INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Landscape structure and the distribution of
Sin Nombre hantavirus in deer mouse
Peromyscus maniculatus populations.

by

Jean Langlois

M.Sc.Thesis

defended 12 December 1996

Department of Biology, Carleton University

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22142-3

The undersigned hereby recommend to the
Faculty of Graduate Studies and Research
acceptance of the thesis,

"Landscape structure and the distribution of
Sin Nombre hantavirus in deer mouse
Peromyscus maniculatus populations."

submitted by
JEAN LANGLOIS, B.Sc.

in partial fulfilment of the requirements
for the degree of Master of Science

Chair, Department of Biology

Thesis Co-Supervisor

Thesis Co-Supervisor

Carleton University

Abstract

The purpose of this study was to investigate the role of landscape structure in the distribution of a virus in its host population. In particular I examined the effect of habitat barriers, landscape composition, and landscape configuration on differences in Sin Nombre hantavirus seroprevalence within and among populations of the deer mouse *Peromyscus maniculatus*, and the effect of the spatial scale of measuring landscape structure. I found heterogeneous distribution of the virus within a putatively homogeneous mouse population, and significant effects of landscape composition and configuration on seroprevalence at sites across Canada. Landscape composition was a more important predictor of seroprevalence among these sites than were seasonal variation, viral strain, climate, buildings, or association with human disease. Epidemiological models with a spatially structured host population can benefit from the explicit consideration of landscape structure.

Acknowledgements

I wish to thank my supervisors Dr. Lenore Fahrig and Dr. Gray Merriam. They are both truly teachers. My advisors Dr. Mark Forbes and Dr. Antoine Morin provided insightful perspectives at several stages of this work. Dr. Harvey Artsob of Health Canada introduced me to the world of Hantavirus (unwittingly sparking the conception of this research), and then guided me through it. Health Canada provided financial support and laboratory facilities. Darryl Dick provided valuable guidance and assistance in the lab. Dave Omond helped with fieldwork and air photos. Thanks to the many people who send rodent sera to be included in Health Canada's hantavirus monitoring program, especially to those whose data were included in this research. Some of these people generously spent time searching through trapping records to supply missing information. Thanks to the people who are the Landscape Ecology lab at Carleton. A special thank you to all the rodents who visited my traps; I hope you were not too inconvenienced during your stay.

Dedication

For Sandy,
whose support takes innumerable forms,
each immeasurably important.

Contents

Title page	i
Acceptance form.	ii
Abstract	iii
Acknowledgements	iv
Contents	v
List of Tables	vi
List of Figures	vii
List of Appendices	viii
1 - GENERAL INTRODUCTION	
Components of probabilities in epidemiological models	1
Subdivided populations	3
Landscape ecology	5
Objectives of study	7
The system: deer mouse and Sin Nombre virus	9
Plan of study	12
2 - LOCAL FIELDWORK	
2.1 Introduction	14
2.2 Methods	20
2.3 Results	23
2.4 Discussion	28
3 - NATIONAL DATABASE	

3.1 Introduction	30
3.2 Methods	38
3.3 Results	50
3.4 Discussion	70
4 - GENERAL DISCUSSION AND CONCLUSIONS	76
REFERENCES	78
APPENDICES	86

List of Tables

Table 2.3 Hantavirus serology of deer mice (<i>Peromyscus maniculatus</i>) from the study area	25
Table 2.4 Statistical tests of landscape effects on hantavirus seroprevalence.	26
Table 2.5 Trapping success.	27
Table 3.3 Stepwise logistic regression results and diagnostics for the four spatial models.	51
Table 3.9 Stepwise logistic regression results and diagnostics for the temporal model.	65

List of Figures

Figure 1.1 Conceptual flowchart.	8
Figure 2.1 Schematic map of study area.	16
Figure 2.2 Forms of replication in experimental designs involving fragmentation.	17
Figure 3.1 Map of trapping locations across Canada.	40
Figure 3.2 Transect sampling method used to characterize landscape structure at each site.	41
Figure 3.4 Overall fit of the model is influenced by the scale of sampling of landscape data.	52
Figure 3.5a Effect of landscape composition on predicted seroprevalence. . .	55
Figure 3.5b Distribution of seroprevalence by index of landscape composition (at the 1 Km scale) for the 101 sites.	56
Figure 3.6a Effect of landscape configuration on predicted seroprevalence. .	57
Figure 3.6b Distribution of seroprevalence by index of landscape configuration (at the 1 Km scale) for the 101 sites.	58
Figure 3.7a Effect of julian date on predicted seroprevalence.	59
Figure 3.7b Distribution of seroprevalence by date of trapping for the 101 sites.	60
Figure 3.8a Effect of mean annual temperature on predicted seroprevalence.	61
Figure 3.8b Distribution of seroprevalence by mean annual temperature for the 101 sites.	62
Figure 3.10a Seasonal trend in seroprevalence.	66
Figure 3.10b Distribution of seroprevalence by trapping date for the 392 site-date subsamples.	67

List of Appendices

Appendix 1	Frequencies of seropositive mice from 101 sites across Canada sorted by landscape composition index.	86
Appendix 2	Frequencies of seropositive mice from 101 sites across Canada sorted by landscape configuration index.	89
Appendix 3	Frequencies of seropositive mice from 101 sites across Canada sorted by trapping date.	92
Appendix 4	Frequencies of seropositive mice from 101 sites across Canada sorted by mean annual temperature.	95
Appendix 5	Complete dataset for the spatial regression models	98
Appendix 6	Complete dataset for the temporal regression model	104

CHAPTER 1 - GENERAL INTRODUCTION

Components of probabilities in epidemiological models

Epidemiological models commonly model the dynamics of infection in a population in terms of the size of the host population, number of infected individuals, the timecourse of infection (from acute to chronic infection to death or immunity), the duration of the infectious period, and the probability of transmission of the disease organism from an infectious host to a susceptible host (Becker 1989, Ahlbom 1993). This transmission probability is often taken as a single parameter, but in fact it has two components:

$$P(\text{transmission}) = P(\text{transmission} \mid \text{contact}) \times P(\text{contact}) \quad \{\text{Eq. 1.1}\}$$

Where:

$P(\text{transmission})$ = probability of transmission of infection from an infectious host to a susceptible host.

$P(\text{transmission} \mid \text{contact})$ = conditional probability of transmission, given contact.
i.e. the probability of transmission of infection from an infectious host to a susceptible host, once they have come into contact.

$P(\text{contact})$ = probability of an infectious host and a susceptible host coming into contact.

Note that in this context contact means the level or nature of contact necessary for disease transmission. Depending on the biology of the disease organism, there may be a number of different forms of contact, each with a specific conditional probability of transmission. In this case:

$$P(\text{transmission}) = \sum [P(\text{transmission}|\text{contact}_i) \times P(\text{contact}_i)] \quad \{\text{Eq. 1.2}\}$$

Where:

$P(\text{transmission})$ = probability of transmission of infection from an infectious host to a susceptible host.

$P(\text{transmission}|\text{contact}_i)$ = conditional probability of transmission of infection from an infectious host to a susceptible host, given that they have contact of the i th type.

$P(\text{contact}_i)$ = probability of an infectious host and a susceptible host coming into contact of the i th type.

The conditional probability of transmission, $P(\text{transmission}|\text{contact})$, is largely determined by the biological attributes of the disease organism. These are the usual epidemiological parameters such as infectivity (the probability of a virus successfully colonising a susceptible individual when it has the opportunity), out-of-host survival (which increases the odds of the virus having an opportunity), and mobility (such as becoming airborne in the case of Sin Nombre).

In contrast, the probability of contact between an infectious and susceptible

host, $P(\text{contact})$, is largely determined by the biological attributes of the host organism. Some factors involved here are population density (with spatial and temporal changes), population dynamics (birth, death, immigration, emigration), specific contact behaviours, and the movements of individuals within the population.

Models that express the transmission probability as a single parameter assume, implicitly or explicitly, that $P(\text{contact})$, or each $P(\text{contact}_i)$, is constant and equal for all individuals in the host population. This assumption may be a sufficient approximation for some host populations at some scales (*e.g.* fish in an aquarium), but not for a subdivided population living in a heterogeneous environment (*e.g.* mice in an agricultural landscape).

Subdivided populations

Several models of subdivided populations have been proposed (MacArthur and Wilson 1967, Levins 1970, Boorman and Levitt 1973, Pulliam 1988, Hanski 1991, Hanski and Gilpin 1991). McCauley (1995) argues that various models of subdivided populations are all points on a continuum of population stability, and for the current discussion it is sufficient to consider them as special cases of a “general metapopulation model”, by which I mean a more generalised model than Levins’s (1970) original “metapopulation”. This general metapopulation model views a regional population as composed of a group of local populations, or subpopulations. Each subpopulation has its own dynamics

which are more or less independent from those of the other subpopulations, and there is some dispersal of individuals between the subpopulations.

The assumption of a universal $P(\text{contact})$ does not hold in the case of a metapopulation. Here $P(\text{contact})$ is greater within a subpopulation and smaller between subpopulations. $P(\text{contact})$ is also greater among subpopulations with more dispersal between them, and smaller among subpopulations with less dispersal between them. I use the term population connectivity to denote the extent of dispersal among subpopulations (note the distinction from landscape connectivity below). Thus $P(\text{contact})$ among subpopulations is directly related to population connectivity.

Epidemiological models recognise explicitly that the “habitat” of the disease organism is distributed in discrete units (individual hosts), with different dynamics within and among these units. Metapopulation models, particularly in a landscape ecology framework (see below), recognise explicitly that the habitat of the host organism is distributed in discrete units (habitat patches), with different dynamics within and among these units.

Predictions about population persistence made by spatially explicit population models sometimes differ importantly from those made by models using simplifying assumptions (see Kareiva and Wennergren 1995). Similarly, more useful predictions may result from epidemiological models that recognise explicitly the hierarchically structured $P(\text{contact})$.

