratio (TempVM)	The variance of landscape populations over the mean landscape population of the final 20 years of simulation for week 30; (measure of temporal dynamics of population at a coarse spatial scale)
Centre cell temporal variance:mean ratio (CentreVM)	The variance of the centre cell population over the mean centre cell population of the final 20 years of simulation for week 30; (measure of temporal dynamics of population at a fine spatial scale)
Spatial variance:mean ratio (SpatVM)	The variance of cell populations over the mean cell populations of the entire landscape for week 30 of the final year; (spatial variability of population dynamics)
Disease Response Variables	**Description; (Value for sensitivity analysis)**
Infected spatial variance:mean ratio (InfectSpatVM)	The variance of cell populations of infected animals over the mean cell populations of infected animals for the entire landscape after the first year of simulation; (spatial measure of diseased population dynamics)
Total number of rabies cases (InfectY1)	Total number of rabies cases one year after the initial infection; (measures the severity of the initial outbreak)
Maximum number of rabies cases (MaxInfect)	The maximum number of rabies cases that occurred during any week of the simulation; (an overall indication of outbreak severity)
Disease duration (Duration)	Number of weeks that incubating or infectious raccoons exist in the simulation; (an overall indication of the severity and temporal intensity of the outbreak)
Rate of disease spread (TimeToCross)	Number of weeks it takes for the first infectious raccoon to cross half the length of the study area (column 10 of 20 columns); (measures the severity of the initial outbreak)

PCA of i) all simulations, ii) infected simulations and iii) non-infected simulations differed by the type of response variables used and their sample size. InfectSpatVM and TimeToCross response variables were not included for PCA of all simulations. This was to ensure a full sample size of 500, because it was not possible to calculate values for InfectSpatVM and TimeToCross from non-infected simulations. PCA for infected simulations used all response variables, but had a reduced sample size of 150. Again, this was because of InfectSpatVM and TimeToCross, which only have 150 of the 250 possible infected observations owing to cases where the disease burned out before InfectSpatVM and TimeToCross were calculated. Only demographic response variables were used for PCA of non-infected simulations, because there were no data for the disease response variables, and consequently had a reduced sample size of 250.

Information Theory Sensitivity Analysis (ITSA)

Component scores generated from the PCA of the response variables (Table 5.2) were used as the response variables for ITSA. ITSA was performed nine times because three principal components were calculated from each of i) all simulations, ii) infected simulations and iii) non-infected simulations. The essential steps for ITSA were to a) run linear regression for each parameter to calculate univariate model values of AIC corrected (AIC_c) for small sample sizes and BIC, and then rank the models using these values, b) run multiple regression analyses for multivariate models designed *a priori* to calculate their values of AIC_c and BIC weights (known as posterior probabilities of models for BIC; Burham and Anderson 2002), and then rank the models using these values. R^2 values were also calculated for steps (a) and (b) to compare with the information criteria.

Scatterplots were used to assess linearity between continuous parameters and PCA derived response variables. Non-linear parameters were transformed into categorical parameters and treated as factors during regression analysis. Transformed parameters were renamed using a prefix of "c" (e.g. P18 to cP18).

Disease parameters (P18, P19, P21, P22) were not used in the ITSA of iii) non-infected simulations because they were not parameterised for these simulations.

AIC_c and BIC values for the univariate and multivariate models were calculated using RSS (Burnham et al. 2000) and R^2 values (Raftery et al. 1997) from least squares regression, respectively:

$$AIC_c = nlog_e(RSS / n) + 2K + C \qquad \text{Equation 4}$$

$$BIC = nlog_e(1-R^2) + (K-2) \bullet log_e(n) \qquad \text{Equation 5}$$

where n is the number of simulations run for each sensitivity testing group. K is the number of estimated regression parameters in the model including the model constant (the y-intercept) and the residual term accounting for variation in the data not explained by the parameters. C is calculated as $2K(K+1) / (n - K - 1)$ to correct for small sample sizes (Burnham and Anderson 2002).

AIC_c and BIC values are unitless, but the values are meaningful on a relative scale when considered together. Small values indicate variation observed in a response variable more succinctly than models with larger values. AIC_c and BIC weights, w_i, were

calculated to define the likelihood of a model being the best model given the data and set of candidate models:

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta_r)} \qquad \text{Equation 6}$$

where Δ_i are the AIC_c or BIC differences, calculated as the difference between the IC value for model i and the lowest IC value: $\Delta_i = AIC_{c,i} - AIC_{c,min}$, or $\Delta_i = BIC_i - BIC_{min}$.

Candidate Multivariate Models

Several types of multivariate models were created *a priori* to explore model parsimony: I) a global model that included all parameters; II) three *semi*-global models, in that, each one lacked one of the superfluous parameters (P22, P23, or P25), to evaluate the effect of the superfluous parameter within a multivariate context; III) a forward stepwise regression model, seeded with the parameter that explained the most variation from the univariate analysis; and IV) a backward stepwise regression model.

Model Stochasticity

ITSA was repeated using the mean of the stochastic results. This reduced sample size by a factor of 10 (because each unique model simulation was run 10 times to capture stochastic variation). The motivation for averaging results was to decrease noise in modelled outcomes, produced by stochasticity, in case this impeded identification of critical parameters.

Statistical analysis was accomplished using R (R Development Core Team 2005). Models of type I - IV were run using the linear models facility. For the stepwise analysis, AIC was used as the criterion for inclusion or exclusion of parameters.

Results

Five continuous parameters were non-linearly related to the response variables, thus, were transformed into categorical parameters: cP5, cP9, cP17, cP18 and cP20. Scatterplots indicate that raccoon populations are larger with higher density values (cP17) and the number of infectious cases in the first year of rabies infection is also greater when contact rates are higher (cP20; Figure 4.1, 4.2).

PCA results of the response variables indicated that the majority of variation was captured in the top two or three components (Table 4.3). The eigenvectors of each component were used to formulate a verbal description (e.g. component 1 for all simulations and infected simulations is most representative of variation caused by "disease intensity"; Table 4.3).

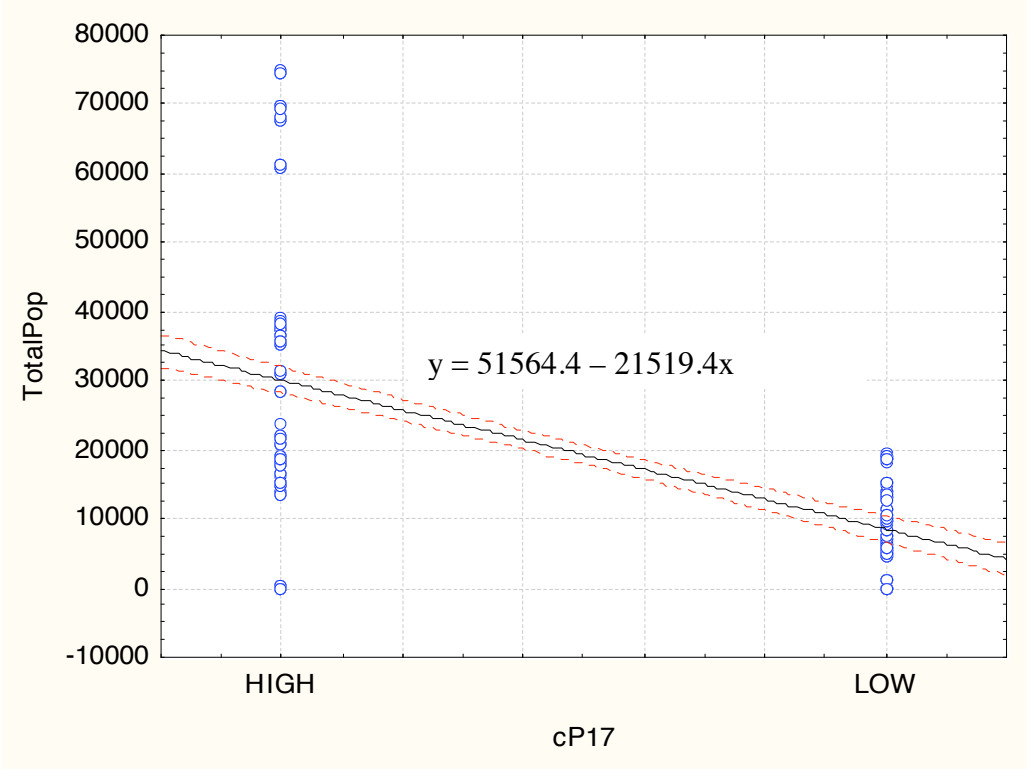


Figure 4.1 Scatterplot illustrating impact of density (cP17) on raccoon population size (response variable: TotalPop). 95% confidence intervals are given.

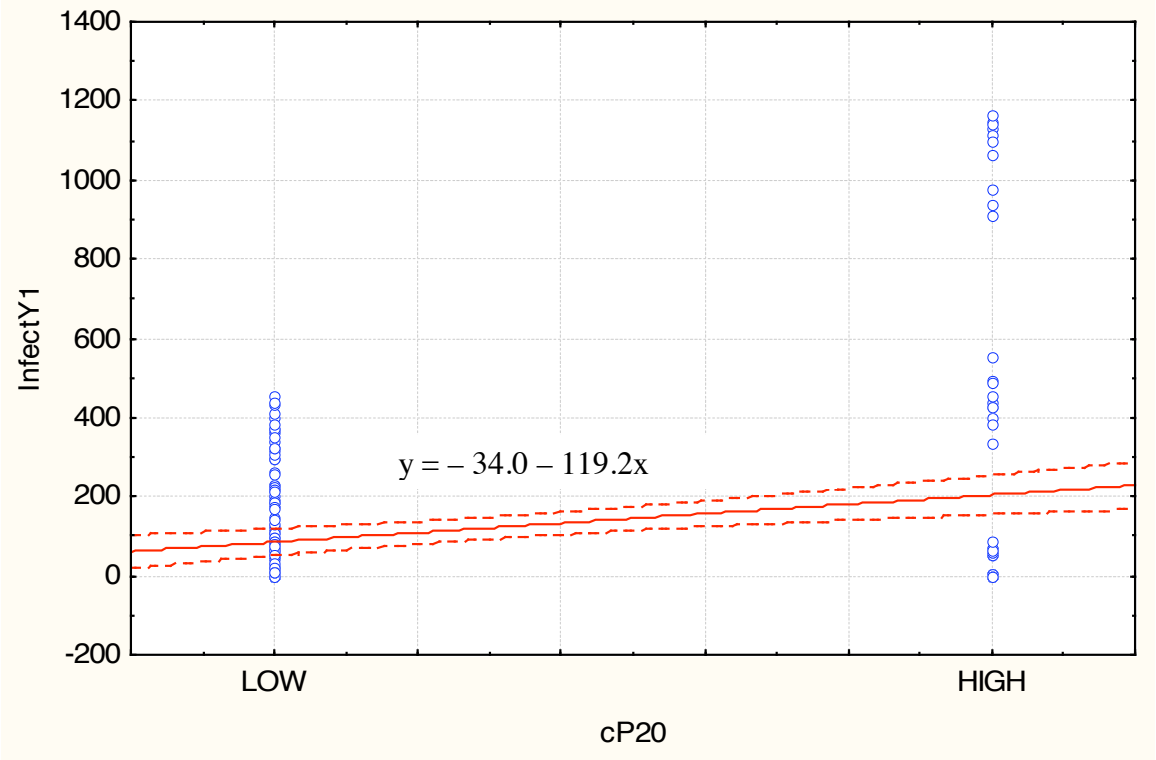


Figure 4.2 Scatterplot illustrating impact of contact rate (cP20) on total number of rabies cases during the first year of infection (response variable: InfectY1). 95% confidence intervals are given.

Table 4.3 Eigenvectors and the cumulative and individual percent variation explained from the first three PCA components (Z1, Z2, Z3) for i) all simulations, ii) infected simulations and iii) non-infected simulations. A verbal description of the factors responsible for the variation in the components is also given.

Response Variable	i) Components for all simulations:			ii) Components for infected simulations			iii) Components for non-infected simulations		
	Z1	**Z2**	**Z3**	**Z1**	**Z2**	**Z3**	**Z1**	**Z2**	**Z3**
TotalPop	0.34	-0.48	0.23	0.28	-0.58	0.25	0.28	0.86	-0.43
TempVM	0.06	0.47	0.38	-0.01	0.13	0.86	-0.60	0.13	-0.21
CentreVM	-0.21	-0.18	0.61	-0.25	-0.55	0.18	-0.49	0.48	0.68
SpatVM	0.12	0.68	-0.18	0.04	0.53	0.31	-0.57	-0.13	-0.56
InfectSpatVM	-	-	-	0.41	0.21	-0.09	-	-	-
InfectY1	0.64	0.02	0.01	0.45	-0.06	0.12	-	-	-
MaxInfect	0.64	0.00	0.04	0.51	-0.06	0.06	-	-	-
Duration	0.02	0.26	0.63	-0.25	-0.07	0.24	-	-	-
TimeToCross	-	-	-	-0.40	0.10	0.05	-	-	-
Cumulative Percentage (and individual percentage)	30.7 (30.7)	48.1 (17.4)	63.6 (15.5)	39.6 (39.6)	55.5 (15.9)	67.3 (11.8)	45.2 (45.2)	70.3 (25.1)	86.5 (16.2)
Component Descriptor	Disease intensity	Spatial and temporal population patterns	Fine scale temporal population patterns	Disease intensity	Spatial and temporal population patterns	Broad scale temporal population patterns	Spatial and temporal population patterns	Population size	Fine scale temporal population patterns

Univariate analyses of parameter effects on the three component scores for i) all simulations, ii) infected simulations and iii) non-infected simulations produced identical parameter rankings from the R^2, AIC_c and BIC values (Table 4.4, 4.5, 4.6). Analytical results were almost identical using data from stochastic simulation runs and the mean of stochastic runs. The exceptions were that the "averaged" results tended to have a higher ranking for the random parameter (P25), and that fewer parameters had significant R^2 values.

Highly ranked parameters for i) all simulations, ii) infected simulations and iii) non-infected simulations were: P4, P7, cP9, P14, cP17, cP18, P24. The disease parameters, cP18, P19 and P21 were ranked most highly for the ii) infected simulations. The random (P25) and superfluous variables (P22, P23) had low rankings for i) all simulations, ii) infected simulations and iii) non-infected simulations. Other parameters with consistently low rankings were cP5, P8, P10, P11, and P12.

For multivariate analyses, the backwards then forwards stepwise regression models were identified as the top two candidate models using AIC_c and BIC weights for i) all simulations. There was no clear trend in the ranks of global and semi-global models. For ii) infected and iii) non-infected simulations, global and semi-global models could not be built because sample size was too small to estimate all parameter coefficients (Table 4.7, 4.8, 4.9).

Table 4.4 Univariate analysis of the PCA components a) Z1, b) Z2 and c) Z3 produced from "all simulations". Univariate models are ranked by AIC_c. *R^2 values with *P*-values < 0.05. Disease parameters are in bold, and the superfluous and random variables are highlighted in grey.

a) Z1							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
cP18	-1001.5	-218.6	0.36*	**cP18**	170.1	-211.3	0.35*
cP17	-878.0	-95.1	0.18*	cP17	289.2	-92.2	0.18*
cP9	-861.4	-78.4	0.16*	cP9	305.3	-76.0	0.15*
P7	-828.1	-45.2	0.10	P7	330.5	-50.9	0.11*
P24	-826.2	-43.2	0.09	P24	339.5	-41.9	0.09*
P14	-808.8	-25.8	0.06	P14	356.4	-25.0	0.06*
P2	-808.4	-25.4	0.06	P2	356.8	-24.6	0.06*
P6	-806.3	-23.4	0.06	P6	358.8	-22.6	0.06*
P3	-804.1	-21.1	0.05	P3	361.0	-20.4	0.05*
cP20	-799.6	-16.7	0.04	cP20	365.3	-16.1	0.04*
P1	-789.2	-6.2	0.02	P1	375.5	-5.9	0.02*
P16	-787.4	-4.5	0.02	P16	377.2	-4.2	0.02*
P15	-786.4	-3.4	0.02	P15	378.2	-3.2	0.02*
P22	-783.5	-0.6	0.01	P22	381.0	-0.4	0.01*
P11	-783.4	-0.5	0.01	P11	381.1	-0.3	0.01
P21	-780.6	2.3	0.01	**P21**	383.8	2.4	0.01
P10	-779.8	3.1	0.01	P10	384.6	3.2	0.01
P19	-779.3	3.7	0.01	**P19**	385.1	3.7	0.00
P12	-779.0	3.9	0.00	P12	385.4	4.0	0.00
P4	-778.3	4.7	0.00	P4	386.1	4.7	0.00
P8	-777.8	5.2	0.00	P8	386.6	5.2	0.00
P23	-777.3	5.7	0.00	P23	387.1	5.7	0.00
P25	-777.0	6.0	0.00	cP5	387.5	6.1	0.00
cP5	-776.8	6.1	0.00	P25	387.6	6.2	0.00

Table 4.4. con't

b) Z2							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P7	-1447.2	-78.5	0.16*	P7	61.4	-36.6	0.08*
P24	-1442.2	-73.5	0.15*	P24	63.4	-34.6	0.08*
cP9	-1408.7	-39.9	0.09*	cP9	80.2	-17.8	0.05*
cP17	-1404.7	-35.9	0.08*	cP17	82.2	-15.8	0.04*
cP18	-1401.6	-32.9	0.08	**cP18**	83.8	-14.2	0.04*
P4	-1398.8	-30.0	0.07	P4	85.3	-12.7	0.04*
P6	-1389.9	-21.2	0.05	P6	89.8	-8.2	0.03*
P15	-1382.6	-13.8	0.04	P15	93.6	-4.3	0.02*
P2	-1379.8	-11.0	0.03	P2	95.1	-2.9	0.02*
P16	-1377.1	-8.4	0.03	P16	96.5	-1.5	0.02*
P1	-1376.6	-7.8	0.03	P1	96.8	-1.2	0.01*
P19	-1375.9	-7.2	0.03	**P19**	97.1	-0.9	0.01*
P3	-1373.0	-4.2	0.02	P3	98.7	0.7	0.01*
cP20	-1372.7	-3.9	0.02	cP20	98.8	0.9	0.01
P23	-1370.3	-1.6	0.02	P23	100.1	2.1	0.01
cP5	-1370.3	-1.5	0.02	P5	100.1	2.1	0.01
P12	-1365.3	3.5	0.01	P12	102.8	4.8	0.00
P14	-1364.6	4.1	0.00	P14	103.1	5.1	0.00
P22	-1364.2	4.6	0.00	P22	103.3	5.4	0.00
P11	-1363.6	5.1	0.00	P11	103.6	5.6	0.00
P21	-1363.5	5.2	0.00	**P21**	103.7	5.7	0.00
P10	-1362.6	6.1	0.00	P25	104.1	6.1	0.00
P8	-1362.6	6.1	0.00	P10	104.2	6.2	0.00
P25	-1362.5	6.2	0.00	P8	104.2	6.2	0.00

c) Z3							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P24	-1657.8	-79.5	0.16*	P24	-52.2	-30.1	0.07*
P14	-1627.8	-49.5	0.11	P14	-39.9	-17.8	0.05*
P16	-1618.8	-40.5	0.09	P16	-36.1	-14.0	0.04*
P15	-1601.9	-23.6	0.06	P15	-28.9	-6.8	0.03*
P3	-1593.7	-15.4	0.04	P3	-25.4	-3.3	0.02*
P22	-1593.0	-14.7	0.04	P22	-25.0	-3.0	0.02*
cP17	-1591.5	-13.2	0.04	cP17	-24.4	-2.3	0.02*
P7	-1585.7	-7.5	0.03	P7	-21.8	0.3	0.01*
P23	-1584.6	-6.4	0.02	P23	-21.4	0.7	0.01*
P19	-1581.0	-2.7	0.02	P19	-19.8	2.3	0.01
P6	-1578.9	-0.7	0.01	P6	-18.9	3.2	0.01
cP18	-1578.6	-0.3	0.01	**cP18**	-18.7	3.3	0.01
P2	-1576.8	1.5	0.01	P2	-17.9	4.1	0.00
P10	-1575.8	2.5	0.01	P10	-17.5	4.5	0.00
P1	-1575.5	2.7	0.01	P1	-17.4	4.7	0.00
P4	-1574.2	4.1	0.00	P4	-16.8	5.3	0.00
P12	-1574.1	4.2	0.00	P12	-16.7	5.3	0.00
P21	-1573.8	4.5	0.00	P21	-16.6	5.5	0.00
cP9	-1573.4	4.9	0.00	cP9	-16.5	5.6	0.00
P8	-1572.8	5.4	0.00	P8	-16.2	5.9	0.00
P11	-1572.6	5.7	0.00	P11	-16.1	6.0	0.00
cP20	-1572.4	5.9	0.00	P25	-16.1	6.0	0.00
cP5	-1572.1	6.2	0.00	cP20	-16.0	6.1	0.00
P25	-1572.1	6.2	0.00	cP5	-15.9	6.2	0.00

Table 4.5 Univariate analysis of the PCA components a) Z1, b) Z2 and c) Z3 produced from "infected simulations". Univariate models are ranked by AIC_c. *R^2 values with *P*-

values < 0.05. Disease parameters are in bold, and the superfluous and random variables are highlighted in grey.

a) Z1							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
cP18	-206.9	-112.9	0.54*	**cP18**	149.5	-107.1	0.53*
P7	-151.7	-57.8	0.34*	P7	201.4	-55.2	0.33*
cP17	-141.7	-47.8	0.30*	cP17	210.9	-45.7	0.29*
cP9	-140.6	-46.7	0.29*	cP9	212.0	-44.6	0.28*
P14	-138.8	-44.9	0.28	P14	213.7	-42.9	0.27*
P22	-130.7	-36.8	0.24	P22	221.3	-35.2	0.24*
P6	-126.2	-32.4	0.22	P6	225.6	-30.9	0.21*
P1	-119.9	-26.0	0.19	P1	231.7	-24.8	0.18*
P2	-119.7	-25.7	0.19	P2	232.0	-24.6	0.18*
cP20	-115.3	-21.4	0.16	cP20	236.1	-20.4	0.16*
P16	-113.5	-19.6	0.15	P16	237.8	-18.7	0.15*
P15	-99.0	-5.1	0.07	P15	251.8	-4.7	0.06*
P3	-98.1	-4.2	0.06	P3	252.7	-3.8	0.06*
P23	-96.4	-2.5	0.05	P23	254.3	-2.3	0.05*
P4	-95.7	-1.8	0.04	P4	255.0	-1.6	0.04*
P19	-94.1	-0.2	0.03	**P19**	256.5	0.0	0.03
P8	-93.1	0.8	0.03	P8	257.5	0.9	0.03
P10	-92.6	1.3	0.02	P10	258.0	1.4	0.02
P25	-91.0	2.9	0.01	P12	261.0	4.4	0.00
P12	-89.5	4.4	0.00	P11	261.1	4.5	0.00
P11	-89.4	4.5	0.00	cP5	261.3	4.7	0.00
cP5	-89.2	4.7	0.00	P25	261.4	4.9	0.00
P21	-88.9	5.0	0.00	**P21**	261.6	5.0	0.00

Table 4.5. con't

b) Z2							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P1	-386.2	-114.8	0.55*	P1	63.2	-66.5	0.38*
cP18	-354.4	-83.0	0.44*	**cP18**	79.9	-49.8	0.31*
P16	-311.6	-40.2	0.26	P16	105.0	-24.7	0.18*
P15	-301.1	-29.6	0.21	P15	111.7	-18.0	0.14*
P4	-300.9	-29.5	0.21	P4	111.8	-17.9	0.14*
cP5	-297.0	-25.6	0.18	cP5	114.3	-15.4	0.13*
P7	-293.0	-21.6	0.16	P7	116.9	-12.8	0.11*
P2	-280.8	-9.4	0.09	P2	124.9	-4.8	0.06*
P19	-276.8	-5.3	0.07	**P19**	127.7	-2.0	0.05*
cP9	-276.3	-4.8	0.06	cP9	128.0	-1.7	0.04*
P11	-275.1	-3.6	0.06	P11	128.8	-0.9	0.04*
P6	-273.6	-2.1	0.05	P6	129.8	0.1	0.03*
P23	-271.8	-0.4	0.04	P23	131.0	1.3	0.02
P3	-270.8	0.6	0.03	P3	131.7	2.0	0.02
P12	-269.9	1.6	0.02	P12	132.4	2.7	0.02
P14	-269.5	2.0	0.02	P14	132.6	2.9	0.01
P8	-267.8	3.6	0.01	P8	133.8	4.0	0.01
P22	-267.7	3.7	0.01	P22	133.8	4.1	0.01
P21	-267.3	4.2	0.01	**P21**	134.1	4.4	0.00
P10	-266.9	4.5	0.00	P10	134.4	4.7	0.00
P25	-266.9	4.6	0.00	P25	134.5	4.8	0.00
cP20	-266.6	4.8	0.00	cP20	134.6	4.9	0.00
cP17	-266.5	4.9	0.00	cP17	134.7	5.0	0.00

c) Z3							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P14	-526.0	-73.2	0.41*	P14	171.1	-4.5	0.06*
P19	-498.2	-45.4	0.29	**P19**	174.0	-1.6	0.04*
cP18	-487.5	-34.7	0.23	**cP18**	175.3	-0.4	0.04*
P8	-487.4	-34.6	0.23	P8	175.3	-0.4	0.04*
P6	-487.4	-34.6	0.23	P6	175.3	-0.4	0.04*
P21	-475.1	-22.3	0.17	**P21**	176.8	1.2	0.03
cP5	-473.9	-21.1	0.16	cP5	177.0	1.3	0.02
P4	-472.0	-19.2	0.15	P4	177.2	1.6	0.02
P11	-471.1	-18.3	0.14	P11	177.4	1.7	0.02
P7	-470.7	-17.9	0.14	P7	177.4	1.8	0.02
P1	-465.0	-12.2	0.11	P1	178.2	2.5	0.02
P3	-464.2	-11.4	0.10	P3	178.3	2.6	0.02
cP20	-460.3	-7.5	0.08	cP20	178.8	3.2	0.01
P23	-458.9	-6.1	0.07	P23	179.0	3.4	0.01
P16	-457.8	-5.1	0.06	P16	179.2	3.5	0.01
cP17	-454.9	-2.1	0.05	cP17	179.6	4.0	0.01
P2	-454.8	-2.0	0.05	P2	179.6	4.0	0.01
cP9	-454.3	-1.5	0.04	cP9	179.7	4.0	0.01
P22	-449.6	3.2	0.01	P22	180.4	4.7	0.00
P15	-448.9	3.9	0.01	P15	180.5	4.8	0.00
P10	-448.8	4.0	0.01	P10	180.5	4.9	0.00
P25	-448.4	4.4	0.00	P12	180.7	5.0	0.00
P12	-447.9	4.9	0.00	P25	180.7	5.0	0.00

Table 4.6 Univariate analysis of the PCA components a) Z1, b) Z2 and c) Z3 produced from "non-infected simulations". Univariate models are ranked by AIC_c. *R^2 values with *P*-values < 0.05. Superfluous and random variables are highlighted in grey.

a) Z1							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P4	-1069.5	-67.9	0.25*	P4	-327.1	-34.8	0.15*
P2	-1059.7	-58.1	0.22	P2	-322.0	-29.8	0.13*
P16	-1049.2	-47.6	0.19	P16	-316.5	-24.2	0.11*
P10	-1044.5	-42.9	0.18	P10	-314.0	-21.7	0.10*
cP17	-1029.1	-27.5	0.12	cP17	-305.5	-13.3	0.07*
P7	-1028.1	-26.5	0.12	cP9	-305.0	-12.8	0.07*
cP9	-1026.2	-24.6	0.11	P7	-304.0	-11.7	0.07
P14	-1026.2	-24.6	0.11	P14	-303.8	-11.5	0.07
cP20	-1020.5	-18.8	0.09	cP20	-300.7	-8.5	0.05
P15	-1020.2	-18.6	0.09	P15	-300.6	-8.3	0.05
P6	-1018.8	-17.2	0.09	P6	-299.8	-7.6	0.05
P1	-1014.0	-12.4	0.07	P1	-297.1	-4.8	0.04
P11	-1014.0	-12.4	0.07	P11	-297.1	-4.8	0.04
P3	-1013.1	-11.5	0.07	P3	-296.6	-4.3	0.04
P25	-1012.9	-11.3	0.07	P12	-296.5	-4.2	0.04
P12	-1009.4	-7.7	0.05	P22	-294.4	-2.2	0.03
P22	-1008.6	-7.0	0.05	cP5	-290.6	1.7	0.02
cP5	-1002.7	-1.1	0.03	P23	-289.9	2.4	0.01
P23	-1001.4	0.2	0.02	P8	-289.4	2.9	0.01
P8	-1000.6	1.0	0.02	P25	-288.9	3.4	0.01

b) Z2							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
cP17	-819.8	-174.6	0.51*	cP17	-174.8	-141.4	0.44*
P16	-690.3	-45.1	0.18	P16	-71.1	-37.7	0.16*
P10	-689.1	-43.9	0.18	P10	-70.1	-36.6	0.16*
cP20	-688.4	-43.1	0.18	cP20	-69.5	-36.0	0.15*
cP9	-686.7	-41.5	0.17	cP9	-68.1	-34.6	0.15*
P25	-677.9	-32.6	0.14	P2	-60.6	-27.1	0.12*
P2	-675.9	-30.7	0.13	P3	-59.0	-25.5	0.12*
P3	-675.8	-30.6	0.13	P8	-50.7	-17.2	0.09*
P8	-666.2	-21.0	0.10	P7	-48.8	-15.4	0.08*
P7	-664.0	-18.8	0.09	P15	-47.4	-13.9	0.07*
P15	-662.3	-17.1	0.09	P22	-45.7	-12.2	0.07*
P22	-660.3	-15.1	0.08	P14	-43.3	-9.9	0.06*
P14	-657.9	-12.7	0.07	P6	-37.4	-4.0	0.04
P6	-650.7	-5.5	0.04	P1	-35.6	-2.1	0.03
P1	-648.6	-3.4	0.03	P25	-35.2	-1.8	0.03
P23	-648.1	-2.9	0.03	P23	-33.4	0.1	0.02
P12	-646.0	-0.8	0.02	P12	-31.6	1.9	0.01
P4	-643.9	1.3	0.02	P4	-30.6	2.9	0.01
P11	-642.5	2.7	0.01	P11	-30.3	3.1	0.01
cP5	-641.8	3.4	0.01	cP5	-29.8	3.7	0.01

Table 4.6. con't

c) Z3							
Stochastic Mean				Stochastic results			
	AIC_c	BIC	R^2		AIC_c	BIC	R^2
P10	-973.8	-56.3	0.22*	P10	-201.3	-24.5	0.11*
P4	-965.8	-48.3	0.19	P4	-197.7	-20.9	0.10*
P16	-960.5	-43.0	0.18	P16	-195.2	-18.4	0.09*
P2	-959.8	-42.3	0.17	P2	-194.9	-18.1	0.09*
cP9	-955.7	-38.2	0.16	cP9	-192.9	-16.1	0.08*
cP17	-954.0	-36.5	0.15	cP17	-192.1	-15.3	0.08*
P7	-949.9	-32.4	0.14	P7	-189.9	-13.1	0.07*
cP20	-945.6	-28.1	0.13	cP20	-188.1	-11.3	0.06*
P14	-938.1	-20.6	0.10	P14	-184.5	-7.6	0.05*
P15	-938.1	-20.6	0.10	P15	-184.4	-7.6	0.05*
P25	-936.6	-19.1	0.09	P11	-183.7	-6.9	0.05*
P11	-930.7	-13.2	0.07	P3	-180.8	-4.0	0.04*
P3	-927.1	-9.6	0.06	P6	-178.9	-2.1	0.03
P6	-926.9	-9.5	0.06	P1	-178.2	-1.4	0.03
P1	-925.5	-8.0	0.05	P8	-177.3	-0.5	0.02
P8	-923.8	-6.3	0.05	P23	-175.4	1.4	0.02
P23	-920.0	-2.5	0.03	cP5	-174.7	2.1	0.01
cP5	-918.7	-1.2	0.03	P12	-173.1	3.7	0.01
P12	-915.4	2.0	0.01	P25	-172.2	4.6	0.00
P22	-913.8	3.7	0.01	P22	-172.1	4.7	0.00

Table 4.7 Parameters included in models from multivariate analysis of i) all simulations. AIC^c, BIC and R^2 values are also given. Models: B = backward stepwise selection, F = forward stepwise selection, G = global, G – P22 = all parameters except P22, G – P23 = all parameters except P23, G – P25 = all parameters except P25; Z1, Z2 and Z3 are response variables derived from the PCA analysis.

	Z1						Z2						Z3					
	B	F	G	G – P25	G – P23	G – P22	B	F	G – P25	G	G – P23	G – P22	B	G – P23	F	G – P25	G	G – P22
P1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P3	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x
P4			x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
cP5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P6	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P7	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P8	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
cP9		x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
P10	x		x	x	x	x	x	x	x	x	x	x	x	x		x	x	x
P11	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x
P12	x		x	x	x	x			x	x	x	x	x	x	x	x	x	x
P14		x	x	x	x	x			x	x	x	x	x	x	x	x	x	x
P15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
cP17	x	x	x	x	x	x		x	x	x	x	x	x	x		x	x	x
cP18	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P19	x	x	x	x	x	x	x	x	x	x	x	x		x		x	x	x
cP20	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
P21		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P22	x	x	x	x	x		x	x	x	x	x		x	x	x	x	x	
P23	x	x	x	x		x	x	x	x	x		x			x	x	x	x
P24	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P25	x	x	x		x	x				x	x	x		x				x
AIC_c w_i	0.49	0.31	0.08	0.07	0.05	0.00	0.95	0.05	0.00	0.00	0.00	0.00	0.43	0.21	0.16	0.12	0.08	0.00
BIC w_i	0.91	0.09	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.99	0.00	0.01	0.00	0.00	0.00
R^2	0.87	0.87	0.87	0.87	0.87	0.87	0.42	0.41	0.42	0.42	0.40	0.39	0.33	0.34	0.33	0.34	0.34	0.32

Table 4.8 Parameters included in the models from the multivariate analysis of ii) infected simulations. AIC_c, BIC and R^2 values are also given. Models: B = backward stepwise selection, F = forward stepwise selection, G = global, G – P22 = all parameters except P22, G – P23 = all parameters except P23, G – P25 = all parameters except P25; Z1, Z2 and Z3 are the response variables derived from the PCA analysis.

	Z1		Z2		Z3	
	F	B	F	B	F	B
P1	x	x	x	x		
P2		x		x		x
P3		x		x		x
P4		x		x		x
cP5		x				x
P6		x	x	x		
P7	x			x		
P8	x	x		x		x
cP9	x	x		x		x
P10						
P11					x	
P12			x			
P14					x	
P15	x					
P16	x		x			
cP17						
cP18	x		x			
P19					x	
cP20			x			
P21						
P22						
P23						
P24						
P25						
AIC_c w_i	0.71	0.29	0.69	0.31	0.88	0.12
BIC w_i	0.91	0.01	0.97	0.03	1.00	0.00
R^2	0.97	0.97	0.68	0.69	0.13	0.14

Table 4.9 Parameters included in the models from the multivariate analysis of iii) non-infected simulations. AIC_c, BIC and R^2 values are also given. Models: B = backward stepwise selection, F = forward stepwise selection, G = global, G – P22 = all parameters except P22, G – P23 = all parameters except P23, G – P25 = all parameters except P25; Z1, Z2 and Z3 are the response variables derived from the PCA analysis.

	Z1		Z2		Z3	
	F	B	B	F	F	B
P1	x		x	x		
P2	x	x	x	x	x	x
P3	x	x	x	x		x
P4	x	x	x		x	x
cP5		x	x	x		x
P6	x	x	x	x	x	x
P7	x	x		x	x	x
P8	x	x	x	x	x	x
cP9	x	x	x	x		x
P10		x	x	x	x	x
P11		x	x	x		x
P12		x	x	x		x
P14	x	x	x	x		x
P15	x	x	x	x	x	x
P16	x	x	x	x	x	x
cP17	x	x	x	x	x	x
cP18	n/a	n/a	n/a	n/a	n/a	n/a
P19	n/a	n/a	n/a	n/a	n/a	n/a
cP20						
P21	n/a	n/a	n/a	n/a	n/a	n/a
P22	n/a	n/a	n/a	n/a	n/a	n/a
P23	x			x	x	
P24						
P25						
AIC_c w_i	0.74	0.26	0.83	0.17	1.00	0.00
BIC w_i	1.00	0.00	0.95	0.05	1.00	0.00
R^2	0.58	0.58	0.87	0.87	0.48	0.51

Overall, as sample size decreased from i) all simulations (n = 500) to iii) non-infected simulations (n = 250) then ii) infected simulations (n = 150), the number of parameters included in the backwards and forwards stepwise selection models also decreased. Backwards stepwise selection appears to perform poorly with the smaller sample sizes because the parameters included in the models are the first few parameters from the order in which they were entered into the analysis.

Models could not be ranked using R^2 because the values were often identical among the candidate models. In contrast, the values from the AIC_c and BIC weights were sufficiently variable to rank the models. Both information criteria indicated the same top candidate models, with BIC more heavily weighting top models.

Discussion

This study incorporated Information Theory into sensitivity analysis. The aim was to overcome disadvantages of the null hypothesis testing approach and to evaluate the benefit of including parameters to increase model explanatory power against increased outcome uncertainty through parameter estimation error. The first stage of ITSA used univariate analysis to rank individual parameters effects on model outcomes. This was informative for quantifying the main effects of parameters. It also provided an objective means of selecting the most explanatory parameter to seed forward stepwise regression analysis, since this can influence the final resulting model (Whittingham et al. 2006)). This problem is exacerbated when multicollinearity exists among the parameters; however, this was not present, which is not surprising since the input parameter values were drawn randomly using LHS.

The second stage of ITSA used multivariate analysis and information criteria to assess model parsimony. Global models (all parameters) were built to compare with semi-global models (all parameters except for a superfluous parameter) to evaluate the contribution of each superfluous parameter independent of other model parameters. However, constructing additional models in an objective manner, *a priori*, to assess parsimony was problematic. Ideally a fully factorial design of every parameter combination would create models for comparison, yet for the ORM, this would result in 10,000's of models to evaluate. Parameters for building models could not objectively be identified using univariate analysis since parameter rankings differed among the principal component response variables (Z1, Z2, Z3) produced from i) all simulations, ii) infected simulations and iii) non-infected simulations. Therefore, forward and backward stepwise selection were used to objectively create models. Unfortunately, stepwise selection models did not meaningfully represent the system because random (P25) and superfluous parameters (P22, P23) were included. Furthermore, backwards stepwise selection models were biased towards including parameters in the order from which they were entered into the analysis, a known problem for stepwise multiple regression (Whittingham et al. 2006).

It is likely sample size was too small for the number of parameters to discern their effects. The greater the number of parameters, the lower the impact of each individual parameter on model outcomes, decreasing the ability to differentiate between important and unimportant parameters; hence the inclusion of random and superfluous parameters in the multivariate models. As a general rule for multivariate analysis, there is a 10:1 ratio of sample size to parameter (Burnham and Anderson 2002). Given 50 unique model

specifications, a sample size of 500 was required. This was achieved for i) all simulations (assuming all observations were completely independent of each other), but not for the ii) infected simulations (n = 150) and iii) non-infected simulations (n = 250).

Model stochasticity may have increased observed variation, obscuring parameter effects. To check for this, mean values were calculated from stochastic runs produced by each unique simulation, and ITSA was repeated using the mean values. Yet, results were identical to the analysis of the complete data. The only difference was that the averaged results had fewer significant univariate models, but this is a consequence of having a reduced sample size, which had the further detriment of being insufficient in sample size for multivariate analysis.