Landscape ecology

A major endeavour of landscape ecology is to define the structural and functional characteristics of the landscape and understand their effect on the movement of individuals and thus the dynamics of populations, and ultimately their persistence. The landscape is considered to be heterogeneous and composed of an assortment of landscape elements; a delimited landscape element is often called a patch, and the term matrix is sometimes used to denote a landscape element which dominates the landscape (see Forman and Godron 1986, Naveh and Lieberman 1994). For example woodlots in an agricultural landscape might be referred to as patches of forest embedded in a matrix of crop land.

I categorise landscape elements under three terms, although they are all points on a gradient:

- (1) habitat : a range of landscape elements which individuals use to varying degrees (breeding habitat, foraging habitat, marginal habitat, etc.).
- (2) non-habitat: landscape elements which individuals do not use, but move through to varying degrees to get to other patches.
- (3) barriers: landscape elements which individuals do not use or move through, this is a limit case of non-habitat.

Habitat, non-habitat, and barriers may each exist in a landscape as either patches or matrix. A commonly described situation is a mosaic of patches of

habitat, non-habitat, and barriers. Others are barrier patches in a matrix of habitat, and patches of habitat in a matrix of non-habitat. If this last example describes the landscape in which a metapopulation exists, then subpopulations occur in habitat patches and dispersal between subpopulations occurs across the non-habitat matrix.

Landscape structure is generally defined in terms of “composition” and “configuration”. These are, respectively, the kinds of patches present in the landscape and their relative abundance, and spatial relationships among them. The sizes and shapes of the patches may be conceived either as an aspect of composition or of configuration. The term fragmentation is used here to mean the extent to which a landscape element is subdivided into separate units.

A functional measure of landscape structure is “connectivity”, the degree to which the landscape facilitates or impedes the movement of individuals among patches (Taylor *et al.* 1993). When the landscape is composed of habitat patches embedded in a matrix used only for dispersal, the connectivity of the landscape for any given composition is thought of as the combined result of the configuration of patches and the ease of movement of individuals through the matrix. Connectivity may vary across such a landscape if the ease of movement of individuals through the matrix varies (Taylor *et al.* 1993) or the configuration of the patches varies.

Under the landscape ecology framework, we expect that seroprevalence could

become uniform throughout a population living in a landscape with high connectivity. The limit situation is one where there is no non-habitat or barriers, but instead the entire landscape is composed of continuous habitat (i.e. one huge patch, no fragmentation). In this situation we expect individuals to be able to move freely across the landscape, $P(\text{contact})$ becomes high, and so does $P(\text{transmission})$.

At the other end of the connectivity spectrum we have barriers. In theory a total barrier is impassible by the animal in question and there is no movement of individuals across this landscape element. In practice, putative habitat barriers are more likely to be partial barriers, or non-habitat that considerably restricts the movement of individuals. We would expect $P(\text{transmission})$ between individuals separated by a habitat barrier to be relatively low, since $P(\text{contact})$ is low.

Objectives of study

The general objective of this study is to answer the question: how does landscape structure affect the pattern of distribution of a virus in a subdivided host population? Figure 1.1 summarizes the theoretical framework for this question developed in the preceding discussion. In particular this study investigates the role of habitat barriers, landscape composition, and landscape configuration on differences in Sin Nombre hantavirus seroprevalence (the proportion of hosts testing positive for antibodies to the virus) within and among

Figure 1.1 Conceptual flowchart summarizing some factors affecting the relationship between measurements of landscape structure and the observed distribution of antibodies to Hantavirus in the rodent population.

deer mouse *Peromyscus maniculatus* populations.

Sin Nombre can be a sensitive tool for assessing the extent of movement of individuals among deer mouse subpopulations. The level of interaction required for viral transmission is spatially specific but temporally general (*i.e.* individuals need to be at the same place but not necessarily at the same time). The antibodies of an exposed individual provide a record of that individual's exposure. Propagation of the virus within the population is much more rapid than is the case for a genetic marker, and can occur within a generation, not only to subsequent generations. Finally screening for antibodies is relatively inexpensive and can be integrated into a well planned serological monitoring program.

The system: the deer mouse and Sin Nombre virus

The deer mouse

The deermouse, *Peromyscus maniculatus*, is ubiquitous throughout most of North America south of the tree line (Carleton 1989 p.47) and has been described as "probably the most widely distributed native small mammal on the continent" (Hooper 1968 p.43).

Peromyscus maniculatus are often described as habitat generalists (Baker 1968), with some habitat specialization among subspecies (Wecker 1963). While there is some contradiction in habitat preferences reported in the

literature (see Kaufman and Kaufman 1989), one generalisation that can be made is that deer mice tend to choose habitats with greater vertical structural diversity (Kaufman and Kaufman 1989), for example choosing mature trees over open areas or ground vegetation (Parren and Cooper 1985, Wolff and Hurlbutt 1982, Barry et al. 1984, Stah 1980, Sekgororoane 1995), choosing coulees over open prairie (Morris 1992; Reynolds, unpublished data), and choosing among clearcuts those with higher diversity of slash and foliage height (Buckner and Shure 1985).

Dispersal occurs primarily in juveniles, and mostly during the breeding season (Wolff 1989). Juvenile dispersal distance averages less than 150m (Baker 1968 p.108), is usually less than 300 metres (>89%, n =413, from data in Stickel 1968 p.395), and is rarely greater than 900 metres (<2%, n=428, from data in Blair 1950 and Stickel 1968). The longest reported dispersal distance by a deer mouse is 1220 m (Dice and Howard 1951 in Stickel 1968, p.395), while the longest reported homing distance by an experimentally relocated deer mouse is 1980 m (Teferi and Millar 1993).

During the summer, deermice nest singly or occasionally in pairs or family groups, but they may nest communally during the winter (Wolff, 1989, p. 282). In Canada seasonal variation of deer mouse population levels is extensive, with "fall" population levels averaging more than twice as high as "spring" population levels (see review by Terman 1968).

Sin Nombre virus

Sin Nombre is a recently recognized hantavirus which is widely distributed across North America (see Rowe *et al.* 1995). The deer mouse *P. maniculatus* is the main host, while other rodents can carry the virus (Childs *et al.* 1994). Sin Nombre virus (also originally referred to as Four Corners or Muerto Canyon virus) was identified after an outbreak of human disease in the southwestern U.S. in 1993 (Nichol *et al.* 1993). It has since been studied intensively because of the public health risk posed by the human disease, Hantavirus Pulmonary Syndrome (HPS). HPS causes severe pulmonary edema, with about half of the cases being fatal. The epidemiology of Sin Nombre virus within the mouse population has been studied very little however, and many important fundamental parameters remain unknown. These include the mode(s) of transmission between mice, and the timecourse of infectivity (Dr. Harvey Artsob, Zoonotic Diseases, LCDC, Health Canada, personal communication).

The primary mode of Sin Nombre infection in humans is by inhalation of airborne virus particles released from the faeces and urine of infected mice, particularly as it dries. The mode or modes of transmission between mice are unknown. Inhalation of airborne virus particles may be the primary mode of transmission, but other possibilities include blood contact during aggressive interactions.

Also unknown are the roles of "vertical transmission", "perinatal transmission", and maternal antibodies, each of which can play an important role in the epidemiology of mammalian viruses. Vertical transmission is the transmission from an infected mother directly to the fetus (or fetuses) through the placenta.

This would result in all offspring of an infected female being infected. Perinatal transmission occurs when a virus which does not cross the placental barrier is transmitted to offspring by blood contact during birthing. This would result in some of the offspring of an infected female being infected. Neonates typically carry maternal antibodies for some time after birth. If the antibodies are effective at combating infection this would result in a low incidence of postnatal infection of neonates. If the offspring remain resistant until they leave the nest, and if they are not infected prenatally or perinatally, then very few of an infected female's offspring would be infected.

The rodent hosts of hantaviruses develop a persistent or chronic infection but with no apparent disease symptoms (LeDuc 1987, Lee and van der Groen 1989).

Plan of study

This study is comprised of two components which address the study objectives at different spatial scales. Chapter 2 reports on local fieldwork designed to test hypotheses about the effect of putative habitat barriers on the observed seroprevalence of Sin Nombre virus in the *Peromyscus* population. Chapter 3 reports on the creation of a national dataset of serological and ecological data on deer mouse populations, and analyses designed to investigate the effect of landscape composition and configuration on the observed seroprevalence of Sin Nombre virus in deer mouse populations. The experimental unit in the local

fieldwork is a habitat barrier (at the scale of about one hundred metres), of which there are several replicates within the same landscape. The experimental unit in the national data analysis is a landscape (at the scale of a few kilometres in diameter), of which there are several replicates across the country.

CHAPTER 2 - LOCAL FIELDWORK

2.1 INTRODUCTION

I set out here to test the fundamental predictions of a metapopulation model consisting of subpopulations living in habitat patches separated by barriers. These are firstly that hantavirus seroprevalence is different across barriers, and secondly that hantavirus seroprevalence is not different across continuous habitat, either within a patch or across a uniform matrix of habitat. The specific objective was to see if the patterns of seroprevalence were consistent with pasture acting as a barrier to the movement of *Peromyscus*. *P. maniculatus gracilis* prefers forest habitat, avoiding grassy vegetation where forests have been cleared (Blair 1950). Forest cleared for roads or pasture seems to behave as a partial habitat barrier for *P. leucopus noveboracensis* (Merriam *et al.* 1989, Merriam and Lanoue 1990, Wegner and Merriam 1990).

Ten of 186 deer mice trapped at locations in and around Algonquin Park, Ontario, during 1994 and 1995 tested positive for Sin Nombre (seroprevalence = 5.4%, Artsob *et al.*, unpublished). This is a large area of forest and logged areas. The landscape changes gradually as one moves outside the park; the forest becomes increasingly fragmented as a result of agricultural clearing. The landscape further to the south and east is an agricultural matrix with remnant forest fragments.