Overall, univariate analysis of ITSA yielded more benefits of sensitivity testing than multivariate analysis. Parameter rankings occurred as expected, demonstrating proper model functionality and supporting ecological theory underlying the disease-host model. This was shown with the original response variables, in that higher density input values create larger raccoon populations (Figure 4.1) and greater contact rates lead to higher incidence (Figure 4.2), and was also shown using the PCA response variables. For example: target cell density (cP17) was a dominant parameter influencing disease intensity (Z1 from Tables 4.4, 4.5) for i) all simulations, ii) infected simulations. Furthermore, disease transmission rate (cP18) was a dominant parameter affecting ii) infected simulations. Both of these parameters were expected to have large effects given their importance in other modelled disease-host systems (Anderson and May 1991). Furthermore, non-ORM parameters were appropriately ranked: P24 (parameter flagging rabid simulations) was a highly explanatory variable for i) all simulations (P24 was not

included for ii) infected simulations and iii) non-infected simulations since its value did not vary), and P25 (random parameter) was consistently ranked as an insignificant parameter. The benefit of these results is to increase confidence that ORM parameters were ranked appropriately. Thus, it is reasonable to use the results to understand the disease-host system. As such, highly ranked parameters (e.g. P7, P9, cP17, cP18) are assumed to have a greater impact on raccoon rabies than lowly ranked parameters (e.g. cP5, P8, P11, P12).

In this regard, sensitivity analysis provides ecological insight about the raccoon-rabies disease host system. For instance, mortality rates (P6, P7) are identified as a critical component of raccoons rabies, as is found with mortality rates in demographic models of other species (Turchin 2003). More specifically, female mortality rate (P7) has a larger impact on the raccoon-rabies system than male mortality rate (P6), since it was most often ranked more highly (Tables 4.4, 4.5, 4.6). This corresponds to expectations of model design because the simulated death of mothers also results in the death of dependent young. A further benefit of this analysis is that field studies can be directed to focus efforts towards collecting data for the critical parameters.

The multivariate aspect of ITSA was designed to be the primary means of assessing model parsimony. To be more effective for the ORM a larger sample size is required, for example, a parameter to sample size ratio of at least 1:20. Another strategy would be to perturb parameter inputs by more extreme values. However, perturbing parameter inputs beyond known variation would be less ecologically informative about the system since any parameter (important or unimportant) is expected to impact model outcomes given a large enough input perturbation.

Fortunately, an assessment of parsimony was possible through univariate analysis. For example, the superfluous nature of P22 and P23 was indicated by their low rankings, which commonly had less explanatory power than the random variable (P25). This approach was subjective because the random variable (P25) cannot be used as an exact exclusion threshold since there is likely variation around its "explanatory" power of model outcomes. Hence, a precise explanatory level by which a parameter should be included or excluded from the model is unknown. Ecological insights gained from sensitivity testing also contribute to achieving model parsimony. For example, distinguishing between juvenile and adult birth rates (P10, P11) was identified as an insignificant system component since both parameters have low univariate explanatory powers. This information can be used to increase model parsimony by redesigning the model to use one overall parameter for birth rates, or by setting equal input values for P10 and P11. However, some insignificant parameters, such as P5, are required for model functionality, so cannot be eliminated. In these cases, their input values can be fixed and blocked from user input.

Future investigations are needed to evaluate using Information Theory as a new means of achieving model parsimony through sensitivity analysis. Aside from obtaining a sufficient sample size for ITSA, there are many types of information criteria that can be used in ITSA, because of differences in their underlying theory and analytical capabilities. For instance, BIC tends to favour parameter-poor models (Raftery et al. 1997), while AIC favours parameter-rich models (Stanley and Burnham 1998), as was evident in this study. Using more than one information criterion reduces the risk of having a biased assessment of results generated from one criterion (Kuha 2004).

Overall, ITSA still yields the traditional benefits of sensitivity analysis. Analysing response variables with regression enables determination of which parameters are responsible for the majority of observed variation. In this regard, the sensitivity analysis guides parameter estimation efforts because it identifies parameters that have the greatest effect on outcomes, and thus require more precise estimates. Other parameters are needed for model functionality, but should more appropriately have their input values restricted from user-modification (e.g. litter size, cP5). Understanding of model processes is increased, and model functionality can be evaluated by checking that parameters produce expected effects on response variables. An important caveat of ITSA is to create response variables that capture a range of spatial-temporal impacts parameters may have on model output so that parameters that differ in their scales of spatial and temporal effects can still be assessed.

It is arguable that the ORM was already a parsimonious model through careful development that only included parameters known to affect the raccoon-rabies disease-host system. However, ecological systems are usually so complex and poorly understood that often developing plausible models *a priori* is not possible (Johnson 1999). Modellers are commonly faced with choosing between a set of models. Sensitivity analysis is a crucial step in model development for assessing whether a model's dynamics realistically represent the system (Saltelli 2000). ITSA provides an objective method for doing this quantifiably and in the absence of having empirical data to validate model outcomes.

The ORM requires appropriate testing before it is used to explore the raccoon rabies disease-host system and to guide rabies control practices. Information Theory

sensitivity analysis contributes substantially to assessing model functionality, parameter estimation, and model parsimony for ORM development. The value of this exercise is being critical of the benefits of modelling, the purpose of modelling, and the purpose of data collection. Benefits of modelling to understand system dynamics are as much about model development as model outcomes.

Chapter 5

Measuring the effect of the Niagara River as a barrier to gene flow in raccoons (*Procyon lotor*) and its implications to the spread of raccoon rabies

Abstract

Landscape barriers have implications for infectious wildlife disease control. Genetic data is used to predict the barrier effect in an area where rabies has yet to occur: from New York State (NY) into Ontario by crossing the Niagara River. To do this, the expansion of a genetically marked raccoon population is simulated to cross the river from NY to Ontario under various barrier scenarios. Since the model records genetics of individual raccoons, neutral mitochondrial DNA haplotype markers are tracked in the expanding population and characterized (at 25 year intervals) the genetic population structure using ϕ_{ST}, Mantel tests and a gene diversity measure. Barrier effects are assessed by comparing the genetic measures to those calculated from haplotypes of 166 raccoons recently sampled from the same landscape. "Best fits" between modelled scenarios and empirical data indicated that the Niagara River reduces movement by 50 percent. Founder effects dominated the colonizing genetic population structure, and, as the river barrier effect increased, genetic diversity decreased. Using gene flow as an analogue to disease spread, it is concluded that the river will cause a similar reduction in movement on the spread of rabies. Including individual genetic markers in simulation modelling benefits investigations of species range expansion and disease spread.

Introduction

Infectious disease modelling has made important contributions to epidemiology and ecology by quantifying infectious disease systems to enable explorations of their fundamental characteristics. Model structure ranges from simulating average populations for determining general trends, to those that simulate variation in disease-host behaviours in heterogeneous environments for addressing more specific questions. In this study, an individual-based spatial simulation model is used to investigate the effect of a major river on the spatial-temporal genetic structuring of raccoons for inferring the rate of raccoon rabies disease spread.

Raccoon rabies is a variant rabies virus specifically adapted to infect raccoons (Winkler and Jenkins 1991). It was first detected in Florida in the 1940's, and a second major epizootic emerged in the late 1970's along the West Virginia/Virginia border (Winkler and Jenkins 1991), spreading northwards at a rate of 30 - 47 km/year (Childs et al. 2000)). It reached Canada near Brockville, Ontario, in 1999 (Wandeler and Salsberg 1999) by crossing the St. Lawrence River from northern New York State (NY) and has the potential to further infiltrate south-central Ontario by crossing the Niagara River from western NY (Rosatte et al. 1997).

The spread of raccoon rabies has occurred as an irregular wave. The spatial-temporal variations are largely attributed to physiography and habitat quality of the landscape, which in turn affect movement patterns and distribution of animals at risk (Childs et al. 2001)). Quantifying the effect of various landscape barriers has, therefore, practical implications for rabies control planning in North America (Slate et al. 2005).

Rabies incidence records have been used to measure landscape barrier effects to disease flow (Smith et al 2002); however there are several difficulties with this approach: (a) assessing landscape effects in areas with no prior record of rabies; and (b) the highly variable quality of rabies incidence data that, in turn, can obscure measures of spread in both space and time. There are a number of potential sources of bias in rabies data in North America. i) Rabies surveillance is passive and typically monitors incidence when humans are at risk. Consequently, the number of reported cases may be influenced by the density of the human population, meaning that many animals die undetected because they are not observed {Childs, Curns, et al. 2001 132 /id}); ii) under reporting is typical once rabies is established in an area because people become complacent about disease presence and the surveillance system is overloaded {Wilson, Bretsky, et al. 1997 12 /id}; iii) A variety of jurisdictions are responsible for rabies surveillance (e.g. province, counties, townships), and these differ in their budgets, mandates and reporting procedures {Childs, Curns, et al. 2001 132 /id}); and iv) most data have been collected on the basis of administrative units such as towns and counties which do not always correlate in size or location with environmental, ecological or biological factors affecting disease incidence {Lawson 2001 150 /id}). There are strategies for controlling the first three reporting biases {Childs, Curns, et al. 2000 130 /id}). In general, however, quality of disease incidence data is problematic when assessing relationships of rabies spread to environmental, ecological or biological factors, and reporting units. Therefore, incidence data tend to be more valuable as an early warning system rather than for a detailed epidemiological assessment.

Genetic data are explored for assessing barrier effects independent of incidence data. The quality of genetic data depends on a sampling design that adequately covers genetic variation observed over space and time. Neutral genetic markers are useful for population genetic analyses because they are not subject to selective pressures. New variations that arise do not affect reproductive or survival rates of animals (fitness), thus variations will be maintained in the population, and will be distributed as reflected by mating and dispersal processes. Therefore, neutral markers can be used as a "tag" to identify spatial-temporal patterns resulting from these processes. This is in contrast to selected markers, where genotypes that increase fitness occur at higher frequencies, because these animals will have more success at spreading their genes. Portions of mitochondrial DNA (mtDNA) contain neutral markers that have considerable variation in frequency of unique genetic sequences (haplotypes) within and among populations, making them suitable for assessing genetic population structure {Harrison 1989 95 /id}).

Is it reasonable to apply neutral markers to track movement patterns for exploring mechanisms influencing spatial-temporal disease patterns? This question arises because: a) rabies propagates and spreads more quickly than raccoon genetics (e.g. raccoon rabies disease spread of 30-47 km/year (Childs et al. 2000)) versus raccoon annual dispersal distances commonly being less than 5 km; Rosatte 2000); and b) rabies affects the population dynamics (Rosatte et al. 2006, 2007a, 2007b). Rabies infection and gene flow are different processes. Any immune susceptible animal can become infected, at any time of year, through successful transmission of the virus caused by direct contact (e.g. bite, scratch), which may even come from another species (e.g. skunk). Spreading of genes requires animals to have reached reproductive maturity, dispersed and successfully

mated in a non-natal location. With regards to population dynamics, gene flow of neutral markers does not affect genetic and demographic processes influencing the distribution of genes (e.g. genetic drift, dispersal, reproduction, mortality); however, the virus has this capability. For instance, there may be a genetic predisposition for surviving the disease. This would influence the genetic population structure if linkage disequilibrium existed between the more "fit" functional genes and the neutral markers. Furthermore, the genetic structure of a population reduced in size by disease could be altered by the mechanisms of genetic drift and founder effects of individuals settling unpopulated areas.

An additional consideration is that the analysis of mtDNA gives a maternal perspective, since it is maternally inherited. This is not an issue when males and females have similar dispersal patterns and mating systems. However, in Ontario male raccoons disperse over greater distances, have larger home ranges and tend to mate with more partners than females {Rosatte 2000 6 /id}). Consequently, model and genetic analysis might underestimate the degree of gene flow and the amount and distance of average raccoon movement.

Despite spatial-temporal differences between disease spread and population dynamics and their respective effects on the genetic population structure, similarities in their fundamental mechanisms still supports the use of gene flow as being informative about disease spread. Both of these systems are highly influenced by raccoon density. Densities must be sufficient to increase the likelihood of interactions leading to copulation or infection. Raccoon movement is also a critical mechanism. Interactive movement behaviours that result in conception can also enable the transmission of the disease. Furthermore, infected young-of-year raccoons dispersing in the fall to new

locations and then mating in the spring, could also become infectious in the spring and transmit the disease to these new locations, in cases where viral incubation periods last several months {Jackson 2002 189 /id}).

In this study genetic data are used to assess a potential barrier through which spread has yet to occur and, therefore, there is no disease incidence data to explore this issue. The Ontario Rabies Model (ORM; Tinline et al. 2007) simulated mtDNA gene flow across the Niagara River (43°N 79°W) at varying levels of permeability. Temporal "snapshots" of the simulated genetic population structure were compared with the one derived from field data. The ORM is a stochastic spatial simulation model currently configured to simulate raccoon population dynamics, to track maternal and bi-parental genetic inheritance, to simulate rabies viral transmission and to model rabies control strategies (vaccination, culling and fertility control; Tinline et al. 2007). The objectives of this study were to i) quantify the effect of the Niagara River as a landscape barrier to maternal gene flow, and ii) infer how the Niagara River will influence the rate of raccoon rabies disease spread from NY into south-central Ontario from the Niagara region, and, in doing so, assess the usefulness of gene flow in measuring barrier effects.

Methods

(a) Study area

The Niagara Region of Ontario and New York State consists of flat rural agricultural lands. Cutting east to west through the region is the Niagara Escarpment, which has a cliff face of approximately 25 metres in height. The Niagara River flows over the escarpment, providing an outflow of water from Lake Erie into Lake Ontario.

The river has an average width of 0.5 km, dropping 99 metres over its course of 58 kilometres. The river is a major source of hydroelectricity for Canada and the United States (*www.energy.gov.on.ca*). Being a fast moving river with only 5 bridges, it is assumed to be the major landscape barrier to raccoon movement in this region.

(b) Ontario Rabies Model

The ORM operates on a lattice of hexagonal cells, each with a radius of 2 km (from cell centre to each vertex of the hexagon), the approximate activity range of raccoons in Ontario (Rosatte 2000). The resulting cell area is 10.39 km^2. The shape and size of this lattice of cells is user-definable. Experiments used a lattice of 2255 cells covering an area of approximately 23 500 km^2 on either side of the Niagara River. Model processes operate at one-week intervals and are stochastically determined. For example, each year in week 18, juvenile (52 – 74 weeks old) and adult females ($\geq$ 75 weeks old) have a chance of producing a number of offspring as determined by a probability distribution.

Probability distributions in the ORM defining demographic behaviours (e.g. mortality rates, dispersal distances, density, litter size) were derived from field and laboratory studies undertaken in Ontario by the Ontario Ministry of Natural Resources {Totton, Rosatte, et al. 2004 #886}). Reported values used within the model fall within variation observed in eastern North America (Chapter 2).

Model experiments typically use an initial raccoon population, which has been grown to a stable population in the whole study area from a pair of raccoons. Every cell has a defined carrying capacity, k. This carrying capacity serves as a target population for a cell and is used to simulate a biological carrying capacity through density dependent

feedback influencing mortality rates. If the cell population is greater than *k*, mean mortality rates are increased. If the cell population is less than *k*, mean mortality rates are decreased. The amount of increase or decrease in mean mortality rates is adjusted by a density dependent mortality control parameter, which derives its value through a model fitting process minimising the deviation between observed and modelled raccoon demographics. *k* can be used to simulate the effect of different habitat types on raccoon densities. However, investigations of raccoon densities and land cover types, based on trapping records, indicate no significant relationship between these two factors in the Niagara region at the scale of the model cells. Therefore, a single *k* value was applied uniformly throughout the study area to represent homogeneous habitat conditions of 5 raccoons / km^2 {Rosatte 2000 6 /id}).

(c) *Experimental design and model specifications*

A population was grown for 250 years to stability on the model landscape, after which all raccoons were removed from the Ontario cells (Figure 5.1). Maternal inheritance of mtDNA was modelled by genetically "tagging" the members of the simulated initial raccoon population with known frequencies of the known haplotypes from the empirical data. Offspring inherited their maternal haplotype.

Mechanisms of novel genetic variation are not simulated for this study because mutation is assumed to be non-existent over the time span of which the model operates (450 years) and recombination is rare or non-existent in mtDNA {Harrison 1989 95 /id}). Assigning all the landscape cells to one of four regional groupings simulated the Niagara River: A) Ontario, B) Navy Island, ON, C) Grand Island, NY, and D) NY (Figure 5.1).

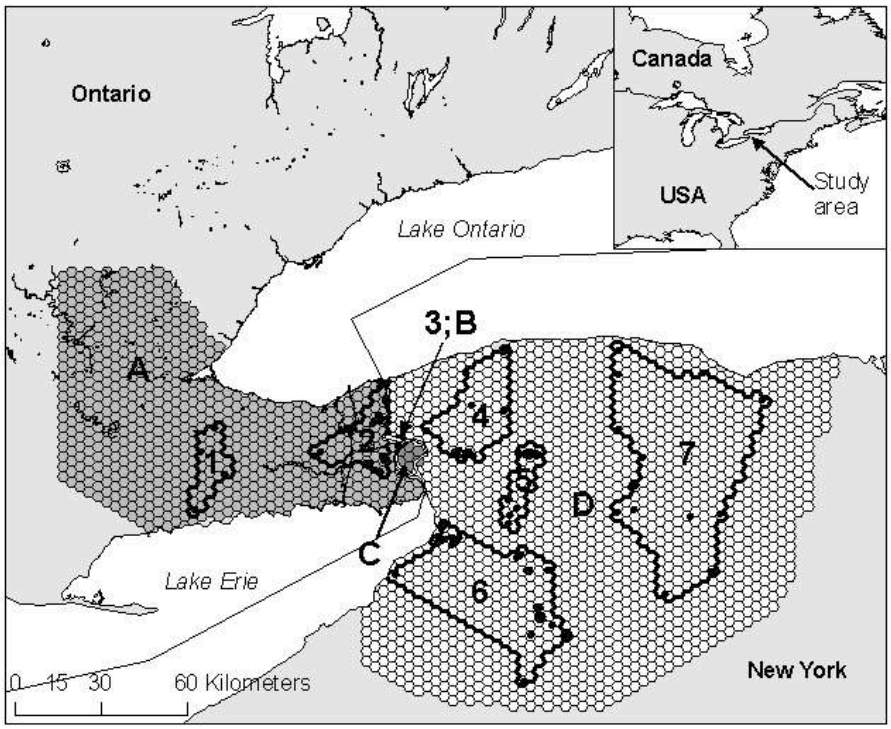


Figure 5.1 Study site in the Niagara Region of Ontario and New York State. Four regional groupings used to create the Niagara River barrier effect are: A) Ontario, B) Navy Island, ON, C) Grand Island, NY, and D) NY. Also shown are the point locations of raccoons sampled from the landscape and the ORM hexagonal cells defining the seven sample groups.

While, raccoon movement within these regions is unimpeded, the animals must overcome a user-defined resistance to move among the regions.

Raccoon presence in south-central Ontario is assumed to be largely from a range expansion across the Niagara River following the last glacial event (Cullingham et al. 2007). The ORM was used to reconstruct this event and simulated results were compared with empirical data. Specifying the river to block 0, 25, 50 and 75 percent of raccoon movement tested the barrier effect of the Niagara River on the genetic population structure. A critical issue for measuring the barrier effect is to decide at which points in time the simulated genetic population structure should be compared with empirical data, since time since colonisation affects the genetic population structure {Hutchinson & Templeton 1999 219 /id}). For this reason, the genetic population structure was measured every 25 years over the course of a 200 year colonisation process that started after the initial population had grown for 250 years (0 to 249); thus, measured at simulation years: 274, 299, 324, 349, 374, 399, 424 and 449. Each unique model input specification was run 10 times to sufficiently capture inherent variation from the stochastic model. Thus, there were 320 model outputs from 4 barrier effects x 10 trials x 8 time intervals (Figure 5.2).

(d) Genetic field samples and laboratory procedures

One hundred sixty-six raccoons sampled in 2003 were obtained from Ontario Ministry of Natural Resources, and fur harvesters and consisted of pelt and hair samples. The raccoon locations were georeferenced to point locations (usually < ± 500m). The samples were stored dry until DNA extraction.