A field site was needed as close as possible to a site with known hantavirus occurrence in Algonquin Provincial Park but with some fragmentation of the forest landscape. The area near Barry's Bay, Ontario (Figure 2.1) met these requirements. The unusual configuration of forest fragments and forest matrix in this area (see below) made possible an experimental design superior to many similar studies.

Typical approaches in field studies of barriers between habitat patches are to compare a fragmented area to a neighbouring area of continuous habitat (*e.g.* Martinsson *et al.* 1993), or to compare differently fragmented areas which can be described in terms of the size or arrangement of habitat patches (*e.g.* Diffendorfer *et al.* 1995). Such a design yields results with little generality, the effective sample size often is one for each treatment (forest matrix *versus* agricultural matrix) because each treatment is not replicated in space (Figure 2.2a). With *Peromyscus* the habitat patches in question are typically woodlots in an agricultural matrix and/or a large area of continuous forest. All the woodlots are typically embedded in the same agricultural matrix, and so each woodlot is a subsample or a pseudoreplicate (Hurlbert 1984).

The landscape in this study is fundamentally different, with the forest fragments separated from but surrounded by a common forest matrix (Figure 2.2b). The area is mostly a matrix of forest, within which a few relatively small pastures have been cleared, with a single woodlot (or in one case, two woodlots) remaining within some of these pastures. Each experimental unit consists of a patch of habitat (woodlot) surrounded by a patch of barrier (pasture), embedded

Figure 2.1 Schematic map of study area near Barry's Bay, Ontario. ■ =forest, □ =pasture, scale is approximate. The area is mostly a matrix of forest, within which a few relatively small pastures have been cleared, with a single woodlot (or in one case, two woodlots) remaining within some of these pastures. The landscape is dominated by an agricultural matrix to the south and east (bottom right). R indicates the areas from which seropositive reference samples were available.

Figure 2.2 Forms of replication in experimental designs involving fragmentation. For example ■ = forest, □ = pasture, x = trapline. (a) subsampling design with pseudoreplication. (b) design with true spatial replication. The test in (a) is fragmented landscape compared to unfragmented landscape, using one paired sample, each with 5 subsamples. In (b) the test is separated from the forest matrix by pasture compared to connected to the forest matrix, using 5 paired samples.

within a matrix of habitat (forest). All 9 experimental units share a common matrix of habitat. At a slightly larger scale this same forest matrix extends to include the locations inside Algonquin Park where deer mice were found to carry the virus.

If the forest matrix has a high connectivity for *Peromyscus* we would expect uniform seroprevalence among the positive sites in the park and the study site, and throughout the forest matrix within the study site. Further, if pasture acts as a barrier we would expect seroprevalence to be different between the woodlots and the forest matrix.

The two species *Peromyscus maniculatus gracilis* (deer mouse) and *P. leucopus noveboracensis* (white-footed mouse) occur sympatrically in this region, and I expected to trap both species with approximately equal probability (Dobbyn 1994, Strickland and Rutter 1994). The white-footed mouse has also been identified as a host for a Hantavirus implicated with Hantavirus Pulmonary Syndrome (Shelter Island or New York Virus, Song *et al.* 1994). Sin Nombre has been shown to infect species other than deermice, but it is not known whether these other species can transmit the virus; infection of other species is thought to occur primarily as a “spill over effect” when prevalence in the main host is unusually high (Childs *et al.* 1994). It is possible that white-footed mice in this area may have antibodies to Sin Nombre, or Hantavirus antibodies that are cross-reactive with Sin Nombre. If Sin Nombre does occur in white-footed mice in this area, we would expect seroprevalence to be lower than among deer mice.

In summary, the hypotheses to test are:

H1- Hantavirus seroprevalence is different between woodlots and the forest matrix.

H2- Hantavirus seroprevalence is not different throughout the forest matrix within the study site.

H3- Hantavirus seroprevalence is not different across the forest matrix between the study site and reference sites in Algonquin Park.

H4- Hantavirus seroprevalence is different between *P. maniculatus* and *P. leucopus*.

2.2 METHODS

Sampling scheme

I used a paired design, in each case trapping mice on a trapline in the woodlot inside the pasture, and on a trapline just outside the pasture in the continuous forest matrix. The treatment was in effect the clearing (historically) of a ring of pasture around a fragment of forest left in the middle. Thus we are examining the barrier effect of a ring of pasture, replicated over several woodlots, and not the effect of fragmentation of the landscape. This treatment was truly replicated in space across my study area. This design yields more generalizable results because it samples spatial heterogeneity, whereas in the pseudoreplicated design spatial heterogeneity confounds treatment effects.

Trapping and sample collection

Live-traps supplied with dacron bedding and sunflower seeds were set at one to three sites each afternoon and checked the following morning. Nest boxes supplied with dacron bedding (Nicholson 1941) were also installed along the traplines and checked concurrently with trapping. Live-traps were used from 9 August to 14 October, and nest boxes from 4 September to 27 October, 1995. Nest boxes were used in an attempt to assess whether infection was common to mice nesting together.

Each mouse was released from the trap directly into an airtight container

containing a swab saturated with metofane anaesthetic, and removed as soon as it was anaesthetized (Olfert *et al.* 1993). A blood sample ($\sim 40 \mu\text{L}$) was taken from the orbital sinus using a heparinized capillary tube (Cunliffe-Beamer 1983). The blood was extruded onto a filter paper strip, which was then bagged and labelled. Standard Monel ear tags were attached in order to identify recaptured mice. The mouse was weighed using Pesola scales. A saliva sample ($\sim 20 \mu\text{L}$) was taken by rinsing the mouth with distilled water using a disposable Pasteur pipet, and stored in a cryotube (Merriam *et al.* 1989). The mouse was identified to species (Smith 1993), was allowed to recover from the anaesthetic, and was released where it was trapped. Blood and saliva samples were stored on ice in the field and at -20°C in the lab until used.

Biosafety precautions and laboratory procedures

Biosafety precautions were necessary for handling potentially infectious rodents in the field. HEPA filter respirators, coveralls, latex gloves, and biosafety and disinfection protocols similar to those described in Mills *et al.* (1995) were used.

I tested mouse sera for the presence of IgG antibodies by enzyme-linked immunosorbent assay (ELISA), using a recombinant nucleocapsid antigen produced from Sin Nombre virus. Laboratory work was conducted at the Zoonoses lab at Health Canada's Laboratory Centre For Disease Control in Ottawa. ELISA procedures were modified from Centres for Disease Control 1994.

Statistical analysis

Fisher's exact tests were used to test for significant differences in seroprevalence between woodlots and forest. Results were also compared between my study area and the several locations within Algonquin Park, from mice which were trapped during the summers of 1994 and 1995 and submitted to Health Canada for Hantavirus screening.

2.3 RESULTS

Hantavirus antibodies were not detected in any of the 124 mouse sera collected from the Barry's Bay study area. The 95% confidence interval for the true seroprevalence in deer mice in this area is 0% to 2.7% (Table 2.3). No difference in seroprevalence was observed between woodlots and forest in the study area; both were statistically not different from zero and their 95% confidence intervals overlapped completely (Table 2.3). A non-significant difference was observed between seroprevalence in the forest in the study area and the concurrent seroprevalence (during 1995) in the reference sites in Algonquin Park (Table 2.4). Seroprevalence was consistent in the reference sites during the 2 years for which data are available (Table 2.4). The data for both years in the reference sites can be combined because they are statistically indistinguishable. Data for forest and woodlots in the study area also can be pooled. The zero overall seroprevalence for the study area is significantly different from the overall seroprevalence in the reference sites ($p=0.01$, Table 2.4).

White-footed mice (*P. leucopus*) were much less abundant than deermice (*P. maniculatus*) (Table 2.5). *P. leucopus* comprised only 13.7% of the *Peromyscus* trapped; the 95% confidence interval for the true proportion is 7.7% to 19.8%.

Hypotheses tested:

H1- We do not reject the null hypothesis of no difference in hantavirus

seroprevalence between woodlots and the forest matrix.

- H2- The all-negative results from the field site do not permit any test of difference throughout the forest matrix within the study site.
- H3- Hantavirus seroprevalence is significantly different across the forest matrix between the study site and reference sites in Algonquin Park.
- H4- The all-negative results from the field site do not permit any test of difference in seroprevalence between *P. maniculatus* and *P. leucopus*.

Table 2.3. Hantavirus serology of deer mice (*P. maniculatus*) from the study area. All mice tested negative for hantavirus antibodies. The 95% confidence interval for the true seroprevalence in the population is shown.

landscape element	Woodlots	Forest	Combined
number of seropositive mice (y)	0	0	0
total number of mice trapped (n)	50	57	107
seroprevalence (y/n)	0.0	0.0	0.0
95 % confidence interval	0 to .058	0 to .051	0 to .027

Table 2.4. Statistical tests of landscape effects on hantavirus seroprevalence. The first comparison shows no demonstrable effect of habitat barriers. The second comparison shows a consistent seroprevalence (~5.4%) over two years in Algonquin Park. The third comparison shows a non-significant difference in contemporaneous samples across continuous habitat. The fourth comparison shows a significantly higher seroprevalence in the control area. p-values shown are for a 2-tailed Fishers Exact Test.

Comparison	proportions	Level of Significance (p-value)
Forest vs. Woodlots	0/57 0/50	no difference (see Table 2.3)
Algonquin Park 1994 vs. Algonquin Park 1995	4/84 6/102	1.00
Forest vs. Algonquin Park 1995	0/57 6/102	.08
Study Area (Forest + Woodlots) vs. Algonquin Park (1994+95)	0/107 10/186	.01

Table 2.5. Trapping success. Deermice were more than 6 times more common than white footed mice.

Species	<i>P. maniculatus</i>	<i>P. leucopus</i>
Woodlots	50	11
Forest	57	6
Total	107	17

2.4 DISCUSSION

The most important result from this field study is that hantavirus seroprevalence is significantly different across continuous habitat between the study site and reference sites in Algonquin Park. (H3). Under simplified assumptions under the landscape ecology framework, we expected uniformity of seroprevalence within a population living in a landscape composed of continuous habitat. This seems to be the first time that a spatially clumped distribution of Sin Nombre has been identified at this spatial scale. Differences in seroprevalence among sites at a similar scale reported in Childs *et al.* (1994) were not statistically significant (my analysis of their Table 4 and Figure 1, $p=.22$, 2-tailed Fishers Exact Test).

The “continuous habitat” of forest may not be as homogeneous as assumed. In this sense these results simply reflect the spatial heterogeneity of the abundance and distribution of an organism living in a structured and dynamic habitat, but the heterogeneity in this system is compounded. The heterogeneous distribution of mice among patches of forest compounds that of viruses among patches of mouse.

The observed pattern may result from stochastic processes. Just as local extinctions and recolonizations may occur as a result of stochastic processes in a metapopulation, so Sin Nombre may become temporarily extinct from some subpopulations of *Peromyscus*. Long term studies would be necessary to determine whether subpopulations which are free of virus become

“recolonized” by the dispersal of an infected mouse into the subpopulation.

Connectivity of this landscape for deer mice may be low, even though it qualitatively seems to be composed mostly of a matrix of habitat.. Recaptures even within sites were very rare, so these data do not allow direct assessment of the movement of individuals within or among sites.

CHAPTER 3 - NATIONAL DATA ANALYSIS

3.1 INTRODUCTION

Health Canada's Zoonoses lab at the Laboratory Centre for Disease Control in Ottawa conducts hantavirus surveillance for all of Canada. This surveillance program relies on samples of rodent sera, or snap-trapped animals, sent to Health Canada for hantavirus testing by a wide range of researchers and agencies from across the country. This program has produced a bank of serological information from several thousand rodents trapped from hundreds of locations across the country. Deer mice, *Peromyscus maniculatus*, are the primary reservoir for Sin Nombre (Childs *et al.* 1994), and make up the bulk of the rodents tested in the serological surveillance program.

My main objective was to use these data to test hypotheses about the effect of landscape structure on the distribution of Sin Nombre in deer mouse populations. The data allowed, and indeed required, an examination of the effect of other variables as well. I developed nine hypotheses regarding the correlation of observed Sin Nombre seroprevalence in deer mouse populations with ecological and epidemiological factors. The first four factors relate to *P. maniculatus* population ecology, *i.e.* $P(\text{contact})$; they are: landscape composition, landscape configuration, spatial scale, and seasonal mouse population dynamics. The next four factors relate to Sin Nombre epidemiology, *i.e.* $P(\text{transmission}|\text{contact})$; these are viral strain and three climatic variables:

temperature, precipitation, and drought. The last three are potentially confounding factors: human dwellings, Hantavirus Pulmonary Syndrome (HPS) cases, and annual within-site variability.

Landscape Composition

Deer mice occur at different densities in different habitat types (Barry et al. 1984, Buckner and Shure 1985, Kaufman and Kaufman 1989, Morris 1992, Parren and Capen 1985, Reynolds unpublished data, Sekgororoane 1995, Stah 1988, Wecker 1963, Wolff and Hurlbutt 1982). I use “preferred habitat” to mean a habitat type where published data show *P. maniculatus* occurring at higher densities compared to other habitat types in the area. The transmission probability of Sin Nombre from one individual to another within a habitat patch may be strongly influenced by the *P. maniculatus* population density within that patch. We would thus expect landscape composition, in terms of the amount of preferred habitat, to affect Sin Nombre epidemiology. Sin Nombre might spread more easily through high density *P. maniculatus* populations living in preferred habitat, and these populations would show a higher prevalence of Sin Nombre infection. More sparse *P. maniculatus* populations living in non-preferred habitat might have a sufficiently small transmission probability to limit prevalence of Sin Nombre infection in the population. The first hypothesis is that seroprevalence is related to the amount of preferred habitat in the landscape.

Landscape Configuration

The transmission probability from one patch to another is largely defined by

the contact probability between the subpopulations, *i.e.* the probability of an infected individual moving into an uninfected patch. This probability may be influenced by landscape configuration. If individuals are less likely to move across a fragmented landscape, such fragmentation may form a partial barrier to the transmission of Sin Nombre, and *P. maniculatus* populations in fragmented landscapes would show a lower prevalence of Sin Nombre infection. Alternatively, individuals may be likely to move further in fragmented habitat, and Diffendorfer et al. (1995) found this to be the case for deer mice. In this case individuals may have a wider range of contacts, increasing the transmission probability. In this case we would expect *P. maniculatus* populations living in fragmented landscapes to show a higher prevalence of Sin Nombre infection. The second hypothesis is that seroprevalence is related to the extent of fragmentation of preferred habitat in the landscape.

Scale

Any discussion of the possible role of landscape structure in the epidemiology of Sin Nombre raises the question of appropriate spatial scale. The scale of a study is appropriate if the landscape is characterized at the scale at which the processes of interest operate. In this case the relevant scale is the scale at which individual deer mice utilise the landscape and move through it. One approach would be to define the appropriate scale based on what is known and/or assumed about the movement of deer mice. In this study I go one step further and characterize the landscapes at the presumed most appropriate scale and also at a larger (double) and smaller (half) scale. By conducting the analysis at several different scales, I can test my assumptions and identify the

most appropriate scale as the one at which the independent variables explain the most about the observed pattern of distribution of Sin Nombre. The third hypothesis is that the relation of seroprevalence to landscape variables is dependent upon spatial scale.

Seasonal change

As discussed in chapter 1, *P. maniculatus* populations typically undergo large temporal fluctuations in numbers. Population density is typically very low in the early spring, rises gradually during the summer, and peaks in the fall. A high overwinter mortality reduces the population to very low density by the following spring. If we presume that transmission probability is a function of population density, we would expect the rate of transmission to accelerate during the summer and fall, and a peak in prevalence of Sin Nombre infection in the fall. If we assume that overwinter mortality is equally high among infected and noninfected individuals we would expect a similar proportion of infected individuals the following spring, that is the same prevalence of Sin Nombre infection over the winter. Thus after several seasons, if the prevalence of Sin Nombre infection reaches an equilibrium, we might not expect a seasonal fluctuation in seroprevalence, but only in the absence of differential mortality and only if offspring are recruited into the population with the same seroprevalence as the parent generation. The former condition seems likely but the latter does not (see chapter 1). If seroprevalence in a population is not (yet) at equilibrium we would expect more new infections in the population in the fall of the year, while population density and the potential for immigration of infected individuals are highest. The fourth hypothesis is that seroprevalence is related

to Julian date.

Climate

Risk of infection for humans is known to be greater when mouse excreta become dry, because virus particles become airborne much more readily in these conditions. Thus if we assume inhalation of airborne virus particles to also be an important mode of transmission between mice (see general introduction), we might expect a greater rate of transmission in drier climates. The rate at which excreta become dry is probably affected more strongly by individual microclimates but this information is not available, and broad patterns could also exist at the level of the overall climate. This would be consistent with the observation that the highest prevalences of Sin Nombre infection in both humans and rodents have been reported in arid regions of the United States. The fifth hypothesis is that seroprevalence is related to some combination of temperature, precipitation, and/or drought.

Strain

Recent work by Morzunov *et al.* (1996) has identified some major phylogenetically distinct groups within the Sin Nombre complex, and suggests coevolution of various strains with phylogenetic groups of *P. maniculatus*. Two major strains of Sin Nombre have been identified in Canada which are also clearly geographically allopatric, an eastern variant, occurring east of the 95th meridian, and a western variant occurring west of the 95th meridian (Morzunov, pers. comm.). There is growing suspicion in the public health field that the existence of these different strains may be responsible for the remarkable

difference in human mortality and morbidity from Sin Nombre in eastern and western North America. The genetic difference between the two strains (15% base-pair variability, Morzunov *et al*, 1996) may be sufficient to account for this, and also suggests that the two strains may differ in their epidemiological characteristics in the mouse populations. The sixth hypothesis is that seroprevalence is related to the strain of Sin Nombre.

Buildings

While most of the mice tested as part of the national surveillance program were trapped in the wild on trap lines as part of ecological research or wildlife surveys, many were snap-trapped in human dwellings or other buildings. These included rural homes, farm buildings, buildings on industrial sites, and buildings in urban centres. The demographic relationship between mice in human buildings and the larger wild population is unknown. In many rural areas there is presumably a free flow of mice in and out of buildings, and in this case the building acts as a large 'trap', drawing in mice from the larger wild population. Any spatial, temporal, or demographic details of such a process are unclear. Mouse populations in human buildings may be more or less demographically distinct from wild populations. The within-patch dynamics of these populations (population density, contact behaviour, communal nesting, *etc.*) may also be quite different from those of wild populations. We might therefore expect a confounding effect from some mice being trapped inside buildings. Statistical procedures must account for this potential confounding effect. The seventh hypothesis is that seroprevalence is related to whether mice are trapped in the wild or in buildings.

Hantavirus Pulmonary Syndrome cases

Some of the mice tested as part of the national surveillance program were trapped in and near the homes of people who had been diagnosed with Hantavirus Pulmonary Syndrome (HPS), the respiratory illness associated with Sin Nombre infection in humans. These sites were sampled specifically because a high probability of finding infected mice there was expected. Statistical procedures must account for this bias. The eighth hypothesis is that seroprevalence is related to whether mice were tested in connection with a human HPS case.