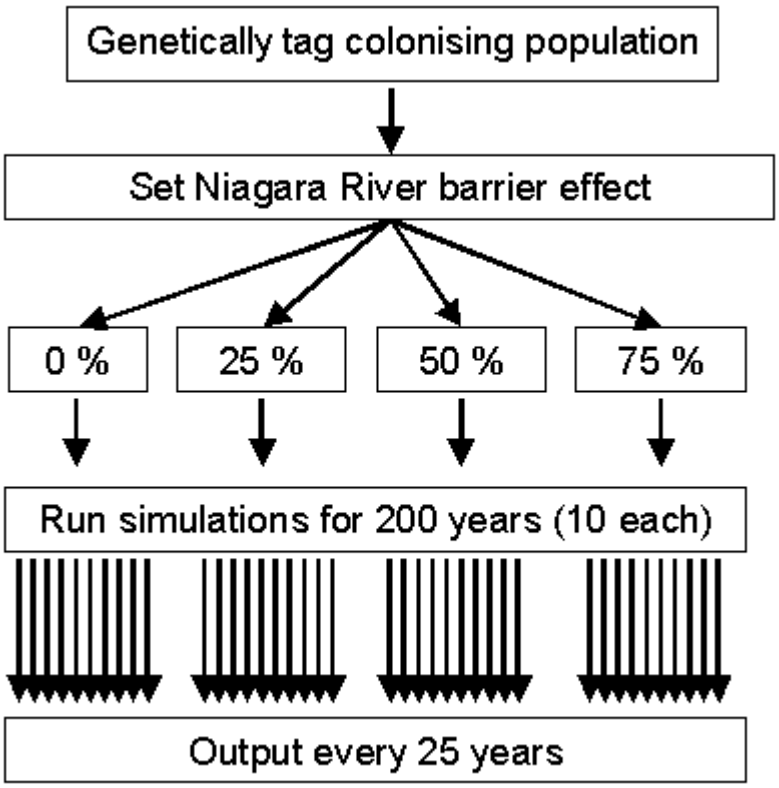


Figure 5.2 Experimental design for modelling the range expansion of raccoons from NY into Ontario, for varying permeabilitites of the Niagara River set to block 0, 25, 50 and 75% raccoon movement.

Samples were digested using 1X lysis buffer (Applied Biosystems) and 600U/mL proteinase K. Extraction was carried out using an automated magnetic bead procedure (Promega) and samples were diluted to 0.5ng/uL using $TE_{0.1}$. Polymerase chain reactions (PCR) were performed using primers L15997 {Ward, Frazier, et al. 1991 299 /id}) and PLO-CRL1 (CGCTTAAACTTATGTCCTGTAACC). The reaction conditions are as follows: standard buffer conditions, 2mM $MgCl_2$, 160uM of each dNTP, 0.3 uM of each primer, and 1 unit of *Taq* DNA polymerase (Invitrogen). The cycling protocol used is 30 cycles following steps: 94ºC for 30 s, 58ºC for 30 s, and 72ºC for 30 s, preceded by 5 min of initial denaturing, and followed by 2 min of final extension. PCR products were purified using ExoSAP-IT (USB) following the manufacturers' instructions. Sequencing using the reverse primer was carried out using the DYEnamic™ ET terminator cycle sequencing kit, and the resulting fragments were analyzed on a MegaBASE 1000 (Amersham-Pharmecia). Fragments were visually inspected, corrected and aligned manually in BioEdit (Hall 1999). Any unique sequences were confirmed with forward sequencing (Cullingham et al. 2007).

(e) *Haplotype diversity*

Saturation curves were calculated using EstimateS (Corwell 2005) to estimate the proportion of total haplotype diversity achieved from the sampled raccoons. The 166 field samples were drawn eight times (n_1 to $n_6 = 21$, n_7 to $n_8 = 20$) without replacement, and a bootstrap estimator was used to calculate diversity {Hellmann & Fowler 1999 282 /id}).

(f) *Genetic analysis*

Seven sample regions were defined within the landscape to spatially correspond with the locations of at least 20 raccoons sampled for genetic analysis and to be adequately spaced to test for Isolation by Distance (IBD; to determine if raccoons become more genetically differentiated from each other with an increase in geographic distance) (Figure 5.1). Arlequin 3 (Excoffier et al. 2005) was used to calculate genetic measures for each sample group (sg) every 25 years. ϕ_{ST} defined genetic distance between pairwise comparisons of samples groups, using haplotype frequencies {Nei 1977 279 /id}), to indicate the difference in haplotype diversity between two populations. To test for IBD, simple Mantel testing assessed whether there was a significant correlation between ϕ_{ST} and geographic distance, and partial Mantel testing assessed this correlation while controlling for the barrier effect. Mantel tests were also used to explore the correlation between ϕ_{ST} and barriers, with the partial test controlling for a geographic distance effect. A "gene diversity" measure for each sg defined the probability of randomly selecting two different haplotypes from the same sg, as an indication of the number of unique haplotyes in the sg {Nei 1987 302 /id}).

Expectations of model behaviour were: 1) The occurrence of an IBD raccoon population genetic structure that would decrease over time as raccoons further dispersed and reproduced; 2) Characteristics of a founder effect would be evident in the newly colonised ON sg's (lower gene diversity measures than the NY sg's), and the force of the founder effect would increase with greater barrier effects; and 3) Smaller sg's would have lower gene diversity than larger sg's; as assessed by regressing the gene diversity measure against the sg area.

To measure the Niagara River barrier effect, genetic measures were calculated for the raccoon field samples and then compared with the model results. Correlation analyses were performed between the simulated and empirical genetic measures, whereby, stochastic variation was addressed by averaging genetic measures for each unique model input specification before comparing with field data. The rationale was that the model barrier effect most closely matching field results would define the magnitude of the Niagara River as a barrier to gene flow. Results were visualised from 3-D surfaces interpolated using Gamma Design Software GS+ version 5.3.2 (*www.gammadesigns.com*) relative to simulation year, barrier effect and genetic measure. Empirical data was tested for IBD using a Mantel's test.

Results

Genetic laboratory analysis found 19 unique haplotypes from the 166 sampled raccoons. An estimated ≥90% of the total haplotypic diversity was captured (19 of 21), as indicated by the asymptotic nature of the haplotype diversity richness curve (Figure 5.3).

In the ORM, the colonizing population required 25 to 50 years to occupy all sg's in the Ontario landscape. Consequently, genetic measures are presented for ≥50 years since the start of the colonisation (model year 299 and onwards).

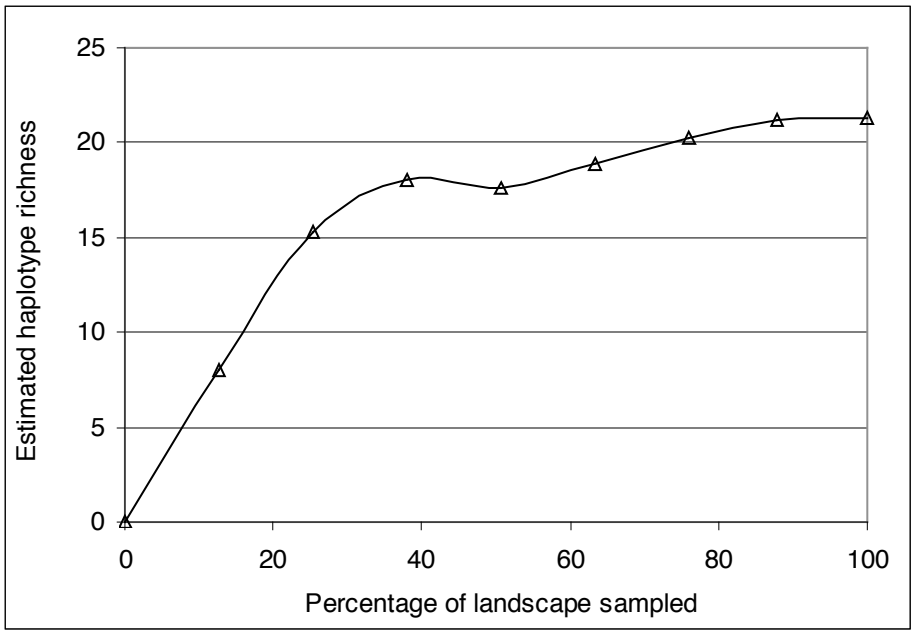


Figure 5.3 Haplotype diversity of the raccoons sampled from the field, as estimated using a bootstrap estimator.

(a) Model behaviour

There was no indication of an IBD genetic population structure, as simple and partial Mantel testing (controlling for barriers) revealed no significant relationship between the correlation of ϕ_{ST} and geographic distance for simulated or empirical data. There was, however, a significant correlation between ϕ_{ST} and barriers for the simulated and empirical data, which was strengthened when using a partial Mantel test to control for the effect of geographic distance (Table 5.1).

A founder effect was evident in that gene diversity measures were lower for ON sg's, and decreased with an increasing barrier effect (Table 5.2). Furthermore, smaller sg's had lower gene diversities than larger sg's (gene diversity = 0.43area + 0.49; adjusted $R^2 = 0.19$, *F*-ratio = 452.43, *P*-value <0.01; Table 5.2).

(b) Measuring the Niagara River barrier effect

Strongest correlations comparing the empirical and simulated genetic measures data was found for a barrier effect of 50% at simulation year 349 (gene diversity: r = 0.91; *P*=0.005; ϕ_{ST}: r = 0.83; *P*<0.001; Table 5.3).

Three dimensional interpolated surfaces illustrate an increasing barrier effect associated with decreasing gene diversity for sg2 and sg3, an effect which is relatively constant over time (Figures 5.4, 5.5). Conversely, the barrier had a negligible effect on sg1, the furthest ON sg from the Niagara River, and no effect of the NY sg's (Figures 5.4, 5.5).

Table 5.1 Simple and partial Mantel correlation and P-values for Niagara River barrier effects of 0%, 25%, 50% and 75%, that test for the correlation between Φ_{ST} and geographic distance (g), Φ_{ST} and Niagara River barrier (b), Φ_{ST} and geographic distance, controlling for Niagara River barrier (g.b), and Φ_{ST} Niagara River, controlling for geographic distance (b.g). *P*-values are in parentheses; significant correlations (≤ 0.05) are in bold.

		Model simulation year						
		299	324	349	374	399	424	449
0%	g	0.33 (0.21)	0.37 (0.21)	0.36 (0.20)	0.42 (0.16)	0.38 (0.17)	0.44 (0.10)	0.39 (0.16)
	b	**0.53 (0.05)**	0.53 (0.08)	0.50 (0.06)	0.52 (0.06)	**0.58 (0.04)**	**0.54 (0.03)**	**0.62 (0.02)**
	g.b	0.22 (0.34)	0.27 (0.32)	0.26 (0.30)	0.34 (0.24)	0.28 (0.28)	0.37 (0.20)	0.32 (0.25)
	b.g	0.50 (0.11)	0.49 (0.12)	0.45 (0.11)	0.48 (0.09)	0.55 (0.07)	0.50 (0.09)	**0.62 (0.05)**
25%	g	0.31 (0.20)	0.34 (0.18)	0.27 (0.22)	0.28 (0.21)	0.39 (0.14)	0.30 (0.22)	0.38 (0.16)
	b	**0.56 (0.02)**	**0.50 (0.03)**	**0.49 (0.03)**	**0.58 (0.03)**	**0.51 (0.05)**	**0.43 (0.05)**	**0.50 (0.03)**
	g.b	0.20 (0.32)	0.23 (0.32)	0.17 (0.34)	0.17 (0.35)	0.32 (0.22)	0.22 (0.33)	0.30 (0.24)
	b.g	0.53 (0.06)	0.46 (0.11)	0.46 (0.08)	0.55 (0.06)	0.47 (0.10)	0.39 (0.12)	0.46 (0.09)
50%	g	0.06 (0.44)	0.07 (0.42)	0.11 (0.33)	0.26 (0.27)	0.16 (0.35)	0.17 (0.33)	0.25 (0.27)
	b	**0.54 (0.03)**	**0.58 (0.03)**	**0.58 (0.02)**	**0.58 (0.04)**	**0.61 (0.01)**	**0.55 (0.04)**	0.54 (0.06)
	g.b	-0.11 (0.61)	-0.13 (0.62)	-0.06 (0.51)	0.11 (0.43)	-0.01 (0.52)	0.02 (0.50)	0.12 (0.40)
	b.g	**0.56 (0.05)**	**0.60 (0.05)**	**0.59 (0.04)**	0.57 (0.06)	**0.64 (0.02)**	0.55 (0.07)	0.52 (0.08)
75%	g	-0.05 (0.50)	0.05 (0.40)	0.12 (0.34)	0.11 (0.34)	0.07 (0.39)	0.14 (0.31)	0.13 (0.34)
	b	**0.71 (0.02)**	**0.73 (0.01)**	**0.70 (0.01)**	**0.74 (0.01)**	**0.71 (0.01)**	**0.72 (0.01)**	**0.73 (0.01)**
	g.b	-0.27 (0.71)	-0.16 (0.64)	-0.10 (0.57)	-0.13 (0.60)	-0.17 (0.65)	-0.08 (0.58)	-0.08 (0.53)
	b.g	**0.75 (0.01)**	**0.75 (0.01)**	**0.71 (0.01)**	**0.76 (0.01)**	**0.73 (0.01)**	**0.73 (0.01)**	**0.75 (0.01)**

Table 5.2 Gene diversity measures: mean from the 10 runs of each unique input specification and its standard deviation, in parentheses, and the overall mean (OM) for the entire 200 years of the simulated colonisation event, as calculated for each sample group (sg) and barrier effect.

Barrier effect	sg	Simulation year							
		299	324	349	374	399	424	449	OM
0%	1	0.40 (0.12)	0.44 (0.13)	0.43 (0.15)	0.44 (0.17)	0.43 (0.20)	0.41 (0.19)	0.42 (0.17)	0.43 (0.16)
	2	0.64 (0.12)	0.62 (0.14)	0.62 (0.12)	0.61 (0.12)	0.59 (0.14)	0.59 (0.15)	0.60 (0.14)	0.61 (0.13)
	3	0.40 (0.22)	0.55 (0.19)	0.53 (0.18)	0.52 (0.18)	0.61 (0.17)	0.50 (0.16)	0.60 (0.16)	0.53 (0.18)
	4	0.70 (0.04)	0.69 (0.06)	0.70 (0.05)	0.69 (0.05)	0.69 (0.06)	0.70 (0.05)	0.69 (0.07)	0.69 (0.05)
	5	0.72 (0.05)	0.73 (0.03)	0.75 (0.05)	0.72 (0.04)	0.70 (0.06)	0.71 (0.05)	0.69 (0.09)	0.72 (0.06)
	6	0.72 (0.04)	0.72 (0.04)	0.72 (0.05)	0.72 (0.05)	0.71 (0.06)	0.70 (0.07)	0.70 (0.08)	0.71 (0.06)
	7	0.74 (0.03)	0.73 (0.03)	0.74 (0.03)	0.72 (0.03)	0.72 (0.04)	0.73 (0.04)	0.74 (0.04)	0.73 (0.03)
25%	1	0.29 (0.23)	0.25 (0.24)	0.24 (0.22)	0.19 (0.22)	0.16 (0.19)	0.16 (0.18)	0.20 (0.19)	0.27 (0.23)
	2	0.53 (0.21)	0.56 (0.22)	0.54 (0.22)	0.52 (0.22)	0.51 (0.23)	0.51 (0.20)	0.49 (0.23)	0.54 (0.20)
	3	0.57 (0.21)	0.59 (0.13)	0.43 (0.18)	0.42 (0.20)	0.63 (0.31)	0.63 (0.13)	0.43 (0.25)	0.56 (0.20)
	4	0.72 (0.04)	0.71 (0.06)	0.70 (0.06)	0.69 (0.06)	0.68 (0.07)	0.68 (0.09)	0.67 (0.11)	0.70 (0.07)
	5	0.69 (0.10)	0.71 (0.09)	0.72 (0.05)	0.70 (0.06)	0.71 (0.06)	0.71 (0.07)	0.71 (0.08)	0.71 (0.07)
	6	0.74 (0.05)	0.73 (0.05)	0.72 (0.06)	0.70 (0.07)	0.70 (0.09)	0.70 (0.08)	0.71 (0.08)	0.72 (0.07)
	7	0.74 (0.04)	0.74 (0.04)	0.74 (0.04)	0.74 (0.03)	0.73 (0.03)	0.73 (0.04)	0.72 (0.04)	0.74 (0.04)

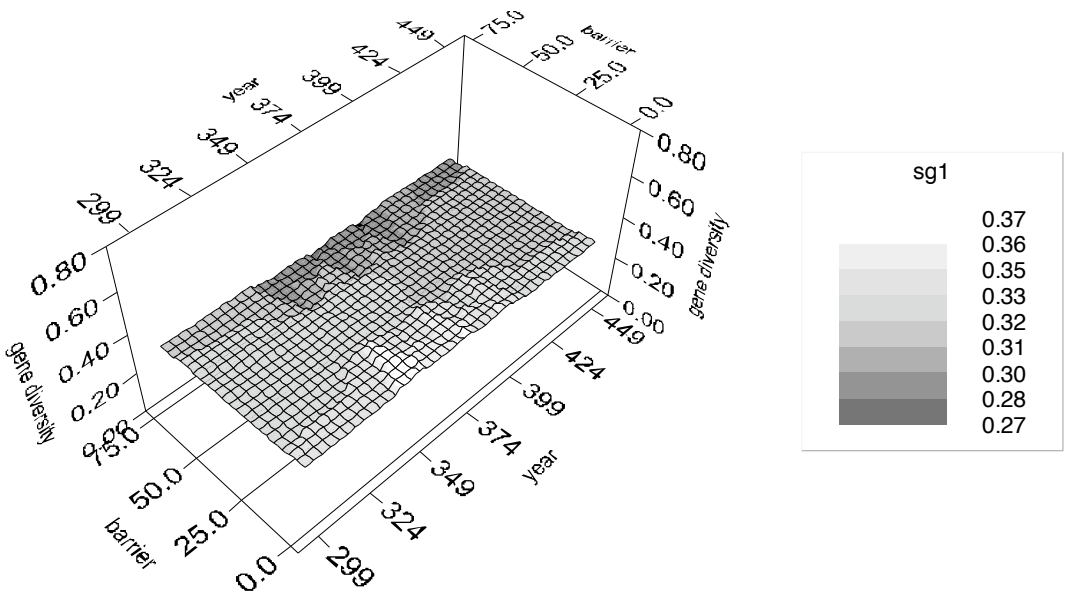
Table 5.2. con't

50%	1	0.42 (0.20)	0.47 (0.20)	0.47 (0.20)	0.44 (0.21)	0.44 (0.21)	0.40 (0.24)	0.39 (0.24)	0.43 (0.21)
	2	0.50 (0.21)	0.52 (0.20)	0.49 (0.22)	0.52 (0.19)	0.48 (0.22)	0.49 (0.20)	0.51 (0.19)	0.50 (0.20)
	3	0.55 (0.19)	0.43 (0.20)	0.35 (0.25)	0.54 (0.15)	0.43 (0.16)	0.59 (0.11)	0.53 (0.21)	0.49 (0.19)
	4	0.71 (0.05)	0.71 (0.06)	0.70 (0.05)	0.68 (0.07)	0.68 (0.06)	0.66 (0.09)	0.68 (0.08)	0.69 (0.07)
	5	0.73 (0.05)	0.72 (0.04)	0.69 (0.07)	0.69 (0.08)	0.67 (0.09)	0.69 (0.10)	0.66 (0.08)	0.69 (0.08)
	6	0.73 (0.03)	0.72 (0.03)	0.73 (0.04)	0.73 (0.04)	0.71 (0.05)	0.71 (0.08)	0.71 (0.06)	0.72 (0.05)
	7	0.75 (0.02)	0.74 (0.03)	0.73 (0.04)	0.72 (0.05)	0.72 (0.06)	0.71 (0.08)	0.70 (0.07)	0.73 (0.05)
75%	1	0.11 (0.19)	0.16 (0.24)	0.22 (0.26)	0.21 (0.25)	0.21 (0.26)	0.21 (0.25)	0.22 (0.26)	0.19 (0.24)
	2	0.30 (0.24)	0.28 (0.24)	0.28 (0.23)	0.27 (0.24)	0.25 (0.26)	0.24 (0.27)	0.21 (0.26)	0.26 (0.24)
	3	0.17 (0.21)	0.10 (0.14)	0.28 (0.28)	0.18 (0.20)	0.20 (0.18)	0.19 (0.20)	0.09 (0.12)	0.17 (0.20)
	4	0.72 (0.03)	0.69 (0.05)	0.69 (0.05)	0.70 (0.06)	0.68 (0.08)	0.68 (0.07)	0.70 (0.07)	0.69 (0.06)
	5	0.72 (0.03)	0.69 (0.05)	0.71 (0.06)	0.72 (0.05)	0.72 (0.04)	0.71 (0.04)	0.70 (0.06)	0.71 (0.05)
	6	0.72 (0.04)	0.70 (0.04)	0.70 (0.04)	0.71 (0.04)	0.68 (0.06)	0.68 (0.06)	0.69 (0.04)	0.70 (0.05)
	7	0.75 (0.02)	0.75 (0.02)	0.75 (0.03)	0.74 (0.03)	0.75 (0.03)	0.75 (0.03)	0.75 (0.03)	0.75 (0.03)