Annual within-site variability

Variability in seroprevalence among years within a site could potentially confound results when examining seasonal variability in seroprevalence. Annual within-site variability must be examined to determine the validity of combining data from several years in the same analysis. The ninth hypothesis is that seroprevalence shows significant within-site variability among years.

In summary, the hypotheses to test are:

- H1. Seroprevalence is related to the amount of preferred habitat in the landscape.
- H2. Seroprevalence is related to the extent of fragmentation of preferred habitat in the landscape.

- H3. The relation of seroprevalence to landscape variables (H1 and H2) is dependent upon spatial scale.
- H4. Seroprevalence is related to Julian date.
- H5. Seroprevalence is related to some combination of temperature, precipitation, and/or drought.
- H6. Seroprevalence is related to the strain of Sin Nombre.
- H7. Seroprevalence is related to whether mice are trapped in the wild or in buildings.
- H8. Seroprevalence is related to whether mice were tested in connection with a human HPS case.
- H9. Seroprevalence shows significant within-site variability among years.

3.2 METHODS

date/location data

Records from Health Canada's Hantavirus surveillance program were searched to identify data which could be used for this analysis. The four criteria for inclusion were:

1. $n \geq 30$, or $y \geq 1$
2. date of trapping was known
3. precise trapping location was known
4. size of trapping area was known

1. A minimum sample size (n) of 30 mice was necessary in order to have a 95% certainty that the virus was in fact absent from sites where the number of seropositive mice (y) was zero.
2. When date or location information was not recorded it could sometimes be found with the assistance of the individual or agency responsible for the trapping.
3. Trapping location was obtained and recorded as latitude/longitude and/or Universal Transverse Mercator (UTM) coordinates. Sites had to be at least 2 Km apart in order to be considered spatially independent for this analysis. This minimum distance was chosen because it exceeds the greatest recorded individual dispersal distance for deer mice and greatly exceeds commonly observed individual dispersal distances (see chapter 1). Thus more than one season would normally be required for infection to be carried between two sites

separated by 2 Km.

4. The size of the trapping area had to be known because in some cases a batch of mice were from a large geographic area ($>2\text{Km}$). These were divided into separate sites if the precise trapping locations could be learned.

No practical criterion could be established to define a separation in time required for temporal independence of samples at the same site. These criteria yielded records from 2837 mice, 101 independent sites across Canada (Figure 3.1), and 392 site-date combinations. Once the 101 independent sites were identified, some biophysical data on the sites could be obtained using their geographic coordinates.

Landscape composition and configuration

Air photos were used to measure indices of landscape composition and configuration for each site. For each site, a point on an air photo was identified corresponding to the precise trapping location. A pair of transects corresponding to 8 Km each in length was established on the air photo(s), intersecting on the point of the trapping location. This defined the landscape associated with each trapping record (Figure 3.2). There were cases of partial overlap between landscapes thus defined. The transects were placed parallel to the edges of the air photos, which were arbitrary in their relation to cardinal directions. Using this method each landscape could be characterized at three

Figure 3.1. Map of trapping locations across Canada. 101 sites were included in the analysis. See text for selection criteria.

Figure 3.2 Transect sampling method used to characterize landscape structure at each site. ■ = preferred habitat, □ = non-preferred habitat. A pair of transects was used, each 8 Km long, intersecting at the point of the trapping location. Using this method each landscape was characterized at three different scales, that is at the scale of a one, two, and four kilometre radius.

different scales, that is at the scale of a one, two, and four kilometre radius. Traditional transect sampling technique avoids intersecting transects since the resulting nonindependence of transects hinders the combining and comparing of transects for analysis (Pielou 1977). In the present case, intersecting transects were deliberately used precisely because this biases the resulting indices (see below) for landscape elements closer to the point of intersection, that is closer to the trapping location. In this way the indices of landscape composition and configuration characterise the landscape surrounding the trapping location, while being particularly sensitive to the landscape closest to the trapping location.

From the air photos, all landscape elements (habitat patches) were classified as either preferred habitat or not. Forest, crop fields, clearcuts, and coulees were all classified as preferred habitat in comparison to pasture, hay, open prairie, bog, marsh, lakes, rivers, and asphalt (Barry *et al.* 1984, Buckner and Shure 1985, Morris 1992, Parren and Capen 1985, Reynolds unpublished data, Sekgororoane 1995, Stah 1980, Wolff and Hurlbutt 1982). Each point was recorded along the transect where it intersected a border between preferred habitat and another landscape element. In the terminology of Pielou (1975), the points of occurrence of two *phases* were recorded along each transect: segments of preferred habitat and gaps comprised of non-preferred habitat.

Landscape composition was represented by the total proportion of the transect comprised of preferred habitat, that is:

$$\text{Comp} = \frac{\sum L_{hi}}{L}$$

Where:

Comp = index of landscape composition, ranges from 0 (no preferred habitat)
to 1 (all preferred habitat)

L_{hi} = length of a transect segment comprised of preferred habitat

$\sum L_{hi}$ = sum of all L_h for the transect

L = length of the transect

The mean value of 'Comp' from the pair of intersecting transects was used for each site.

I developed an index of Landscape configuration which reflects the degree of fragmentation of the preferred habitat. The index needed to range from 0 to 1, and be independent of landscape composition (Comp). A review of the literature found no such index which was applicable to the transect method employed here (Gustafson and Parker 1994, O'Neill *et al.* 1988, Wegner 1994). The index is developed by (1) identifying the limit of resolution of the transect sampling procedure (see below) and expressing the length of the transect L as a number of limit-of-resolution units, (2) calculating the theoretical maximum possible number of habitat segments for each landscape, based on these quantities and the proportion of preferred habitat, and (3) expressing the extent

of fragmentation as the proportion of the observed number of habitat segments to the maximum possible number of habitat segments.

(1) The limit of resolution is defined as the shortest unit of length along the transect which can be resolved and classified as either preferred habitat or not, or in other words, the shortest possible segment or gap. This was 1 mm on the air photo transects (which is equivalent to 50 m on the ground), and so segments expressed in mm are also conveniently expressed in limit-of-resolution units.

(2) $n_{\max}$ is the maximum number of distinguishable segments of habitat which could occur for a given total amount of habitat on a given transect. L = length of transect in "limit-of-resolution units", $\sum L_{hi}$ = total amount of habitat on transect in "limit-of-resolution units". Since L is an even number for my transects:

$$n_{\max} = \sum L_{hi} , \text{ when } \sum L_{hi} \leq 1/2 L$$

$$n_{\max} = L - \sum L_{hi} + 1 , \text{ when } \sum L_{hi} > 1/2 L \text{ (i.e. = maximum number of gaps +1)}$$

(3) Landscape configuration was expressed as the proportion:

$$\text{Conf} = \frac{n-1}{n_{\max}-1}$$

Where:

Conf = index of landscape configuration, ranges from 0 (unfragmented)
to 1 (maximally fragmented)

n = observed number of habitat segments

$n_{\max}$ = maximum possible number of habitat segments

The mean value of 'Conf' from the pair of intersecting transects was used for each site.

Other variables

Strain

This was addressed by using a dummy variable, labelled Strain. All sites east of 95° longitude were classified as being in the range of the eastern variant (strain = 0). All sites west of 95° longitude were classified as being in the range of the western variant (strain = 1). The virus strain is of course perfectly correlated to longitude. It would thus be impossible to make a confident inference about any real effect of virus strain, but it would be equally misleading not to include strain data.

Climate

Ecoclimatic zone maps and climate data (Ecoregions Working Group 1989) were used to estimate the mean annual temperature, mean total annual precipitation, and mean annual period of relative drought (the number of months during which mean temperature in °C exceeds mean precipitation in

mm) for each site.

Hantavirus Pulmonary Syndrome

Rather than exclude these samples from the analysis because of this bias, all sites were classified as to whether they were sampled in response to a human Hantavirus Pulmonary Syndrome case (HPS = 1) or not (HPS = 0), in order to detect and control for any effect of this bias.

Buildings

All sites were classified by a dummy variable labelled Building to test for a possible difference between wild-trapped mice (Building = 0) and mice trapped in human dwellings (Building = 1).

Annual within-site variability

Prior to the logistic regression analysis, a Fisher's exact test was used to test for significant difference in seroprevalence among years for all sites where data existed for multiple years (hypothesis H9).

Logistic regression analyses

SAS software (release 6.09, ©1989 SAS Institute Inc., Cary NC, USA) was used to perform a total of 5 logistic regression analyses. The dependent variable in all 5 models was the logit transformation (see Ashton 1972) of the ratio y/n , where y is the number of seropositive mice in each sample of n mice.

The first four regression models were termed “Spatial Models”, as the purpose was to identify the differences among sites that account for the difference in seroprevalence among sites (hypotheses H1, H2, H5, H6, H7, H8). Because temporal autocorrelation was assumed among samples taken throughout the season at each site, all serological data for each site were pooled, yielding one y/n proportion for each of the 101 sites. The mean trapping date of all mice trapped at each site was used to control for differences among sites due to different trapping dates.

The first analysis included no landscape variables (composition, configuration) in the model. The analysis was then repeated with the landscape variables using data from the 1 Km radius around the trapping location, then from the 2 Km and 4 Km radii. Comparing the regression results from these four models allowed an examination of the effect of the scale of sampling of the landscape. The objective was to identify the most appropriate scale, that is the scale at which landscape composition and configuration most affect the epidemiology of Sin Nombre in the mouse population (hypothesis H3).

The fifth regression model, the “Temporal Model”, was designed to examine the relation of seroprevalence to trapping date (hypothesis H4) in more detail by fitting a polynomial of julian trapping date. This model used the 392 site-date combinations as data points. Samples taken at different sites on the same date were considered as spatially independent replicates of the same date. The model included 7 dummy variables which coded for the 101 sites (*i.e.* each site

was represented by a unique combination of the 7 dummy variables), to control for differences among dates that may be attributable to a difference among sites. This coding was the best compromise for the purpose of examining the temporal component of the relationship, between using a unique dummy variable for each site and not controlling for spatial variation at all. Coding with 7 dummy variables controlled partially for spatial variation while focussing analytical power on the temporal variation.