Table 5.3 Correlation (and *P*-values) from comparing empirical to model data for each simulation year and barrier effect. Significant correlations (*P*-value ≤ 0.05) are in bold.

barrier effect	year	gene diversity	ϕ_{ST}
0%	299	**0.81 (0.027)**	**0.48 (0.027)**
	324	0.63 (0.131)	0.43 (0.051)
	349	0.66 (0.105)	**0.47 (0.032)**
	374	0.69 (0.087)	**0.49 (0.024)**
	399	0.42 (0.342)	**0.53 (0.013)**
	424	0.70 (0.083)	**0.44 (0.047)**
	449	0.45 (0.310)	**0.64 (0.002)**
25%	299	0.44 (0.321)	**0.49 (0.026)**
	324	0.38 (0.405)	0.38 (0.089)
	349	0.64 (0.125)	0.42 (0.056)
	374	0.61 (0.149)	**0.58 (0.006)**
	399	0.23 (0.149)	**0.51 (0.019)**
	424	0.23 (0.614)	0.35 (0.117)
	449	0.59 (0.165)	**0.57 (0.007)**
50%	299	0.60 (0.155)	**0.55 (0.009)**
	324	**0.86 (0.014)**	**0.67 (0.001)**
	349	**0.91 (0.005)**	**0.83 (<0.001)**
	374	0.62 (0.140)	**0.51 (0.018)**
	399	**0.80 (0.029)**	**0.71 (<0.001)**
	424	0.42 (0.344)	**0.48 (0.026)**
	449	0.57 (0.182)	**0.56 (0.009)**
75%	299	**0.77 (0.043)**	**0.54 (0.012)**
	324	**0.84 (0.018)**	**0.57 (0.007)**
	349	0.74 (0.056)	**0.52 (0.016)**
	374	**0.81 (0.028)**	**0.62 (0.003)**
	399	**0.79 (0.035)**	**0.55 (0.010)**
	424	**0.79 (0.035)**	**0.50 (0.022)**
	449	**0.85 (0.16)**	**0.58 (0.006)**

a)



b)

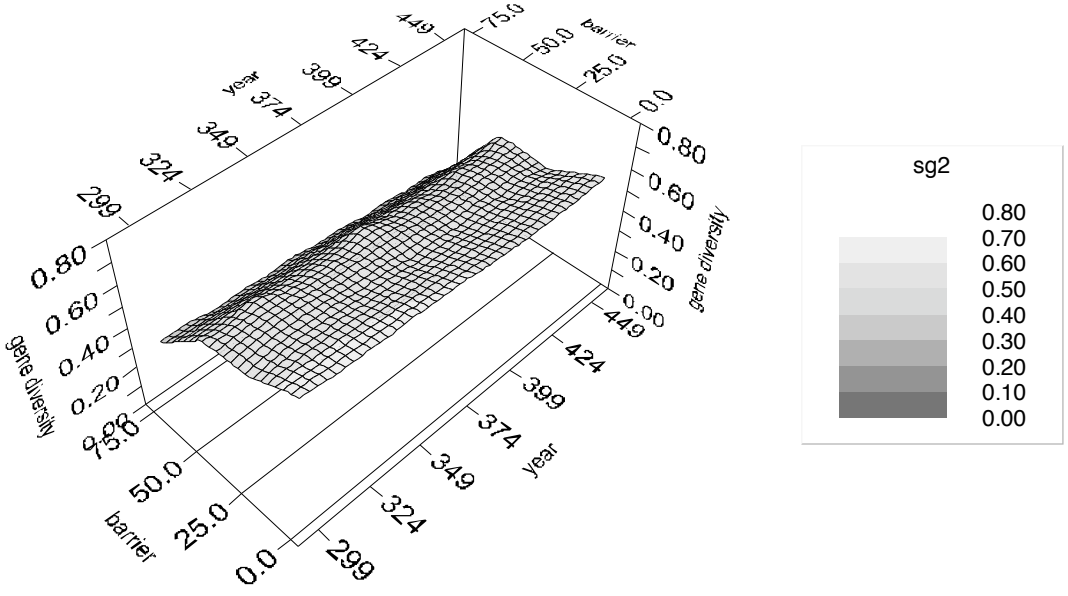
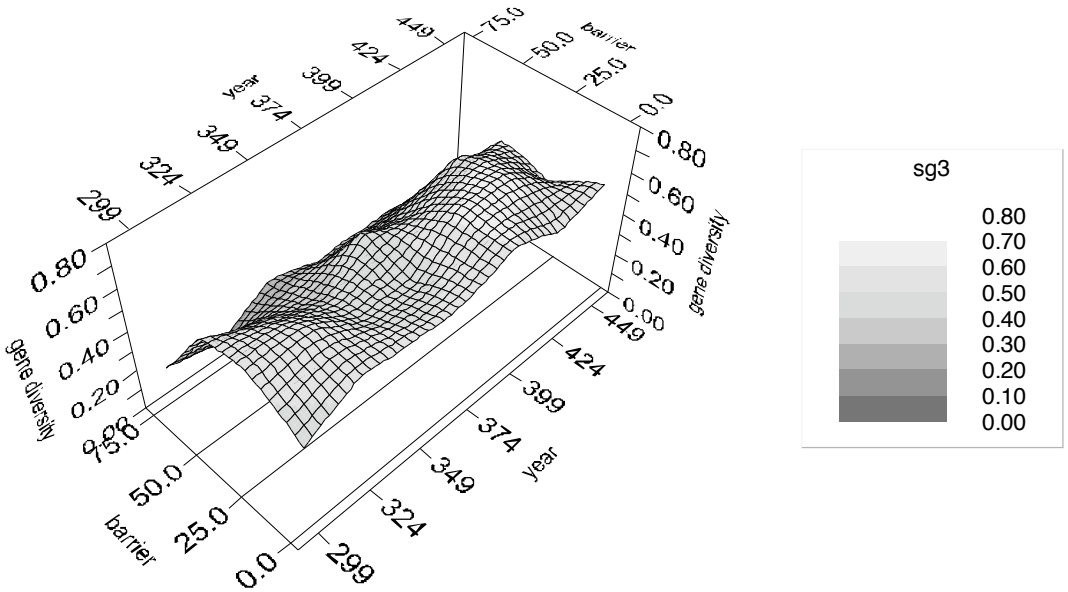


Figure 5.4 Surfaces interpolated over time (simulation years 299 – 449), barrier effect (0, 25, 50, 75%) and gene diversity: a) sg1, and b) sg2.

a)



b)

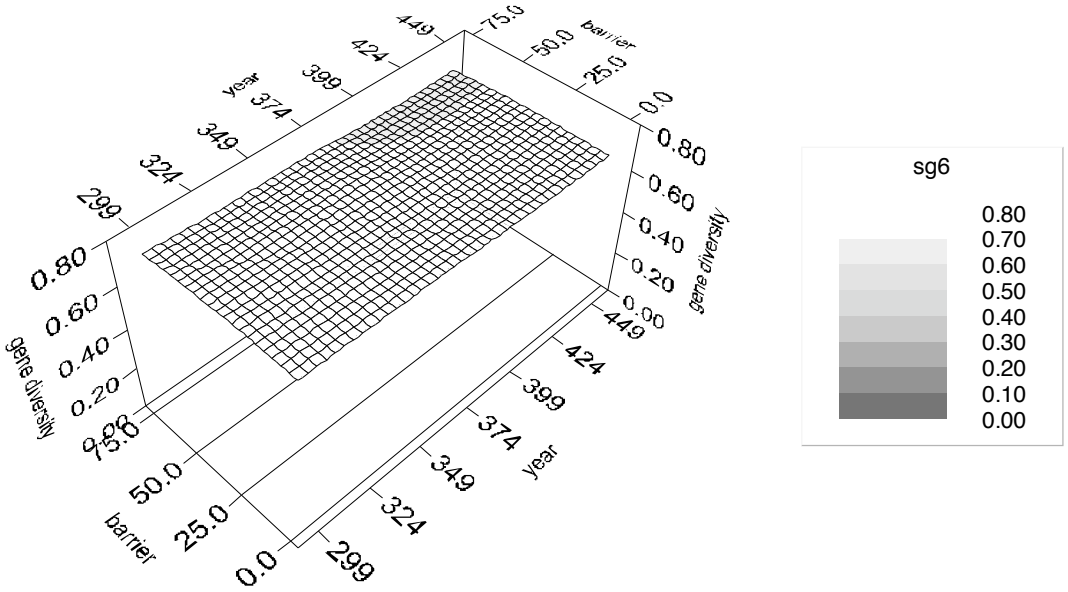


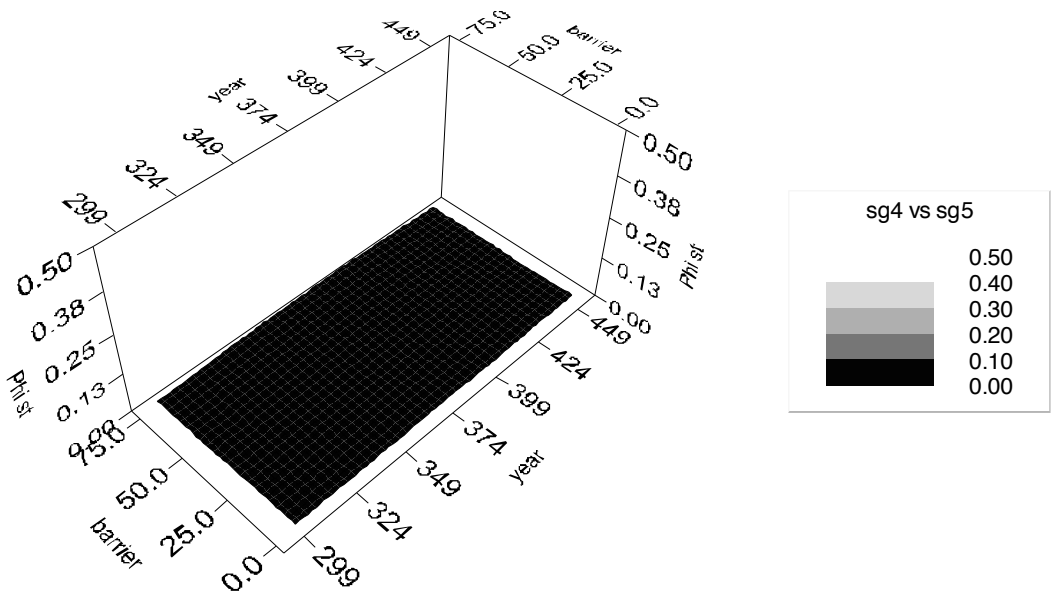
Figure 5.5 Surfaces interpolated over time (simulation years 299 – 449), barrier effect (0, 25, 50, 75%) and gene diversity: a) sg3 and b) sg6, a typical surface from a NY sg.

Three dimensional interpolated surfaces also illustrate an increasing barrier effect associated with increasing genetic differentiation, as quantified by ϕ_{ST}, for pairwise comparisons of any sg with sg2 or sg3; whereas, all other pairwise comparisons excluding sg2 or sg3 do not show a barrier effect. And with all cases, the barrier effect or non-effect is constant over time (Figures 5.6, 5.7, 5.8).

Discussion

Genetic data enabled an independent means of validating ORM raccoon demographic behaviours because these data were not used in its construction and calibration. Model behaviours occurred as expected, except for the absence of an IBD genetic population structure. Yet, IBD was also not present in the empirical data. This type of genetic population structure was expected because mating and dispersal occurs over a limited range (e.g. usually <10 km{Rosatte 2000 6 /id}, 2006), relative to the species range continuously spread over 1000's of kilometres {Zeveloff 2002 #889}). Though it appears that at the scale of this study, simulated and actual gene flow are sufficient to counter genetic drift and homogenise the population; while IBD is present at a larger spatial extent of several 1000's of kilometres (Cullingham et al. 2007). IBD may only be evident in a particular cardinal direction; however, Lake Ontario and Lake Erie confine the study area to an east-west orientation. Sub-sampling of data could enable a north-south IBD analysis; however, raccoon sample size is too small to have a sufficient number of sample groups and samples per group for assessing IBD.

a)



b)

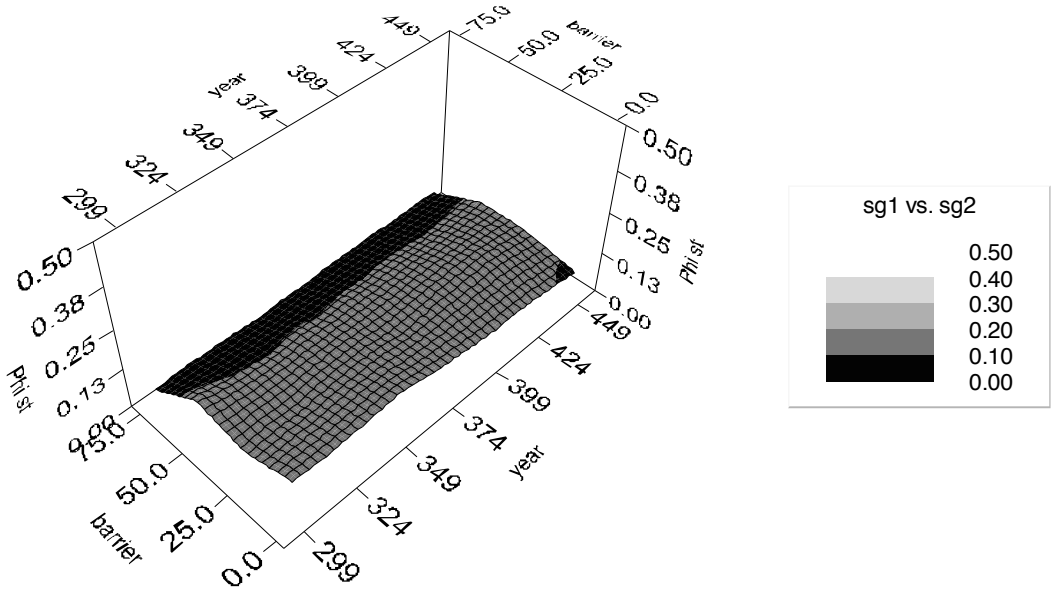
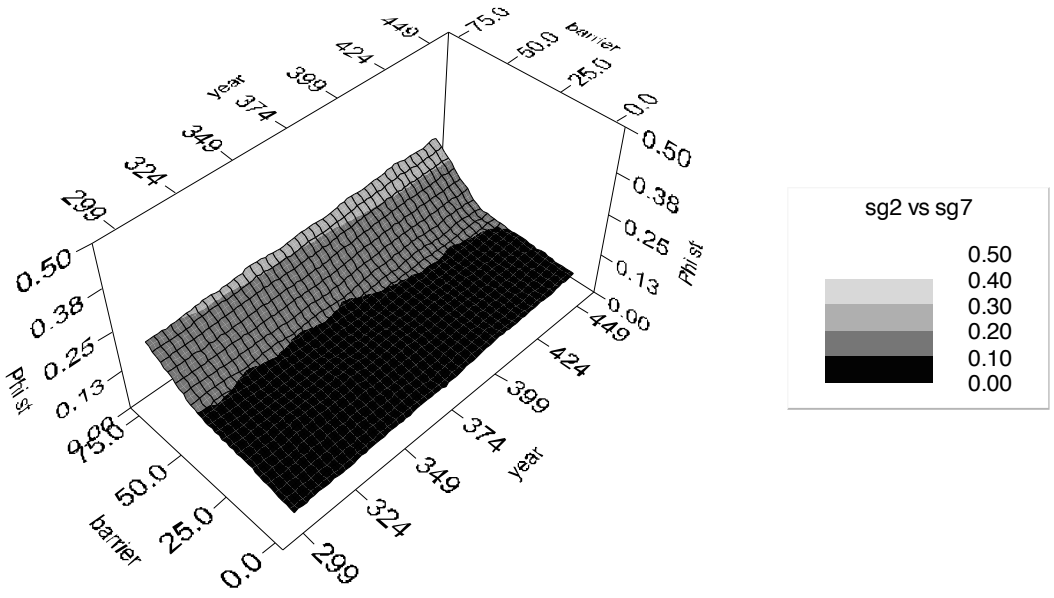


Figure 5.6 Surfaces interpolated over time (simulation years 299 – 449), barrier effect (0, 25, 50, 75%) and Φ_{ST}. a) A surface typical for pairwise comparisons on the NY side of the barrier; b) A slight barrier effect exhibited by the pairwise comparison of sg1 and sg2.

a)



b)

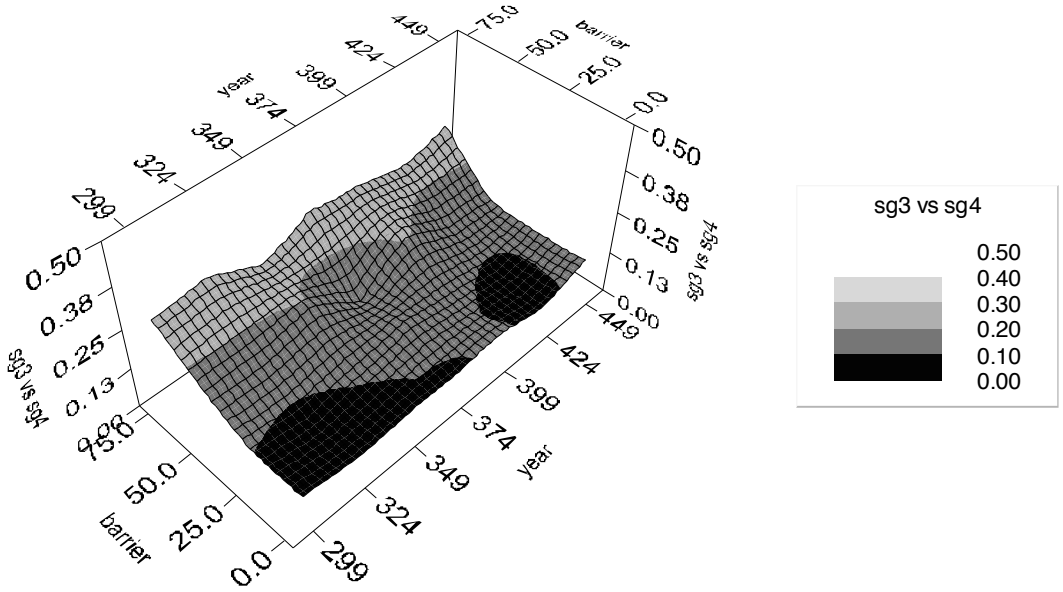


Figure 5.7 Surfaces interpolated over time (simulation years 299 – 449), barrier effect (0, 25, 50, 75%) and Φ_{ST}. Surfaces typical for pairwise comparisons of sg's separated by the barrier and showing a barrier effect for a) ON vs. NY and b) sg3 vs. NY.