The Spatial Regression Models

The dependent variable was the logit transformed ratio y/n for each of the 101 sites, where y is the number of seropositive mice and n is the total number of mice tested from the site. The independent variables were Date, Comp, Conf, Temp, Precip, Drought, the square of each of these variables, Strain, HPS, and Urban, where:

Date = the weighted mean trapping date, that is the mean of the julian trapping dates of all mice trapped at the site

Comp = index of landscape composition defined above

Conf = index of landscape configuration defined above

Temp = mean annual temperature

Precip = mean annual precipitation

Drought = mean annual number of months of relative drought

Strain = dummy variable, 0= east of 95° longitude, 1= west of 95° longitude

HPS = dummy variable, 0= not an HPS case, 1= HPS case

Building = dummy variable, 0= wild-trapped mice, 1= mice trapped in buildings

The Temporal Regression Model

The dependent variable was the logit transformed y/n ratio for each of the 392 site-date samples. The independent variables were the set of 7 spatial dummy variables and ten powers of the julian trapping date (i.e. date to date¹⁰). The analysis was a logistic regression with forward selection, with the 7 spatial dummy variables in the model and stepwise selection applied to the ten powers of Date.

3.3 RESULTS

Annual Within-Site Variability

A Fisher's exact test detected no significant difference in seroprevalence among years for the 5 sites for which data from 2 years were available. Data from all years were therefore pooled for subsequent analyses.

The Spatial Regression Models

The model resulting from the stepwise logistic regression procedure was generally similar for all four spatial models (see Table 3.3). Composition squared was a significant variable at all three spatial scales at which the landscape was measured ($p < .001$ to $p < .01$). Composition was significant at the 1 Km scale ($p < .002$). Configuration squared was also a significant variable at all three spatial scales at which the landscape was measured ($p = .0001$ to $p < .005$). Configuration was significant at the 2 Km scale ($p = .02$). Mean annual temperature squared was significant in all four models ($p < .001$ to $p < .005$). Building was consistently significant in all four models ($p = .0001$). Date was significant in the three models that included landscape data ($p < .01$ to $p < .02$). Mean annual precipitation was significant at the 4Km scale ($p < .03$). The shape of the relationship (sign of the parameter and curvilinearity) was consistent for each variable among all the models in which it was significant. Date squared, strain, precipitation squared, temperature, drought, drought squared, and HPS

Table 3.3. Stepwise logistic regression results and diagnostics for the four Spatial Models. The level of statistical significance (p-value) is shown for each variable which was included in the model by the stepwise selection procedure, with the sign (+/-) of the estimated parameter for that variable. n.a.= not applicable. The 1Km scale model had the best fit (highest % concordance and least significant residual Chi-squared). The overall fit of the model is significant at $p=.0001$.

Variables	Spatial scale of landscape data (radius)			
	None	1 Km	2 Km	4 Km
Intercept	-0.0001	-0.0001	-0.0001	-0.0001
composition	n.a.	-0.0011		
composition2	n.a.	0.0002	0.0075	0.0015
configuration	n.a.		-0.0200	
configuration2	n.a.	0.0044	0.0014	0.0001
date		-0.0110	-0.0148	-0.0083
date2				
strain				
precipitation				-0.0275
precipitation2				
temperature				
temperature2	-0.0007	-0.0014	-0.0021	-0.0039
drought				
drought2				
HPS				
building	0.0001	0.0001	0.0001	0.0001
% concordant	56.5	69.8	67.3	67.2
% discordant	30.7	25.4	28.6	28.5
% tied	12.8	4.8	4.1	4.3
p(Residual Chi2)	0.0918	0.7909	0.3787	0.5933

Figure 3.4 Overall fit of the model is influenced by the scale of sampling of landscape data. The model with no landscape data had the lowest predictive power (56.5% concordance) and the lowest explanatory power ($p=.09$ for residual Chi-squared). The model which included landscape data from a 1Km radius had the most predictive power (69.8 % concordance) and the most explanatory power ($p=.79$ for residual Chi-squared). Enlarging the scale of sampling beyond 1 Km did not produce better models.

were not significant variables in any of the models. The overall fit for each of the four models was significant at $p=.0001$.

Among the four spatial models, the model including landscape data from a 1 Km radius around the trapping location yielded the best regression fit (Table 3.3, Figure 3.4). This spatial model had a 69.8% concordance and a p-value of .79 for residual Chi-Square. The regression equation for this model is:

$$\ln \frac{P}{1-P} = -2.2527 - 4.1221 h + 4.6839 h^2 + 3.6159 c^2 - 0.0051 d - 0.0190 t^2 + 1.4052 u \quad \{\text{Eq. 3.1}\}$$

where

P = seroprevalence

h = landscape composition

c = landscape configuration

d = trapping date

t = mean annual temperature of a site

u = dummy variable for "building" (0 = wild-trapped mouse, 1 = house-trapped mouse)

, or,

$$P = \frac{1}{1 + e^{(-2.2527 - 4.1221 h + 4.6839 h^2 + 3.6159 c^2 - 0.0051 d - 0.0190 t^2 + 1.4052 u)}} \quad \{\text{Eq 3.2}\}$$

Shape of relationships

Figures 3.5a to 3.8a help to visualize the nature of the relationship of seroprevalence to each of the variables in the model. The seroprevalence P predicted by equation 3.2 was plotted against the value of each variable as it is varied across its range, while all other variables in the model are held constant at their mean value. These mean values are landscape composition index 0.61, landscape configuration index 0.10, julian trapping date 214.9, mean annual temperature 2.23 °C, and the dummy Urban set to 1. The data plots in Figures 3.5b to 3.8b can be difficult to interpret without the raw frequencies for the datapoints. These are presented in Appendices 1 to 4. The logistic regression procedure uses the raw frequencies (the numerator and denominator of the y/n fraction) while a single plot axis can only represent the proportion (the quotient of y/n).

For example a point with $y/n = 2/100$ would carry more weight in a regression analysis than a point with $y/n = 1/5$ but both would appear equal at $y/n = 0.20$ on a plot.

Composition

The relationship between seroprevalence and landscape composition indicated by the regression model (Equation 3.1) is illustrated in Figure 3.5. This relationship is such that, all else being equal, predicted seroprevalence is lowest at sites composed of 40% to 50% preferred habitat (intermediate composition) and highest at sites with a very small or very large proportion of preferred habitat.

Figure 3.5a Effect of landscape composition on predicted seroprevalence. With other variables in the regression model (Eq. 3.1) held constant, predicted seroprevalence is lowest (.286) at Composition = 0.44 and highest (.635) at Composition = 1. The range of predicted seroprevalence values is .349.

Figure 3.5b Distribution of seroprevalence by index of landscape composition (at the 1 Km scale) for the 101 sites. n is the number of mice tested from each site, y is the number of these which tested positive for Hantavirus antibodies.

Figure 3.6a Effect of landscape configuration on predicted seroprevalence. With other variables in the regression model (Eq. 3.1) held constant, predicted seroprevalence is lowest (.015) at Configuration = 0 and highest (.102) at Configuration = 0.75. The range of predicted seroprevalence values is .087.

Figure 3.6b Distribution of seroprevalence by index of landscape configuration (at the 1 Km scale) for the 101 sites. n is the number of mice tested from each site, y is the number of these which tested positive for Hantavirus antibodies.

Figure 3.7a Effect of julian date on predicted seroprevalence. With other variables in the regression model (Eq. 3.1) held constant, predicted seroprevalence is highest (.508) at Date = 56 and lowest (.230) at Date = 302. The range of predicted seroprevalence values is .278.

Figure 3.7b Distribution of seroprevalence by date of trapping for the 101 sites. n is the number of mice tested from each site, y is the number of these which tested positive for Hantavirus antibodies.

Figure 3.8a Effect of mean annual temperature (T) on predicted seroprevalence. With other variables in the regression model (Eq. 3.1) held constant, predicted seroprevalence is highest (.335) at $T = 0^{\circ}\text{C}$ and lowest (.070) at $T = 10^{\circ}\text{C}$. The range of predicted seroprevalence values is .265.

Figure 3.8b Distribution of seroprevalence by mean annual temperature for the 101 sites. n is the number of mice tested from each site, y is the number of these which tested positive for Hantavirus antibodies.

Configuration

The relationship between seroprevalence and landscape configuration indicated by the regression model (Equation 3.1) is illustrated in Figure 3.6. This relationship is such that, all else being equal, predicted seroprevalence is lowest in unfragmented landscapes and increases geometrically with fragmentation.

Date

The relationship between seroprevalence and date indicated by the regression model (Equation 3.1) is illustrated in Figure 3.7. This relationship is such that, all else being equal, predicted seroprevalence decreases linearly during the breeding season.

Temperature

The relationship between seroprevalence and mean annual temperature indicated by the regression model (Equation 3.1) is illustrated in Figure 3.8. This relationship is such that, all else being equal, predicted seroprevalence is highest at sites with an intermediate temperature (0°C) and lower at cooler or warmer sites.

Building

The relationship between seroprevalence and Building indicated by the regression model (Equation 3.1) is such that, all else being equal, predicted seroprevalence is higher among mice trapped in buildings than among mice trapped in the wild. The predicted values are 0.651 and 0.314, respectively,

with a range of 0.337.

Relative importance of variables

Each of the variables included in the model contribute significantly to explaining the variation in seroprevalence among sites. The intensity of the effect of each variable on seroprevalence is reflected in the range of seroprevalence values predicted for the range of each variable (Figures 3.5 to 3.8). The variables ranked in decreasing order of size of their effect on predicted seroprevalence are: Composition (.349), Building (.337), Date (.278), Temperature (.265), and Configuration (.087).