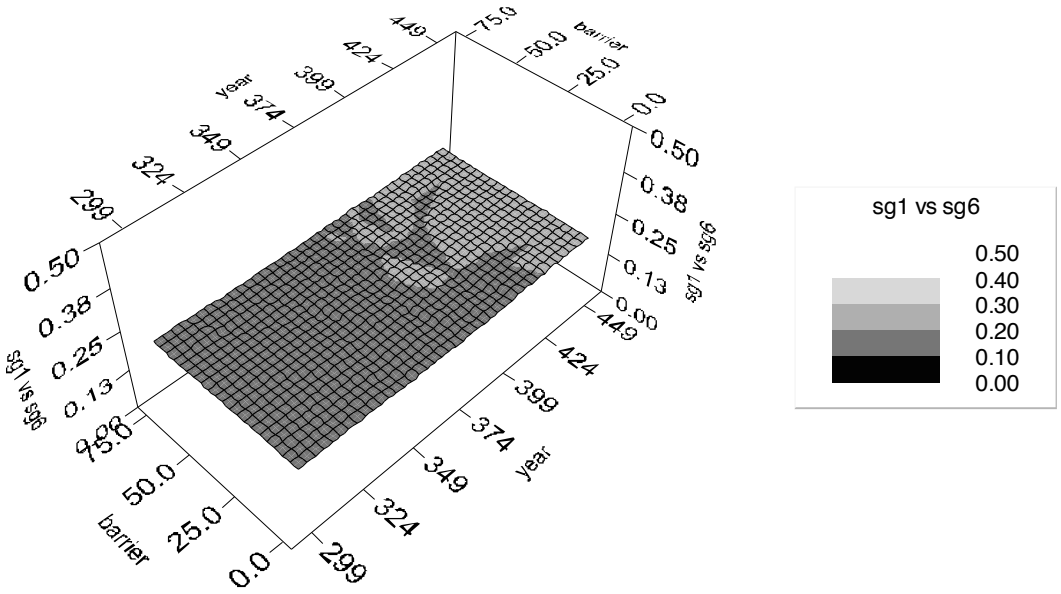


Figure 5.8 Surfaces interpolated over time (simulation years 299 – 449), barrier effect (0, 25, 50, 75%) and Φ_{ST}. A surface typical to pairwise comparisons of sg1 and sg's on the NY side of the river, showing no barrier effect.

Correlations between simulated and empirical genetic measures indicate a 50% barrier effect of the Niagara River to raccoon movement. A potential difficulty with this method is distinguishing between different barrier effects at multiple points in time when there is more than one genetic measure exhibiting a "best fit" with the empirical data. The degree of fit between modelled and simulated data depends, in part, on how well the model represents system behaviours. Model developmental techniques (e.g. sensitivity analysis) can be used to minimise uncertainty by ensuring the model has a parsimonious structure and is appropriately parameterised. Quality of the genetic data is another factor affecting the degree of fit between simulation and reality. It is important to sufficiently sample animals on either side of the landscape feature to capture the total genetic diversity and to be within the spatial scale for which a barrier affects gene flow. Furthermore, "best fit" can be determined from a consensus of a variety of genetic measures which characterise different aspects of gene flow. Also, an accurate estimate of time since colonisation (e.g. through historic records, fossil evidence) can help by determining the simulation "year" and then by default indicating the barrier effect.

Physiography is known to affect the spatio-temporal rate of infectious wildlife disease spread {Lucey, Russell, et al. 2002 148 /id}, {Smith, Lucey, et al. 2002 110 /id} {Russell, Smith, et al. 2004 280 /id}. Mountains may act as an effective barrier because animal populations are absent or too low in density for disease persistence. Rivers and valleys may either impede or facilitate disease flow depending on whether their orientation obstructs or falls in line with animal movements and interactions. The value of the proposed method is to give a relative indication of barrier strength to disease

spread, before the region becomes infected, as a form of risk assessment. By quantifying the effect of the Niagara River as a barrier to gene flow, its influence can be utilised to increase effectiveness and efficiency of rabies control programs.

Forewarning of factors affecting disease spread has obvious applications for designing effective control strategies. Landscape features can be used in conjunction with control measures to increase a barrier effect. For example, in this study the Niagara River blocks about 50% of raccoons crossing from NY into ON. This effect extends at least 30 km, but less than 60 km from the river into ON, as sg2 shows a barrier effect but sg1 does not. Thus, one strategy would be to apply control within 30 km along the length of the river to increase the barrier effect to sufficiently prevent disease from spreading into ON from NY. It then becomes necessary to determine the intensity and type of control, the width of its application, and the most effective side of the river to apply control (or both sides). These questions can be explored by modelling disease and various control strategies using the ORM.

A genetic perspective on the spread of infectious wildlife diseases is a valuable and complementary strategy to current techniques, especially because high quality genetic data have some advantages over disease incidence data. Population genetics has been thoroughly embraced as a strategy for understanding the effect of landscape barriers on animal and plant movements (Soltis et al. 2006), though its application to understanding infectious wildlife disease spread is still developing. The current study demonstrates the utility of the ORM as a genetic simulation model for exploring raccoon rabies, and its utility for exploring other disease-host systems, species invasions and colonisation events.

Chapter 6

Discussion

The motivation for my PhD research was to further develop the Ontario Rabies Model (ORM) to become a "genetic" individual based model (IBM) so that simulated and empirical genetic population structures could be compared as a novel means for understanding the spread of raccoon rabies. To achieve this goal I believed it was necessary to continue evaluating the ORM, because there were no published studies documenting its use. Even though model results may match closely with reality, this does not necessarily mean the model properly represents system behaviours; the correct result may have occurred by chance (Oreskes et al. 1994, Rykiel 1996). Validation would increase the confidence of users, beyond the ORM builders, that the ORM is a valuable tool for exploring disease-host systems. As such, my research involved using multiple modelling approaches to further assess the ORM but also had the benefit of contributing to understanding of raccoon ecology and raccoon rabies.

Model validation was also an important step in model development because the ORM is an IBM. IBM's arose in the 1970's through the advancements of computing power and programming languages (Breckling et al. 2006). The primary motivation for the ORM being an IBM was the ability to incorporate a spatial dimension and model the genetics of individuals, which was necessary for the purpose of the modelling exercise: to understand factors in the environment affecting the raccoon-rabies system. Yet, classical state variable models (SVMs) using differential equations are firmly entrenched as a valuable tool for developing ecological theory. Hence, there is an active debate as to the benefits and concerns of using IBMs (Uchmanski and Grimm 1996), (Fahse et al. 1998),

(Grimm 1999), (Lomnicki 1999, Oreskes 2003), thus, I believe using an IBM necessitated addressing their issues.

Firstly, IBMs are criticised for their high complexity (Grimm 1999). Parameter-rich models result from the "bottom-up" approach. It is important to determine whether parameters contribute more to improving the accuracy of system representations rather than increasing outcome uncertainty. The high degree of realism in IBMs makes this a challenging exercise because the impact of parameters on modelled outcomes depends on the response variables being assessed. For example, autumn dispersal is not expected to affect the spread of rabies if evaluated during the winter season when raccoons are sedentary in their dens, though autumn dispersal would be a major factor when considering rabies spread over multiple years. Therefore, when assessing ORM complexity through sensitivity analysis (SA; Chapter 4), parameter effects were quantified using a variety of response variables to consider various characteristics of the raccoon-rabies system. The most important result of SA was demonstrating that the ORM functioned as expected. This meant model complexity was "appropriate" for simulating raccoon demographics over the spatial-temporal scale required for measuring the effect of the Niagara River on gene flow (Chapter 5) - "appropriate" meaning that the model was able to explore the issue for which it was designed.

A second major criticism of IBMs is their failure to contribute to ecological theory (Grimm 1999). This is because the motivation for creating most IBMs is for pragmatic applications (e.g. comparing the rate of disease spread in two different locations); however, IBMs can contribute to ecological theory. For instance, by accepting that the ORM is an appropriate representation of the raccoon-rabies system, SA

results can be used to identify important components in the system (e.g. target cell density affecting the rate of disease spread; Grimm 1999). Additional theoretical insight is found by characterising system-level behaviours (Grimm 1999, (Breckling et al. 2006). This is demonstrated by measuring the degree of isolation by distance (IBD) in the mitochondrial genetic population structure (Chapter 5). The simulated and observed population genetic structures do not show IBD over the same spatial-temporal scale; hence, the modelled mating and dispersal systems are more likely to be realistic representations of raccoon demographics. Hypotheses testing what is disrupting the development of IBD is discussed below.

Thus, when justifying the use of ORM to explore the raccoon-rabies system it is important to be clear that IBMs and SVMs are used for different purposes, and that IBMs can still contribute to ecological theory while also being predictive tools at finer spatial-temporal scales. Furthermore, the higher degree of realism in the ORM lends it to more easily communicate model structure and results to wildlife managers. The ORM is largely mechanistic because parameters represent known processes of the system (e.g. chance of mating, chance of giving birth, litter size), as opposed to a more phenomenological approach of aggregating many processes into one parameter (e.g. growth rate). Many biologists and ecologists appreciate this design approach because they can mimic their detailed understanding of the system by creating parameters that explicitly represent system processes (Grimm 1999). Also, ORM realism makes it easier to validate model outcomes through pattern-oriented modelling. That is, by modelling known scenarios and comparing simulated results with observed results.

My research generates many important benefits, ORM and non-ORM specific. For instance, REDB was used to check that the ORM default values fell within known variation. Confirming the appropriateness of parameter values and documenting their sources is an important part of model evaluation (Bart 1995), (Conroy et al. 1995). Furthermore, the REDB provides parameter input data for parameterising the ORM in geographic regions beyond southern Ontario. The REDB contains useful data being used to explore other ecological studies (Laura Bigler, pers. comm.), and as was demonstrated by the meta-analysis in Chapter 2. Additionally, the REDB is spurring the development of other species ecological databases in the Ontario Ministry of Natural Resources (OMNR; Bruce Pond, pers. comm). The REDB data model provides a design example for constructing, populating, managing and querying these other databases. Model development in these projects will benefit by having a populated ecological database. For instance, consider the benefits if the ORM builders had used the REDB; The REDB acts as a poll among researchers as to critical aspects inferred about a system, hence, components to include during model development. Furthermore, the REDB would be an efficient tool for acquiring information to guide parameter input values and give sources for documenting their values.

As mentioned earlier, SA was most valuable at further confirming that the ORM was functioning as expected. There is also value from integrating information from the REDB and with the SA results. For example, meta-analysis of litter size and degrees latitude indicates that a significantly lower number of kits are produced at lower latitudes (Chapter 2). In this regard, it would seem appropriate to lower the litter size default value when parameterising the ORM to run scenarios at lower latitudes of the raccoon species

range. However, from the SA, litter size does not affect model outcomes (Chapter 4); thus, it is not necessary to alter the default value, or spend additional resources in the field to acquire more accurate litter size estimates.

More accurately parameterising the ORM to different geographic regions or landscapes also led me to improving raccoon density estimates, because this parameter is used to simulate habitat heterogeneity, but also because density is a critical factor in infectious disease models (Anderson and May 1991). Densities were calculated for the St. Lawrence region, but I applied the model to the Niagara region. This happened for several reasons. OMNR capture-mark-recapture data were more geographically extensive in the St. Lawrence, and this region has more variation in habitat than the Niagara region. Consequently, there were more landcover types available for testing for their effects on capture probability, estimated population size and density. However, there was a finer spatial resolution of raccoon samples processed for genetic analysis in the Niagara region. So by default, the Niagara region became the landscape to apply the ORM as a tool for measuring the effect of a landscape barrier to gene flow.

Density estimates used in the ORM for the Niagara / New York modelling exercise may have been low. Target cell density was set to 5 raccoons / km^2, as found for the St. Lawrence region (Chapter 3), but the Niagara region may support even higher densities (Rick Rosatte, pers. comm). Thus, an interesting future modelling experiment would be a sensitivity analysis of raccoon densities on barrier effects to gene flow. It is likely higher densities would necessitate a stronger barrier effect for achieving the highest correlation between simulated and empirical genetic population structures. This is because, at higher densities, a stronger barrier effect would be needed to match the

number of raccoons crossing the river and generating a genetic population structure with a diversity matching the empirical data, as occurred for a lower density and weaker barrier effect.

Another consideration is whether the density estimation technique would be applicable to other species. For this study capture-mark-recapture data were extensive, in that thousands raccoons where caught multiple times; however, large mammals often have lower population densities (e.g. 0.011 – 0.0283 bears/km^2 (Stoen et al. 2006) versus 5 – 6 raccoons/ km^2; Chapter 3), hence, have lower encounter probabilities. The ORM could be used as a tool to determine the threshold number of individuals being captured multiple times that is required for estimating density, given that the explicit movement tracking system in the ORM would enable exact calculations of density.

There are advantages and disadvantages of field and simulation modelling studies (Table 6.1). Models are useful tools for running many different kinds of experiments that would be too costly in resources or infeasible to recreate in the field. ORM simulations last several minutes to several hours, depending on the size of the landscape, length of the simulation, density of the population, number of model outputs, and number of runs specified to capture stochastic variation. A noteworthy caveat is that the ORM produces a huge volume of output files. Individual location and demographics are tracked on a weekly basis through their lifetime for all individuals in the population. Disease dynamics are also recorded, including location and time an individual becomes infected and symptomatic, and from whom the infection was acquired. Consequently, computer programs that automate the creation of response variables beyond the basic population and disease characteristics are essential. Furthermore, computer programs are often

Table 6.1 Advantages and disadvantages of field and simulation modelling studies.

Type of Study	Advantages	Disadvantages
Field	realistic test conditions	one sample
		uncontrolled
		not repeatable
		problems with field measurements
		expensive
		Slow
		limited testing scenarios
Simulation Modelling	multiple samples	appropriate definition of system
	Controlled	computer processing constraints
	Repeatable	validation of results
	complete data of results	
	Inexpensive	
	Fast	
	variety of testing scenarios	
	Predictive	

necessary to analyse model outcomes when a large number are created from running multiple input specifications.

Model stochasticity is an important benefit of the ORM even though it adds to the duration of model runs and number of outputs to analyse. In the "real world" events occur with a sample size of one because time is continually progressing. In the "model" world experiments can be repeated using identical starting conditions and stochastically defined behaviours can produce a distribution of possible outcomes. This is advantageous given uncertainty in values of parameter inputs. Hence, a likely distribution of parameter values can be defined, and input values can be randomly selected to feed model processes. It is then possible to infer "real world" variation from variation produced by multiple stochastic runs. And, if the model operates correctly, the empirical results lie within the simulated variation. Furthermore, the distribution in outcome variation produced by stochasticity enables quantifying the probability of certain scenarios (e.g. chance of spread into a new region), and for infectious-disease modelling, this is beneficial for risk assessment.

The culmination of my PhD research opens up many possible ORM studies for using neutral genetic markers to explore factors affecting genetic population structure. This approach depends on using the model to create the empirical genetic population structure. If this is achieved, then model specifications (e.g. duration of simulation) and parameterisations (e.g. barrier effects) can be used to test factors affecting the known genetic structure; for instance, testing and developing theories pertaining to the effects of gene flow and genetic drift on genetic population structure. For example, isolation by distance (IBD) is commonly used as a null hypothesis to investigate factors affecting

gene flow among populations (Crispo and Hendry 2005). Populations exhibiting IBD are assumed to be in equilibrium for the loss of unique alleles through genetic drift and the acquisition of new alleles entering the population through gene flow (Hutchinson and Templeton 1999). Field samples indicated that raccoons in the Niagara Region do not have an IBD genetic population structure (Chapter 5). The ORM can be used to test hypotheses that may be responsible for this result: barrier effects of physiographic features (e.g. Niagara River; Slatkin 1994), insufficient time since colonisation (Crispo and Hendry 2005), panmixia (Hutchinson and Templeton 1999) or translocation of animals or raccoon dispersal and mating sufficient to homogenise the genetic population structure (Slatkin 1994, Hutchinson and Templeton 1999).

My research has increased value of the ORM as an additional tool, to field and laboratory studies, for increasing understanding of the raccoon-rabies system and advising policy decisions. It is true that I did not run rabies in an application of the ORM, except for SA. However, there are multiple factors influencing the spread of raccoon rabies in North America as evidenced by variable spatial-temporal patterns of the disease front. Implicated variables include disease-host biology and ecology, land cover variation and landscape barriers, and anthropogenic factors such as translocated animals and rabies control tactics. It is through the understanding of many factors, for which the ORM is one available tool, that knowledge is increased about the raccoon-rabies system. Thus, my PhD research contributes one part to this greater endeavour.

There are many perspectives for defining and analysing a scientific question. The assumptions of each perspective generate uncertainty in understanding. The title of my thesis "Approaches to Modelling Raccoon Rabies" represents my philosophy to be

perceptive to all outcome possibilities by investigating questions from a variety of angles. IBMs are a complementary tool to SVMs for understanding mechanisms generating system level processes. Stochastic models account for input uncertainty by defining their values as a distribution, and account for outcome uncertainty by producing a variation in results. Furthermore, it is important to use measures and analytical methods that appropriately characterise and assess system behaviours, given the purpose of the investigation (e.g. multiple Information Criteria in sensitivity testing; Chapter 4), because a single method may not be analytically sufficient. Therefore, a collaboration of techniques strengthen understanding through consensus.

Appendix

The Ontario Rabies Model (ORM) is an individual-based spatially explicit model that simulates raccoon demographics, maternal and bi-parental inheritance of genetic markers, raccoon rabies disease transmission, and infectious disease control methods (e.g. depopulation, vaccination, fertility control). The ORM is written in Visual Basic following an object-oriented structure that enables it to simulate other disease-host systems. Values for model processes are stochastically determined and operate over a user-defined size and configuration of hexagonal cells arranged in a lattice. Users interface with the ORM using ESRI ArcMap (*www.esri.com*) or Microsoft Excel. Model output is written as text, XML (extensible markup language) or Microsoft Access database files. Refer to Tinline et al. (2007) for a detailed description of the ORM.

The ORM operates following a rule-based framework where all processes operate at the individual level. Values defining model behaviours are pre-determined from a probability or randomly drawn from probability distribution functions (pdf's) characterising behaviour (Table A.1). Parameter values were derived from Ontario field data (e.g. Rosatte 2000).

Model processes occur over the course of a 52-week year using a temporal resolution of one week. For example, every week a raccoon has a probability of dying as defined by its sex and age class (P6, P7). Default ORM P6 and P7 input values are equal. Dispersal is discriminated by sex and age classes. It is permitted once per year for specific time periods and distances (P14 – P16, and P1 – P4, respectively). Mating occurs at week 9, such that mates are randomly selected within a cell. Mate selection is required for defining genetics of bi-parentally inherited markers.

Table A.1. A description of ORM parameters and their default values. *Parameter values defined by a probability distribution function.

Parameter	Description	ORM default
P1	Juvenile/adult male dispersal distance*	1.36 km
P2	Juvenile/adult female dispersal distance*	0.68 km
P3	Young-of-year male dispersal distance*	2.12 km
P4	Young-of-year female dispersal distance*	1.02 km
P5	Average litter size	4
P6 (P7)	Year 0 male (female) mortality rate	0.6
P6 (P7)	Year 1 male (female) mortality rate	0.4
P6 (P7)	Year 2 male (female) mortality rate	0.3
P6 (P7)	Year 3 male (female) mortality rate	0.3
P6 (P7)	Year 4 male (female) mortality rate	0.3
P6 (P7)	Year 5 male (female) mortality rate	0.6
P6 (P7)	Year 6 male (female) mortality rate	0.6
P6 (P7)	Year 7 male (female) mortality rate	0.6
P8	Density dependent mortality control	0.2
P9	Age of independence	20 weeks
P10	Juvenile birth rate	60%
P11	Adult birth rate	95%
P12	Birth week	calendar week 18 (end of April)
P13	Litter size variance	1
P14	Male juvenile/adult movement weeks	calendar weeks: 8-43 (spring, summer, autumn)
P15	Female juvenile/adult movement weeks	calendar weeks: 12-17, 38-43 (spring, autumn)
P16	Male young-of-year movement weeks	calendar weeks: 38-43 (autumn)
P16	Female young-of-year movement weeks	calendar weeks: 38-43 (autumn)
P17	Target cell population density	5 raccoons / km^2
P19	Incubation period*	6.47 weeks
P20	Contact rate (contact within cell)	77.80%
P21	Time of infection	user-defined week and year
P22	Reflective disease spread	user-defined activation: yes/no
P23	Infectious period*	1.00 week
P24	Adult age	75 weeks
P25	Sex ratio at birth	50:50 male:female
P26	Chance of spread	1.50%

Birth week occurs during week 18. Juvenile (52 to 74 weeks) and adult females (>75 weeks) give birth with a probability defined by P10 and P11, respectively. Population size is regulated by density dependent control affecting mortality risk (P8). P8 defines the magnitude of an adjustment to mortality risk; hence P6 and P7 are increased with cell populations above the target cell carrying capacity (P17) and decreased with cell populations below P17.

Rabid simulations require users to define the year, week, cell(s) and percentage of infected animals. Rabies can be re-introduced into the model as often as desired. Duration of incubation periods is defined by P19. The ORM uses a default value of one-week for the infectious period (P23). Currently, all rabid animals die with 100% certainty. Rabies control strategies are also defined by year and week for duration and frequency and location by cell(s). Vaccination lasts for a user-defined time period with a user-defined efficacy.

Individual genetic markers are defined using a source population database before the start of a simulation. Genetic markers do not influence model behaviours or mutate to new forms. During bi-parental inheritance, offspring are equally likely to inherit one of two markers from each parent.

A lattice of hexagonal cells forms the ORM landscape. For every week of simulation, locations of each raccoon are known at the level of a cell. Each cell is internally homogeneous. Habitat heterogeneity and landscape features are defined at the level of the cell; for instance, by using P17 (e.g. agricultural areas have 5 raccoons/km^2 and water bodies have 0 raccoons/km^2) and by restricting movement among cells (e.g. 50% of raccoons are blocked from crossing large rivers).

References

Anderson, D.P., Forester, J.D., Turner, M.G., Frair, J.L., Merrill, E.H., Fortin, D., Mao, J.S., and Boyce, M.S. 2005. Factors influencing female home range sizes in elk (*Cervus elaphus*) in North American landscapes. Landscape Ecology 20:257-271.

Anderson, D.R., Burnham, K.P., and Thompson, W.L., 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64(4): 912-923.

Anderson, R.M. and R.M. May. 1985. Vaccination and herd immunity to infectious diseases. Nature 318:323-329.

Anderson,R.M. and May,R.M., 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, 757 pp.

Arnqvist, G. and Wooster, D., 1995. Meta-analysis: synthesizing research findings in ecology and evolution. Trends in Ecology and Evolution 10(6): 236-240.

Asano, M., Matoba, Y., Ikeda, T., Suzuki, M., Asakawa, M. and Ohtaishi, N., 2003. Reproductive Characteristics of the Feral Raccoon (Procyon Lotor) in Hokkaido, Japan. Journal of Veterinary Medical Science 65 (3): 369-373.

Bar-David, S., J.O. Lloyd-Smith, and W.M. Getz. 2006. Dynamics and management of infectious disease in colonizing populations. Ecology 87(5): 1215-1224.

Barlow, N.D., 1996. The ecology of wildlife disease control: simple models revisited. Journal of Applied Ecology 33: 303-314.

Bart, J. 1995. Acceptance criteria for using individual-based models to make management decisions. Ecological Applications 5(2)411-420.

Bauchau, V., 1997. Is there a file drawer problem in biological research? Oikos 79(2): 407-409.

Belant, J.L., 1992. Homing of Relocated Raccoons, Procyon-Lotor. Canadian Field-Naturalist 106(3): 382-384.

Bender, D.J., T.A. Contreras, and L. Fahrig. 1998. Habitat loss and population decline: a meta-analysis of the patch size effect. Ecology 79: 517-533.

Berec, L. 2002. Techniques of spatially explicit individual-based models: construction, simulation, and mean-field analysis. Ecological Modelling 150: 55-81.

Bigler, W. J., Hoff, G.L., and Johnson, S.A., 1981. Population characteristics of Procyon lotor marinus in estuarine mangrove swamps of southern Florida. Florida Scientist 44: 151-157.

Bilenca, D.N., Zuleta, G.A., Kravetz, F.O., 1994. Latitudinal variation in litter size and length of the breeding season in populations of pampas mice, *Akodon azarae*. Mammalia 58:563-568.

Blackburn, T.M. and K.J. Gaston. 1996. Abundance-body size relationships: the area you census tells you more. Oikos 75: 303-309.

Blower, S.M. and Dowlatabadi, H., 1994. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. Int. Stat. Rev. 62(2): 229-243.

Boulanger, J., G. Stenhouse, and R. Munro. 2004. Sources of heterogeneity bias when DNA mark-recapture sampling methods are applied to grizzly bear (*Ursus arctos*) populations. Journal of Mammalogy 85(4): 618-624.

Breckling, B., Middelhoff, U., and Reuter, H. 2006. Individual-based models as tools for ecological theory and application: Understanding the emergence of organisational properties in ecological systems. Ecological Modelling 194: 102-113.

Broadfoot, J. D., Rosatte, R. C. and O'Leary, D. T. 2001 Raccoon and skunk population models for urban disease control planning in Ontario, Canada. Ecological Applications 11: 295-303.

Buckland, S.T., D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas. 2001. Introduction to Distance Sampling: Estimating abundance of biological populations. Oxford University Press, Oxford, U.K. 432 pp.

Burnham, K.P. and Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag Inc., New York, 488 pp.

Burnham, K.P., Anderson, D.R., and Thompson, W.L. 2000. Model selection and multimodel inference: a practical information-theoretic approach. Journal of Wildlife Management 64(4): 912-923.

Cagle, F.R., 1949. Notes on the raccoon, Procyon lotor megalodus Lowery. Journal of Mammology 30: 45-47.

Caro, T.M., Shargel, J.A., and Stoner, C.J., 2000. Frequency of medium-sized mammal road kills in an agricultural landscape in California. American Midland Naturalist 144: 362-369.

Chamberlain, M.J. and Leopold, B.D., 2002. Spatio-Temporal Relationships Among Adult Raccoons (Procyon Lotor) in Central Mississippi. American Midland Naturalist 148 (2): 297-308.

Chamberlain, M.J., Conner, L.M., Leopold, B.D. and Hodges, K.M., 2003. Space Use and Multi-Scale Habitat Selection of Adult Raccoons in Central Mississippi. Journal of Wildlife Management 67(2): 334-240.

Chan,K., Saltelli,A., and Tarantola,S. 1997. Sensitivity analysis of model output: Variance-based methods make the difference. Edited by S.Andradottir, K.J.Healy, D.H.Withers, and B.L.Nelson. Environment Institute European Commission Joint Research Centre, Ispra, Italy. pp. 261-268.

Childs, J. E., Curns, A. T., Dey, M. E., Real, L. A., Feinstein, L., Bjornstad, O. N. and Krebs, J. W. 2000 Predicting the local dynamics of epizootic rabies among raccoons in the United States. Proceedings of the National Academy of Science USA 97: 13666-13671.

Childs, J. E., Curns, A. T., Dey, M. E., Real, L. A., Rupprecht, C. E. and Krebs, J. W. 2001 Rabies Epizootics Among Raccoons Vary Along a North-South Gradient in the Eastern United States. Vector Borne and Zoonotic Diseases 1: 253-267.

Cockburn, A., Lee, A.K., and Martin, R.W., 1983. Macrogeographic variation in litter size in *Antechinus* (Marsupalia: Dasyuridae). Evolution 37:86-95.

Conaway, C.H., Sadler, K.C., and Hazelwood, D.H., 1974. Geographic variation in litter size and onset of breeding in cottontails. Journal of Wildlife Management 38:473-481.

Conn, P.B., A.D. Arthur, L.A. Bailey, and G.R. Singleton. 2006. Estimating the abundance of mouse populations of known size: promises and pitfalls of new methods. Ecological Applications 16(2): 829-837.

Conner, E.F., A.C. Courtney, and J.M. Yoder. 2000. Individuals-area relationships: the relationship between animal population density and area. Ecology 81(3): 734-748.

Conroy, M.J., Cohen, Y., James, F.C., Matsinos, Y.G., and Maurer, B.A. 1995. Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. Ecological Applications 5(1): 17-19.

Cooke, B.D., A.J. Robinson, J.C. Merchant, A. Nardin, and L. Capucci. 2000. Use of ELISAs in field studies of rabbit haemorrhagic disease (RHD) in Australia. Epidemiology and Infection 124(3): 563-576.

Corwell, R. K. 2005 EstimateS: Statistical estimation of species richness and shared species from samples. Version 7.5. User's Guide and application published at: http://purl. oclc. org/estimates

Cowan, W.F., 1973. Ecology and life history of the raccoon (Procyon lotor hirtus Nelson and Goldman) in the northern part of its range. University of North Dakota. PhD Dissertation. University of North Dakota, Grand Forks.

Coyne, M.J., Smith, G., and McAllister, F.E., 1989. Mathematic model for the population biology of rabies in raccoons in the mid-Atlantic states. American Journal of Veterinary Research 50(12): 2148-2154.

Crispo, E., and Hendry, A., 2005. Does time since colonization influence isolation by distance? A meta-analysis. Conservation Genetics 6: 665-682.

Cullingham C.I., Kyle C.J., Pond B.A., and White B.N. 2007. Examination of subspecific designations of raccoons (Procyon lotor) using mitochondrial DNA control sequences. Journal of Mammalogy (in review)

Dice, L.R. 1938. Some census methods for mammals. Journal of Wildlife Management 2(3): 119-130.

Dorney, R.S., 1954. Ecology of marsh raccoons. Journal of Wildlife Management 18: 217-225.

Efford, M. 2004. Density estimation in live-trapping studies. Oikos 106: 598-610.

Efford, M., D.K. Dawson, and C.S. Robbins. 2004. DENSITY: software for analysing capture-recapture data from passive detector arrays. Animal Biodiversity and Conservation 27(1): 217-228.

Endres, K.M. and W.P. Smith. 1993. Influence of Age, Sex, Season and Availability on Den Selection by Raccoons Within the Central Basin of Tennessee. American Midland Naturalist 129(1): 116-131.

Erb, J.D., and Boyce, M.S. 1999. Distribution of population declines in large mammals. Conservation Biology 13:199-201.

Excoffier, L., Laval, G. and Schneider, S. 2005 Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online

Fa, J., Sharples, C., Bell, D., and DeAngelis, D., 2001. An individual-based model of rabbit viral haemorrhagic disease in European wild rabbits (*Oryctolagus cuniculus*). Ecological Modelling 144: 121-138.

Fahse, L., Wissel, C., and Grimm, V.,1998. Reconciling classical and individual-based approaches in theoretical population ecology: a protocol for extracting population parameters from individual-based models. The American Naturalist 152(6): 838-852.

Ferguson, N.M., Keeling, M.J., Edmunds, W. J., Gani, R., Grenfell, B.T., Anderson, R.M., and Leach, S., 2003. Planning for smallpox outbreaks. Nature 425: 681-685.

Ferguson, N.M., May, R., and Anderson, R.,. 1997. Chapter 5: Measles: Persistence and synchronicity in disease dynamics. In: Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions, D. Tilman, P. Kareiva (Editors), Princeton University Press, Princeton, New Jersey, USA. pp. 137-157.

Fiero, B.C. and Verts, B.J., 1986. Age-Specific Reproduction in Raccoons in Northwestern Oregon. Journal of Mammalogy 67(1): 169-172.

Fritzell, E. K., 1978b. Aspects of raccoon (Procyon lotor) social organization. Canadian Journal of Zoology 56: 260-271.

Fritzell, E.K., 1978a. Reproduction of raccoons (Procyon Lotor) in North Dakota. The American Midland Naturalist 100(1): 253-256.

Gachet, S., Vela, E., and Tatoni, T., 2005. BASECO: a floristic and ecological database of Mediterranean French flora. Biodiversity and Conservation 14: 1023-1034.

Gaston, K.J. and Matter, S., 2002. Individuals-area relationships: Comment. Ecology 83(1): 288-293.

Gaston, K.J., T.M. Blackburn, and B.T. Grenfell. 1999. Does variation in census area confound density comparisons? Journal of Applied Ecology 36: 191-204.

Gates, S. 2002. Review methodology of quantitative reviews using meta-analysis in ecology. Journal of Animal Ecology 71: 547-557.

Gehrt, S.D. 2002. Evaluation of Spotlight and Road-Kill Surveys as Indicators of Local Raccoon Abundance. Wildlife Society Bulletin 30(2): 449-456.

Gehrt, S.D., 2003. Raccoon: Procyon lotor and allies. Wild mammals of North America: Biology, Management and Conservation. The John Hopkins University Press, Baltimore, pp. 611-634.

Gehrt, S.L.F. 2004. Spatial patterns and dynamic interactions among raccoons in Eastern Kansas. The Southwestern Naturalist 49(1): 116-121.

Gehrt, S. D. and E. K. Fritzell., 1996. Sex-Biased Response of Raccoons (Procyon Lotor) to Live Traps. American Midland Naturalist 135(1): 23-32.

Gehrt, S. D. and E. K. Fritzell., 1997. Sexual Differences in Home Ranges of Raccoons. Journal of Mammalogy 78(3): 921-931.

Gehrt, S.D. and E.K. Fritzell. 1998. Duration of Familial Bonds and Dispersal Patterns for Raccoons in South Texas. Journal of Mammalogy 79(3): 859-872.

Gehrt, S.D. and Fox, L.B., 2004. Spatial Patterns and Dynamic Interactions Among Raccoons in Eastern Kansas. Southwestern Naturalist 49(1): 116-121.

Ginot, V., Gaba, S., Beaudouin, R., Aries, F., and Monod, H., 2006. Combined use of local and ANOVA-based global sensitivity analyses for the investigation of a stochastic dynamic model: application to the case study of an individual-based model of a fish population. Ecological Modelling 193: 479-491.

Grigione, M.M., Beier, P., Hopkins, R.A., Neal, D., Padley, W.D., Schonewald, C.M., and Johnson, M.L. 2002. Ecological and allometric determinants of home-range size for mountain lions (*Puma concolor*). Animal Conservation 5:317-324.

Grimm, V. 1999. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecological Modelling 115: 129-148.

Gurevitch, J. and Hedges, L.V., 1999. Statistical issues in ecological meta-analyses. Ecology 80(4): 1142-1149.

Guthery, F.S., Brennan, L.A., Peterson, M.J., and Lusk, J.J., 2005. Information theory in wildlife science: critique and viewpoint. Jounral of Wildlife Management 69(2): 457-465.

Hable, C., Hamir, A., Snyder, D., Joyner, R., French, J., Nettles, V., and Hanlon, C., 1992. Prerequisites for oral immunization of free-ranging raccoons (Procyon lotor) with a recombinant rabies virus vaccine: study site ecology and bait system development. Journal of Wildlife Diseases 28(1): 64-79.

Hall, T. A. 1999 BioEdit:a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Ser 41, 95-98.

Hamby, D. M. 1995. A comparison of sensitivity analysis techniques. Health Physics Society 68(2): 195-204.

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring Assessment 32, 135-154.

Hansson, L. 1969. Home range, population structure and density estimates of removal catches with edge effect. Acta Theriologica 14:153-160.

Harrison, R. G. 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution , 6-11.

Hartman, L.H. and Eastman, D.S., 1999. Distribution of Introduced Raccoons (*Procyon Lotor*) on the Queen Charlotte Islands: Implications for Burrow-Nesting Seabirds. Biological Conservation 88(1): 1-13.

Hartman, L.H., A.J. Gaston and D.S. Eastman. 1997. Raccoon Predation on Ancient Murrelets on East Limestone Island, British Columbia. Journal of Wildlife Management 61(2): 377-388.

Hellmann, J. J. and Fowler, G. W. 1999 Bias, precision, and accuracy of four measures of species richness. Ecological Applications 9: 824-834.

Hengeveld, R., and Haeck, J. 1982. The distribution of abundance. I. Measurements. Journal of Biogeography 9:303-316

Henning, J., D.U. Pfeiffer, P.R. Davies, J. Meers, and R.S. Morris. 2006. Temporal dynamics of rabbit haemorrhagic disease virus infection in a low-density population of wild rabbits (*Oryctolagus cuniculus*) in New Zealand. Wildlife Research 33(4): 293-303.

Higgins, J.P.T., Green ,S., editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 (updated September 2006). http://www.cochrane.org/resources/handbook/hbook.htm

Hodges, K. M., Chamberlain, M.J., and Leopold, B.D., 2000. Effect of summer hunting on ranging behavior of adult raccoons in central Mississippi. Journal of Wildlife Management 64(1): 194-198.

Hoffmann, C.O. and Gottschang, J.L., 1977. Numbers, distribution, and movements of a raccoon population in suburban residential community. Journal of Mammalogy 58: 623-636.

Hutchinson, D. W. and Templeton, A. R. 1999 Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53: 1898-1914.

Innes, D.G.L., 1978. A reexamination of litter size in some North American microtines. Canadian Journal of Zoology 56:1488-1496.

Jackson, A. C. 2002. Pathogenesis *In* Rabies, 1st edition. Jackson, A. C. and Wunner, W. H. (Academic Press, San Diego), pp. 245-282.

Johnson, D., 1999. The insignificance of statistical significance testing. Journal of Wildlife Management 63(3): 763-772.

Jones, M.B., Schildhauer, M.P., Reichmann, O.J., and Bowers, S. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Reviews of Ecological Systems 37:519-544.

Junge, R.E. and Sanderson, G.C., 1982. Age-Related Reproductive Success of Female Raccoons. Journal of Wildlife Management 46(2): 527-529.

Kamler, J.F. and Gipson, P.S., 2003. Space and Habitat Use by Male and Female Raccoons, *Procyon Lotor*, in Kansas. Canadian Field-Naturalist 117(2): 218-223.

Kaufmann, J. H. 1982., Raccoon and allies *In* Wild mammals of North America: Biology, management, and economics, J. A. Chapman and G. A. Felhamer, John Hopkins University Press, Baltimore, pp. 567-585.

Kennedy, M.L., G.D. Baumgardner, M.E. Cope, F.R. Tabatabai and O.S. Fuller. 1986. Raccoon (Procyon-Lotor) Density as Estimated by the Census-Assessment Line Technique. Journal of Mammalogy 67(1): 166-168.

Kermack, W.O. and A.G. McKendrick. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London A 138: 55-83.

Kissell, R.E. and Kennedy, M.L., 1992. Ecologic Relationships of Co-occurring Populations of Opossums (*Didelphis-Virginiana*) and Raccoons (*Procyon-Lotor*) in Tennessee. Journal of Mammalogy 73(4): 808-813.

Kotiaho, J.S., 2002. Meta-analysis, can it ever fail? Oikos 96(3): 551-553.

Kuha, J. 2004. AIC and BIC: comparisons of assumptions and performance. Sociological Methods and Research 33(2): 188-229.

Lawson, A. B. 2001. Statistical Methods in Spatial Epidemiology. John Wiley and Sons, LTD, Chichester, 277 pp.

Leberg, P.L., and Kennedy, M.L., 1988. Demography and habitat relationships of raccoons in Western Tennessee. Proceedings Annuals Conference Southeastern Association of Fish and Wildlife Agencies 42: 272-282.

Leung, B. and Grenfell, B.T., 2003. A spatial stochastic model simulating a scabies epidemic and coyote population dynamics. Ecological Modelling 166: 41-52

Lipsey, M.W., and Wilson, D.B., 2001. Practical Meta-Analysis. Sage, Thousand Oaks, California, 247 pp.

Lomnicki, A. 1999. Individual-based models and the individual-based approach to population ecology. Ecological Modelling 115:191-198.

Lucey, B. T., Russell, C. A., Smith, D. L., Wilson, M. L., Long, A., Waller, L. A., Childs, J. E. and Real, L. A. 2002 Spatiotemporal analysis of epizootic raccoon rabies propagation in Connecticut, 1991-1995. Vector Borne and Zoonotic Diseases 2: 77-86.

Lynch, G., 1967. Long-range movement of a raccoon in Manitoba. Journal of Mammalogy 48(4): 659-660.

MacArthur, R.H. and Wilson, E.O. 1967. The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA, 224 pp.

MacArthur, R.H. and Wilson, E.O. 1972. Density compensation in island faunas. Ecology 53: 330-342.

Manly, B.F.J., 1994. Multivariate statistical methods: A primer. 2nd Ed. Chapman and Hall, NY, 215 pp.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Matter, S.F. 2000. The importance of the relationship between population density and habitat area. Oikos 89(3): 613-619.

Mayor, S.J. and Schaefer, J.A. 2005. The many faces of population density. Oecologia 145: 276-281.