Using parameter estimates or standardized parameter estimates to compare the relative importance of variables in the model would have been inappropriate here. Firstly parameter estimates are sensitive to the size of units of measurement of each variable, and these differ arbitrarily relative to each other. Secondly, it is difficult to compare parameter estimates when some phenomena are represented by a single linear variable (*e.g.* Date), some by an exponential variable (*e.g.* Temp²), and some by more than one variable, each with a separate parameter estimate (*e.g.* Comp, Comp²).

The Temporal model

The best fit temporal polynomial model is summarized in Table 3.9. Date, date2, and date6 were all significant at $p < .005$. The model fit is significant at $p = .0001$. The regression equation for this model is:

Table 3.9 Stepwise logistic regression results and diagnostics for the Temporal Model. The analysis was a logistic regression with stepwise selection among 10 powers of date (date to date 10). The dummy variables coding for the sites(s1 to s7) were all forced in the model. Overall fit of the model is significant at $p=.0001$.

Variable	Parameter Estimate	Standardized Parameter	Significance
Intercept	3.2301		0.0043
s1	-0.1558	-0.0398	0.5352
s2	0.4904	0.1302	0.0341
s3	0.3824	0.1037	0.0584
s4	0.5093	0.1400	0.0199
s5	-0.0721	-0.0198	0.7263
s6	0.0937	0.0258	0.6543
s7	-0.2878	-0.0794	0.1591
date	-0.0890	-2.7087	0.0001
date2	0.0003	3.5619	0.0001
date6	0.0000	-0.9384	0.0027
% concordant	66.5		
% discordant	29.3		
% tied	4.2

Figure 3.10a. Seasonal trend in seroprevalence estimated by the temporal model. The fitted curve shows the temporal component of a stepwise logistic regression of seroprevalence on date and site, based on 2837 tested deermice grouped into 392 unique site-date subsamples, and is significant at $p=.0001$.

Figure 3.10b. Distribution of seroprevalence by trapping date for the 392 site-date subsamples. n is the number of mice tested from each site, y is the number of these which tested positive for Hantavirus antibodies.

$$\ln \frac{P}{1-P} = 3.2301 + 0.1558 s_1 + 0.4904 s_2 + 0.3824 s_3 + 0.5093 s_4 + 0.0721 s_5 + 0.0937 s_6 - 0.2878 s_7 - 0.0890 d + 0.00028 d^2 + 7.7 \times 10^{-15} d^6 \quad \{\text{Eq. 3.3}\}$$

where

P = seroprevalence

$s_1 \dots s_7$ = dummy variables for site

$d \dots d^6$ = powers of julian trapping date

The temporal component of this regression equation is illustrated in Figure 3.10a. The relationship is bimodal with a large peak in February and a small peak in September. Minimum predicted seroprevalence is in June and November.

In summary, the hypotheses tested are:

- H1. Seroprevalence is significantly related to the amount of preferred habitat in the landscape.
- H2. Seroprevalence is significantly related to the extent of fragmentation of the preferred habitat in the landscape.
- H3. The relation of seroprevalence to landscape variables (H2 and H3) is dependent upon scale. A 1 Km radius is the most appropriate scale.
- H4. Seroprevalence is significantly related to julian date, the relationship is bimodal, with peaks in March and August and troughs in June and November.

- H5. Seroprevalence is significantly related to mean annual temperature.
- H6. These data show no significant relation of seroprevalence to the strain of Sin Nombre.
- H7. Seroprevalence is significantly related to whether mice were trapped in the wild or in buildings.
- H8. These data show no significant relation of seroprevalence to whether or not mice were tested in connection with a human Hantavirus Pulmonary Syndrome case.
- H9. These data show no significant within-site variability in seroprevalence among years.

3.4 DISCUSSION

It is important when considering these results to remember that that these are regressions. A statistically significant regression does not establish causation, or even the existence of a causal relationship. Regression analysis does provide a precise and well defined form of observation, and observation is the necessary first step in the scientific method. The purpose of an observational study such as this one should always be to provide the best possible examination of the existing data, that is to identify trends, patterns, and apparent relationships which together with theoretical knowledge of the system can form the basis for formulating optimal hypotheses to be tested experimentally. This discussion considers the results in this context.

Landscape composition

The result that seroprevalence is higher in landscapes with a more extreme composition index (either mostly preferred habitat or mostly non-preferred habitat) indicates in other words that seroprevalence can reach higher levels in more uniform landscapes (uniform in terms of composition). For a habitat generalist such as *P. maniculatus*, the proportion of marginally preferred habitat *per se* may be less important than the uniformity of habitats. Given a highly uniform landscape (as defined by a mouse) an individual moving through the landscape to meet its needs may move more randomly. An individual presented with a highly nonuniform landscape on the other hand, must make choices between the various landscape elements (patches). To a mouse, the choices are undoubtedly much more subtle than the dichotomous preferred /

non-preferred classification used here, and choices are likely to differ among individuals. The lower seroprevalence observed in nonuniform landscapes (intermediate composition index) may be the result of less contact among individuals choosing different habitat elements. This is different from the process we would expect with a habitat specialist in the same conditions.

Landscape configuration

This same process may be responsible for the observed relationship with fragmentation. When a landscape is minimally fragmented, for example large areas of each of two habitat types, there may be less contact between individuals choosing each of the habitat types. When a landscape is maximally fragmented, with much intermingling of the habitat types, contact between individuals choosing each habitat type may be more frequent, which would permit a higher seroprevalence in the population sampled in the landscape as a whole.

This points out the utility of sampling at the patch scale, because patterns of seroprevalence at the landscape scale may well result from dynamics within or among patches in addition to dynamics among individuals. This was not done here because trapping records generally were not precise enough to assign trapping locations to specific patches within each landscape. More generally, these results at the landscape scale suggest hypotheses about mechanisms operating at the patch scale in addition to the scale of individual mice. For example do patches or patch types operate as separate epidemiological units and does this result in differing seroprevalence among patches or patch types?

Spatial Scale

These results identify the scale of a one kilometre radius as the appropriate scale for testing hypotheses about effects of landscape structure on the epidemiology of Sin Nombre in the deer mouse population. This provides a basis for choosing to design future experiments and/or sampling schemes at this scale. In many cases however, it would likely be more informative to repeat the method employed here, given the relative ease of simultaneous sampling and analysis at several scales. One kilometre may be a more accurate estimate of the minimum distance required for spatial independence than is the 2 Km distance which I used in defining sites, but any dichotomous definition of spatial independence based on a minimum separation distance is a pragmatic simplification. The extent of spatial independence between sites is a continuous function of the distance between them. Further, this function varies with the characteristics of movement of individuals across the landscape, which will differ among landscapes.

Seasonal change

We know that the total mouse population increases during the breeding season from spring to autumn, while I found seroprevalence to decrease during this period (Figure 3.7a). This indicates that the number of uninfected individuals increases more rapidly than does the number of infected individuals during this period. This could be due to any combination of:

- higher emigration of infected individuals than of noninfected individuals
- lower immigration of infected individuals than of noninfected individuals

- higher mortality of infected individuals than of noninfected individuals
- lower prevalence among offspring than among adults

Of these, the first two are contradictory when we consider the population at the largest scale. If more infected individuals emigrate from a location more, not fewer, will immigrate at a neighbouring location.

Differential mortality of infected deermice has not been directly investigated, but seems unlikely. The rodent hosts of hantaviruses develop a persistent or chronic infection but with no apparent disease symptoms (LeDuc 1987, Lee and van der Groen 1989). Also, hantaviruses have undergone co-evolution with their rodent hosts worldwide, and there is evidence that Sin Nombre has co-evolved with *Peromyscus maniculatus* (Chizhikov *et al.* 1996, Morzunov *et al.* 1996). A virus is rarely lethal to its primary host since this characteristic would be selected against as a direct result of this differential mortality of the host. This is particularly true of a co-evolved virus-host system (see any introductory virology text, *e.g.* Dimmock and Primrose 1988).

The temporal pattern makes sense if adults are more likely to be infected by other adults than are offspring to be infected by their mother. This may well be the case if “vertical transmission” and “perinatal transmission” are low and if resistance of neonates to respiratory infection in the nest is high, for example due to maternal antibodies. If reproductive rate of infected and uninfected females is assumed to be the same then anything less than 100% infection of the offspring of infected females would result in a decreasing seroprevalence

during the reproductive season. During the overwintering period, communal nesting among adults (including those which were unsusceptible juveniles earlier in the season) would permit high transmission rates and build up seroprevalence to a higher level again the following spring. Field studies using nest boxes could resolve these questions, but antigen testing would likely be required in addition to antibody screening because neonates may be seropositive due to maternal antibodies but not be infected with the virus.

The bimodal seasonal trend shown in figure 3.10a suggests a possible link between seroprevalence and breeding cycles, with the arrival of litters in the population potentially coinciding with and/or being responsible for changes in population seroprevalence. Bimodal peaks in breeding in long-season environments has been documented in the *Peromyscus* genus (Millar 1989, p.189), but not in *Peromyscus maniculatus* (see Drickamer 1978). The length of the breeding season of *P. maniculatus* varies significantly with latitude and longitude across North America (Millar, 1989, p.190) however, and breeding seasons recorded in Canada range from 5 weeks in June and July in Yukon (Gilbert and Krebs, 1981) to 40 weeks, from February to November, in southwestern British Columbia (Redfield *et al.*, 1977). Since the temporal model used data from across Canada, and therefore with a wide range of asynchronous breeding seasons, it is not possible to relate the fitted curve to the timing of litters. It would be well worth examining this timing in a particular population in detail, since individual females almost always produce two or three litters in a breeding season regardless of the length of the breeding season (Millar, 1989, p.189). A bimodal seasonal trend in seroprevalence in a

specific population is a possibility.