McCallum, H. 2000. Population Parameters: Estimation for Ecological Models. Blackwell Science Ltd., Oxford, U.K., 348 pp.

McCleery, R.A., Foster, G.W., Lopez, R.R., Peterson, M.J., Forrester, D.J. and Silvy, N.J., 2005. Survey of Raccoons on Key Largo, Florida, USA, for Baylisascaris procyonis. Journal of Wildlife Diseases 41(1): 250-252.

McKay, M.D., Conover, W.J., and Beckman, R.J., 1979. The epidemiology and transmission dynamics of HIV/AIDS. Technometrics 21: 239-245.

McKeever, S., 1958. Reproduction in the raccoon in the southwestern United States. Journal of Wildlife Management 22: 211.

Mech, L.D., and Turkowski, F.J., 1966. Twenty-three raccoons in one winter den. Journal of Mammalogy 47: 529-530.

Michener, W.K., 2006. Meta-information concepts for ecological data management. Ecological Informatics 1: 3-7.

Michener, W.K., Brunt, J.S., Helly, J., Kirchner, T.B., and Stafford, S., 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 7: 330-342.

Moore, D. A. 1999 Spatial diffusion of raccoon rabies in Pennsylvania, USA. Preventative Veterinary Medicine, 19-32.

Mosillo, M., Heske, E. J., and Thompson, J.D., 1999. Survival and Movements of Translocated Raccoons in Northcentral Illinois. Journal of Wildlife Management 63(1): 278-286.

Murray, M. 2002. Determinants of cluster distribution in the molecular epidemiology of tuberculosis. Proceedings of the National Academy of Science USA 99(3): 1538-1543.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA, 512 pp.

Nei, M. 1977 F-statistics and analysis of gene diversity in sub-divided populations. The Annals of Human Genetics 41: 225-233.

Nottingham, B.G., Johnson, K.G., Woods, J.W., and Pelton M.R., 1982. Population characteristics and harvest relationships of a raccoon population in east Tennessee. Proceedings Annual Conference Southeastern Association of Fish and Wildlife Agencies 36: 691-700.

Onstad, D.W., 2007. EDWIP: Ecological Database of the World's Insect Pathogens. Champaign, Illinois: Illinois Natural History Survey. http://insectweb.inhs.uiuc.edu/Pathogens/EDWIP

Oreskes, N. 2003. The Role of Quantitative Models in Science. in The Role of Quantitative Models in Science. pp. 13-31.

Oreskes, N., Frechette-Shrader, K., and Belitz K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263: 641-646.

Otis, D.L., K.P. Burnham, G.C. White, and D.R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62: 1-135.

Parmenter, R.R., T.L. Yates, D.R. Anderson, M.A. Burgman, J.L. Dunnun, A.B. Franklin, M.T., Friggens, B.C. Lubow, M. Miller, G.S. Olson, C.A. Parmenter, J. Pollard, E. Rexstad, T.M. Shenk, T.R. Stanley, and G.C. White. 2003. Small-mammal density estimation: A field comparison of grid-based vs. web-based density estimators. Ecological Monographs 73(1): 1-26.

Pedlar, J.H., L. Fahrig and H.G. Merriam. 1997. Racoon habitat use at 2 spatial scales. Journal of Wildlife Management 61(1): 102-112.

Perry, B.D., Garner, N., Jenkins, S.R., McCloskey, K.M., and Johnston, D.H., 1989. A study of techniques for the distribution of oral rabies vaccine to wild raccoon populations. Journal of Wildlife Diseases 25(2): 206-217.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics 56: 434-442.

Prairie, Y.T. and D.F. Bird. 1989. Some misconceptions about spurious correlation problem in the ecological literature. Oecologia 82: 285-288.

Prange, S., Gehrt, S.D., and Wiggers, E.P., 2004. Influences of Anthropogenic Resources on Raccoon (Procyon Lotor) Movements and Spatial Distribution. Journal of Mammalogy 85(3): 4834-90.

Pullin, A.S., Knight, T.M., Stone, D.A., and Charman, K., 2004. Do conservation managers use scientific evidence to support their decision-making? Biological Conservation 119: 245-252.

R Development Core Team, 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

R Development Core Team. 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rademaker V., and Cerqueira, R., 2006. Variation in the latitudinal reproductive patterns of the genus *Didelphis* (Didelphimorphia: Didelphidae). Austral Ecology 31:337-342.

Raftery, A.E., Madigan, D., and Hoeting, J.A. 1997. Bayesian Model Averaging for Linear Regression Models. Journal of the American Statistical Association 92: 179-191.

Ratnaswamy, M. J., Warren, R.J., Kramer, M.T., and Adam, M.D., 1997. Comparisons of Lethal and Nonlethal Techniques to Reduce Raccoon Depredation of Sea Turtle Nests. Journal of Wildlife Management 61(2): 368-276.

Ratnayeke, S., Tuskan, G.A., and Pelton, M.R., 2002. Genetic Relatedness and Female Spatial Organization in a Solitary Carnivore, the Raccoon, *Procyon Lotor*. Molecular Ecology 11(6): 1115-1124.

Real, L.A., Henderson, C.J., Biek, R., Snaman J., Lambert Jack, T., Childs, J.E, Stahl, E., Waller, L.A., Tinline, R.R., and Nadin-Davis, S.A. 2005. Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proceedings of the National Academy of Science USA 102(34): 12107-12111.

Reineking, B. and Schroder, B., 2006. Constrain to perform: Regularization of habitat models. Ecological Modelling 193: 675-690.

Riley, S.P.D., Hadidian, J., and Manski, D.A., 1998. Population Density, Survival, and Rabies in Raccoons in an Urban National Park. Canadian Journal of Zoology-Revue Canadienne De Zoologie 76(6): 1153-1164.

Ritke, M.E., 1990a. Quantitative Assessment of Variation in Litter Size of the Raccoon *Procyon Lotor*. American Midland Naturalist 123(2): 390-398.

Ritke, M.E., 1990b. Sexual Dimorphism in the Raccoon (Procyon-Lotor) - Morphological Evidence for Intrasexual Selection. American Midland Naturalist 124(2): 342-351.

Rivest, P. and Bergeron, J.M., 1981. Density, Food-Habits, and Economic Importance of Raccoons (Procyon-Lotor) in Quebec Agrosystems. Canadian Journal of Zoology-Revue Canadienne De Zoologie 59(9): 1755-1762.

Roberts, P.D., Stewart, G.B., and Pullin, A.S., 2006. Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. Biological Conservation 132: 406-423.

Robinson, D.H. and Wainer, H., 2002. On the past and future of null hypothesis significance testing. Journal of Wildlife Management 66(2): 263-271.

Root, R.B. 1973. Organization of a plant-arthropod association in simple and diverse habitats: the fauna of collards (*Brassica oleraceae*). Ecological Monographs 45: 95-120.

Rosatte, R., 2000. Management of raccoons (*Procyon lotor*) in Ontario, Canada: Do human intervention and disease have significant impact on raccoon populations? Mammalia 64(4): 369-390.

Rosatte, R., Donovan, D., Allan, M., Howes, L., Silver, A., Bennett, K., MacInnes, C., Davies, C., Wandeler, A., and Radford, B., 2001. Emergency response to raccoon rabies introduction into Ontario. Journal of Wildlife Diseases 37(2): 265-279.

Rosatte, R., MacDonald, E., Sobey, K., Donovan, D., Bruce, L., Allan, M., Silver, A., Bennett, K., Brown, L., MacDonald, K., Gibson, M., Buchanan, T., Stevenson, B., Davies, C., Wandeler, A., and Muldoon, F., 2007a. The elimination of raccoon rabies from Wolfe Island, Ontario: Animal density and movements. Journal of Wildlife Diseases 43: 242-250.

Rosatte, R. and MacInnes, C., 1989. Relocation of city raccoons. Ninth Great Plains Wildlife Damage Control Workshop Proceedings: USDA Forest Service General Technical Report: Great Plains Agricultural Council, 87-92.

Rosatte, R., MacInnes, C., Williams, R. and Williams, O. 1997. A proactive prevention strategy for raccoon rabies in Ontario, Canada. Wildlife Society Bulletin 25: 110-116.

Rosatte, R., Power, M., and MacInnes, C., 1991. Ecology of urban skunks, raccoons, and foxes in Metropolitan Toronto *In* Wildlife Conservation in Metropolitan Environments,

L. Adams and D. Leedy (editors). National Institute for Urban Wildlife, Columbus, Maryland, USA pp. 31-38.

Rosatte, R., Power, M., and MacInnes, C., 1992. Density, Dispersion, Movements And Habitat Of Skunks (Mephitis Mephitis) And Raccoons (*Procyon Lotor*) In Metropolitan Toronto. Wildlife 2001: Populations. Elsevier Applied Science, London, pp. 932-944.

Rosatte, R., Sobey, K., Donovan, D., Allan, M., Bruce, L., Buchanan, T., and Davies, C., 2007b. Raccoon density and movements in areas where population reduction programs were implemented to control rabies. Journal of Wildlife Management 71(7): (in press).

Rosatte, R., Sobey, K., Donovan, D., Bruce, L., Allan, M., Silver, A., Bennett, K., Gibson, M., Simpson, H., Davier, C., Wandeler, A., and Muldoon, F., 2006. Behavior, movements, and demographics of rabid raccoons in Ontario, Canada: management implications. Journal of Wildlife Diseases 42: 589-605.

Rosatte, R., Tinline, R., and Johnston. 2007c. Rabies control in wild carnivores *In* Rabies. A. Jackson and W. Wanner (editors) 2nd edition. Academic Press. San Diego, California, USA.

Roscoe, D. E., Holste, W.C., Sorhage, F.E., Campbell, C., Niezgoda, M., Buchanaa, R., Diehl, D., Niu, H.S. and Rupprecht, C.T., 1998. Efficacy of an Oral Vaccinia-Rabies Glycoprotein Recombinant Vaccine in Controlling Epidemic Raccoon Rabies in New Jersey. Journal of Wildlife Diseases 34(4): 752-763.

Rupprecht, C. E. and Smith, J. S. 1994 Raccoon rabies: the re-emergence of an epizootic in a densely populated area. Virology 5: 155-164.

Russell, C. A., Smith, D. L., Waller, L. A., Childs, J. E. and Real, L. A. 2004 A priori prediction of disease invasion dynamics in a novel environment. Proceedings of the Royal Society of London B Biological Sciences 271: 21-25.

Rykiel, E.J.Jr. 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90: 229-244.

Saltelli, A. 2000. What is sensitivity analysis? Pages 3-13 in: A. Saltelli, K. Chan and E. Marian Scott, editors. Sensitivity Analysis. John Wiley & Sons, Ltd., Chichester, U.K.

Sanchez, M.A. and Blower, S.M., 1997. Uncertainty and sensitivity analysis of the basic reproductive rate. American Journal of Epidemiology 145(12): 1127-1137.

Sanderson, G.C. 1987. Raccoon. Pages 486-499 in: M. Novak, J.A. Baker, M.E. Obbard and B. Malloch, editors. Wild furbearer management and conservation in North America. Ministry of Natural Resources, Peterborough, Ontario, Canada.

Sanderson, G.C., 1950. Methods of measuring productivity in raccoons. Journal of Wildlife Management 14: 389-402.

Scheffer, V.B., 1950. Notes on the raccoon in southwest Washington. Journal of Mammalogy 31: 444-448.

Schneider, D., Mech, L., and Tester, J., 1971. Movements of female raccoons and their young as determined by radio-tracking. Animal Behaviour Monographs 4(1): 1-43.

Schubert, C. A., Barker, I.K., Rosatte, R.C., MacInnes, C.D., and Nudds, T.D., 1998. Effect of Canine Distemper on an Urban Raccoon Population: an Experiment. Ecological Applications 8(2): 379-387.

Seidensticker, J., Johnsingh, A.J.T., Ross, R., Sanders, G. and Webb, M.B., 1988. Raccoons and Rabies in Appalachian Mountain Hollows. National Geographic Research 4(3): 359-370.

Simberloff, D., 1988. The contribution of population and community biology to conservation science. Annual Review of Ecological Systems 19: 473-511.

Slatkin, M., 1994. Gene flow and population structure *In*: Ecological Genetics. L. Real (editor), Princeton University Press, Princeton, NJ, USA, pp. 3 - 17.

Slate, D., Rupprecht, C. E., Rooney, J. A., Donovan, D., Lein, D. H. and Chipman, R. B. 2005 Status of oral rabies vaccination in wild carnivores in the United States. Virus Research 111: 68-76.

Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E. and Real, L. A. 2002. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. Proceedings of the National Academy of Science USA 99, 3668-3672.

Smith, G.C. and Cheeseman, C.L.A., 2002. Mathematical model for the control of diseases in wildlife populations: culling, vaccination, and fertility control. Ecol. Modelling 150: 45-53.

Smith, H. T. and Engeman, R.M., 2002. An Extraordinary Raccoon, *Procyon Lotor*, Density at an Urban Park . Canadian Field-Naturalist 116(4): 636-639.

Smith, R.A., and Kennedy M.L., 1985. Demography of the raccoon (*Procyon lotor*) at Land Between the Lakes. Kentucky Academy of Science 46: 44-45.

Smith, W.P., Borden, D.L., and Endres, K.M., 1994. Scent-Station Visits as an Index to Abundance of Raccoons - an Experimental Manipulation. Journal of Mammalogy 75(3): 637-647.

Snowling, S.D. and Kramer, J.R., 2001. Evaluating modelling uncertainty for model selection. Ecological Modelling 138: 17-30.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S. and Soltis, P. S. 2006 Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15: 4261-4293.

Sonenshine, D.E., and Winslow, E.L., 1972. Contrasts in distributions of raccoons in two Virginia localities. Journal of Wildlife Management 36(3): 838-847.

Stanley, T.R. and Burnham, K.P. 1998. Information-Theoretic Model Selection and Model Averaging for Closed-Population Capture-Recapture Studies. Biometrical Journal 40(4): 475-494.

Sterner, R.T. and G.C. Smith. 2006. Modelling wildlife rabies: Transmission, economics, and conservation. Biological Conservation 131: 163-179.

Stevens, J., Nixon, C., and Suver, J., 1995. Some aspects of raccoon life history in Lee County, Illinois. Transactions of the Illinois State Academy of Science 88, 1 and 2: 49-59.

Stoen, O.G., Zedrosser, A., Saebo, S., and Swenson, J.E. 2006. Inversely density-dependent dispersal in brown bears. Oecologia 148: 356-364.

Stuewer, F.W., 1943. Raccoons: their habits and management in Michigan. Ecological Monographs 13: 203-257.

Tabatabai, F. and Kennedy, M., 1989. Movements of relocated raccoons (*Procyon lotor*) in western Tennessee. Journal of the Tennessee Academy of Science 64(4): 221-224.

Thomas, L., J.L. Laake, S. Strindberg, F.F.C. Marques, S.T. Buckland, D.L. Borchers, D.R. Anderson, K.P. Burnham, S.L. Hedley, J.H. Pollard, and J.R.B. Bishop. Distance 4.1. Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/

Thulke, H.H., Tischendorf, L., Staubach, C., Selhorst, T., Jeltsch, F., Muller, T., Schluter, H., and Wissel, C., 2000. The spatio-temporal dynamics of a post-vaccination resurgence of rabies in foxes and emergency vaccination planning. Preventive Veterinary Medicine. 47(1-2):1-21.

Tischendorf, L., Thulke, H.H., Staubach, C., Muller, M.S., Jeltsch, F., Goretzki, J., Selhorst, T., Muller, T., Schluter, H., and Wissel, C. 1998. Chance and risk of controlling rabies in large-scale and long-term immunized fox populations. Proceedings of the Royal Society of London Series B-Biological Sciences. 265(1399):839-846.

Tilman, D., Lehman, C.L. and Kareiva, P. 2002. Population dynamics in Spatial Habitats. *In*: Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions, D. Tilman, P. Kareiva (Editors), Princeton University Press, Princeton, New Jersey, USA. pp. 3-20.

Tinline, R.R., Ball, D., and Broadfoot, J. 2007. The Ontario Rabies Model. Wildlife Research and Development Section, Ontario Ministry of Natural Resources, Peterborough, Canada.

Tinline, R.R., Rosatte, R.C., and MacInnes, C.D., 2002. Estimating the incubation period of raccoon rabies: a time-space clustering approach. Preventative Veterinary Medicine 56: 89-103.

Totton, S. C., Tinline, R. R., Rosatte, R. C. and Bigler, L. L. 2002 Contact rates of raccoons (*Procyon lotor*) at a communal feeding site in rural Eastern Ontario. Journal of Wildlife Diseases 38: 313-319.

Totton, S.C., Rosatte, R.C., Tinline, R.R. and Bigler, L.L., 2004. Seasonal home ranges of raccoons, Procyon loto, using a common feeding site in rural Eastern Ontario: Rabies management implications. Canadian Field-Naturalist 118(1): 65-71.

Turchin,P. 2003. Complex Population Dynamics: A Theoretical/Empirical Synthesis. Princeton University Press, New Jersey. 450 pp.

Twichell, A.R., and Dill, H.H., 1949. One hundred raccoons from one hundred and two acres. Journal of Mammalogy 30: 130-133.

Uchmanski, J. and Grimm, V. 1996. Individual-based modelling in ecology: what makes the difference? Trends in Ecology and Evolution 11(11): 437-441.

Urban, D., 1970. Raccoon populations, movement patterns, and predation on a managed waterfowl marsh. Journal of Wildlife Management 34(2): 372-382.

Viet, A.F., Fourichon, C., Seegers, H., Jacob, C., and Guihenneuc-Jouyaux, C., 2004. A model of the spread of the bovine viral-diarrhoea virus within a dairy hear. Preventive Veterinary Medicine. 63(3-4):211-236.

Voigt, D.R., Tinline, R.R., and Broekhoven, L.H., 1985. A spatial simulation model for rabies control *In* Population Dynamics of Rabies in Wildlife, P.J. Bacon (Editor), Academic Press Inc. Ltd., London. pp. 311-349.

Wandeler, A. and Salsberg, E. 1999 Raccoon rabies in eastern Ontario. Canadian Veterinary Journal 40: 731.

Ward, R. H., Frazier, B. L., Dew-Jager, K. and Paabo, S. 1991 Extensive mitochondrial diversity within a single Amerindian tribe. Proceedings of the National Academy of Science USA 88: 8720-8724.

Weiner, J. 1982. A neighbourhood model of annual-plant interference. Ecology 63(5): 1237-1241.

Weiner, J. 1984. Neighbourhood interference amongst *Pinus rigida* individuals. Journal of Ecology 72: 183-195.

White, E.P., Ernest, S.K.M., Kerkhoff, A.J., and Enquist, B.J. 2007. Relationships between body size and abundance in ecology. TRENDS in Ecology and Evolution 22:323-330.

White, G.C. and K.P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46(supplement): 120-138.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., and Freckleton, R.P., 2006. Why do we still use stepwise modeling in ecology and behaviour? Journal of Animal Ecology 75: 1182-1198.

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

Williams, C.K., Ives, A.R., and Applegate, R.D. 2003. Population dynamics across geographical ranges: time-series analysis of three small game species. Ecology 84:2654-2667.

Wilson, K.R. and D.R. Anderson. 1985. Evaluation of two density estimators of small mammal population size. Journal of Mammalogy 66(1): 13-21.

Wilson, M. L., Bretsky, P. M., Cooper-JR, G. H., Egbertson, S. H., Van Kruiningen, H. J. and Cartter, M. L. 1997 Emergence of raccoon rabies in Connecticut, 1991-1994: Spatial and temporal characteristics of animal infection and human contact. American Journal of Tropical Medicine and Hygiene 57, 457-463.

Winkler, W. and Jenkins, S. 1991. Raccoon Rabies *In* The Natural History of Rabies, ed. Baer, G. M. (CRC Press, Boca Raton, Florida), pp. 325-240.

Wood, J.E., 1955. Notes on the reproduction and rate of increase of raccoons in the post oak region of Texas. Journal of Wildlife Management 19: 409-410.

Zeveloff, S. I. 2002. Raccoons: a natural history. UBC Press, Vancouver, Canada, 240 pp.

Zeveloff, S.I. and Doer, P.D., 1981. Reproduction of raccoons in North Carolina. J. Elisha Mitchell Sci. Soc. 97: 194-199.