It should be noted that the predictive power for the temporal model was less than that of the best spatial model. While the temporal model does offer a more detailed examination of the seasonal trend at these sites, it does not provide any more power to predict seroprevalence.

Temperature

The pattern of maximum seroprevalence at sites with an intermediate mean annual temperature (MAT) suggests that the optimal MAT for Sin Nombre is within the range represented by these 101 sites. This is unexpected because higher seroprevalence rates have been reported from warmer regions in the United States. It is possible that these data represent a local rather than global optimum MAT for Sin Nombre.

Buildings

As discussed in 3.1, the dynamics of subpopulations in buildings may be quite different from those of wild subpopulations. The very conditions that probably attract deer mice into buildings, a concentrated food supply and moderate microclimate, may increase transmission of Sin Nombre by increasing mouse population density and out-of-host survival of the virus, respectively.

4 - GENERAL DISCUSSION AND CONCLUSIONS

Several authors have observed the similarities between epidemiological models and metapopulation models (*e.g.* Nee 1994). The link between landscape structure and various metapopulation models is a popular area of research (*e.g.* Fahrig and Merriam 1994, Henein and Merriam 1990). This study has attempted to integrate these three concepts and clarify the relationships among them.

The distribution of a virus in a subdivided host population is probably influenced more directly by population structure, including population connectivity, rather than by landscape structure *per se* (Figure 1.1). Population connectivity is more easily (but less accurately) implied from landscape structure than measured directly. In the absence of direct measurement of host population connectivity, I have demonstrated some of the possible effects of landscape structure on the distribution of a virus in its host population. I have found that landscape composition can be the most important among several ecological variables affecting the distribution of Sin Nombre hantavirus in the deer mouse population. I have also found that landscape configuration can be a significant factor. An applied epidemiological model which would not account for landscape structure could overlook important aspects of the system.

When the host population is spatially subdivided, epidemiological models must consider not only the demographic structure of the host population, but its spatial structure as well, and this can in some cases be effectively inferred from

landscape structure.

Conclusions

- Landscape structure can affect the pattern of distribution of a virus in its host population.
- Landscape composition can have a more important effect on the distribution of the virus than other ecological variables such as climate and seasonal change.
- The choice of spatial scale can affect the usefulness of landscape data in epidemiological models.
- Epidemiological models with a spatially structured host population can benefit from the explicit consideration of landscape structure.

References

- Ahlbom A. 1993. Biostatistics for epidemiologists. Boca Raton, Florida, Lewis Publishers.
- Artsob H. Dick D. Thompson I. and Barker I. unpublished. Identification of Sin Nombre hantavirus in deer mice in Algonquin Park, Ontario, Canada.
- Ashton W.D. 1972. The logit transformation, with special reference to its uses in bioassay. Griffin's statistical monographs & courses, no. 32. Charles Griffin and Company, London.
- Baker R.H. 1968. Habitats and distribution, in King J.A. (editor) Biology of *Peromyscus* (RODENTIA). Special publication no.2. The American Society of Mammalogists. 593p.
- Barry R.E. Jr., Botje M.A. and Grantham L.B. 1984. Vertical stratification of *Peromyscus leucopus* and *P. maniculatus* in southwestern Virginia. Journal of Mammalogy 65:145-148.
- Becker N.G. 1989. Analysis of infectious disease data. Chapman, London.
- Blair W.F. 1950. Ecological factors in speciation of *Peromyscus*. Evolution 4: 253-275.
- Boorman S.A. and Levitt P.R. 1973. Group selection on the boundary of a stable population. Theoretical Population Biology 4:85-128.
- Buckner C.A. and Shure D.J. 1985. The response of *Peromyscus* to forest opening size in the southern Appalachian Mountains. Journal of Mammalogy 66: 299-307.
- Carleton M.D. 1989. Systematics and evolution, in Kirkland G.L. Jr. and Layne J.N. (editors) Advances in the study of *Peromyscus* (RODENTIA). Texas

- Tech University Press. Lubbock Texas. 366p.
- Centres for Disease Control, 1994, Enzyme Immunoassay for Detection of IgG Antibody to Hantavirus in Rodents.
- Childs J.E. Ksiazek T.G. Spiropoulou C.F. Krebs J.W. Morzunov S. Maupin G.O. Gage K.L. Rollin P.E. Sarisky J. Enscoe R.E. Frey J.K. Peters C.J. Nichol S.T. 1994. Serologic and Genetic Identification of *Peromyscus maniculatus* as the Primary Rodent Reservoir for a New Hantavirus in the Southwestern United States. *The Journal of Infectious Diseases*, 169:1271-1280.
- Chizhikov V.E. Hörling J. Lundkvist Å. Jonsson M. Ivanov L.I. Niklasson B. Tkachenko E.A. Peters C.J. Nichol S.T. 1996. Genetic and serologic analysis of hantaviruses isolated from *Microtus fortis* trapped in far eastern Russia. American Society for Virology 15th annual meeting, July 13-17, 1996, London Ontario.
- Cunliffe-Beamer T.L., 1983, Biomethodology and Surgical Techniques, in Foster H.L., Small J.D., and Fox J.G., *The Mouse in Biomedical research Volume III Normative Biology, Immunology, and Husbandry*. Academic Press, New York.
- Diffendorfer J.E., Gaines M.S., Holt R.D. 1995. Habitat fragmentation and movements of three small mammals (*Sigmodon*, *Microtus*, and *Peromyscus*). *Ecology* 76(3): 827-839.
- Dimmock N.J. and Primrose S.B. 1988. Introduction to modern virology, 3rd edition. Blackwell Scientific Publications. Oxford 347p.
- Dobbyn, J., 1994, Atlas of the Mammals of Ontario. 120p., Federation of Ontario Naturalists, Don Mills Ontario.

- Drickamer L.C. 1978. Annual reproduction patterns in populations of two sympatric species of *Peromyscus*. *Behavioral Biology* 23:405-408.
- Ecoregions Working Group. 1989. Ecoclimatic regions of Canada, first approximation. *Ecological Land Classification Series No. 23*. Environment Canada.
- Fahrig L. and Merriam G. 1994. Conservation of fragmented populations. *Conservation Biology* 8(1): 50-59.
- Forman R. T. T. and Godron M. 1986. *Landscape ecology*. Wiley, New York.
- Gilbert B.S., Krebs C.J. 1981. Effects of extra food on *Peromyscus* and *Clethrionomys* populations in the southern Yukon. *Oecologia* 51: 326-331.
- Gustafson E.J. and Parker G.R., 1994. Using an index of habitat patch proximity for landscape design, *Landscape and Urban Planning* 29:117-130.
- Hanski I. 1991. Single-species metapopulation dynamics: Concepts, models and observations. *Biological Journal of the Linnean Society* 42:17-38.
- Hanski I. and Gilpin M. 1991. Metapopulation dynamics: Brief history and conceptual domain. *Biological Journal of the Linnean Society* 42:3-16.
- Henein K. and Merriam G. 1990. The elements of connectivity where corridor quality is variable. *Landscape Ecology* 4:157-170.
- Hooper E.T. 1968. Classification, in King J.A. (editor) *Biology of Peromyscus (RODENTIA)*. Special publication no.2. The American Society of Mammalogists. 593p.
- Hurlbert S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54(2): 187-211.
- Kareiva P. and Wennergren U. 1995. Connecting landscape patterns to

- ecosystem and population processes. *Nature* 373:299-302.
- Kaufman D.W. and Kaufman G.A. 1989. Population biology, in Kirkland G.L. Jr. and Layne J.N. (editors) *Advances in the study of Peromyscus* (RODENTIA). Texas Tech University Press. Lubbock Texas. 366p.
- LeDuc J. 1987. Epidemiology of hantaan and related viruses. *Laboratory Animal Science* 37: 413-418.
- Lee H.W. and van der Groen G. 1989. Hemorrhagic fever with renal syndrome. *Progress in Medical Virology* 36: 62-102
- Levins R. 1970. Extinction. in Gerstenhaber M., ed. *Lectures on mathematics in the life sciences*. Vol. 2: 76-107. American Mathematical Society, Providence, Rhode Island.
- MacArthur R.H. and Wilson W.O. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey.
- Martinsson B. Hansson L. and Angelstam P. 1993. Small mammal dynamics in adjacent landscapes with varying predator communities. *Ann. Zool. Fennici* 30: 31-42.
- McCauley D.E. 1995. Effects of population dynamics on genetics in mosaic landscapes. in Hansson, Fahrig, and Merriam, eds. *Mosaic landscapes and ecological processes*. Chapman and Hall, London. 356p.
- Merriam G. and Lanoue A. 1990. Corridor use by small mammals: field measurements for three experimental types of *Peromyscus leucopus*. *Landscape Ecology* 4:123-131.
- Merriam H.G., Kozakiewicz M., Tsuchiya E., Hawley K. 1989. Barriers as boundaries for metapopulations and demes of *Peromyscus leucopus* in farm landscapes. *Landscape Ecology* 2(4): 227-235.

- Millar J.S. 1989. Reproduction and development, in Kirkland G.L. Jr. and Layne J.N. (editors) *Advances in the study of Peromyscus* (RODENTIA). Texas Tech University Press. Lubbock Texas. 366p.
- Mills J.N., Yates T.L., Childs J.E., Parmenter R.R., Ksiazek T.G., Rollin P.E., Peters C.J., 1995. Guidelines for working with rodents potentially infected with hantavirus. *Journal of Mammalogy* 76(3):716-722.
- Morris D.W. 1992. Scales and costs of habitat selection in heterogeneous landscapes. *Evolutionary Ecology* 6:412-432
- Morzunov S.P. St Jeor S.C. Rowe J. Artsob H. Ksiazek T.G. Peters C.J. Nichol S.T. 1996. North American *Peromyscus*-borne hantaviruses: co-evolution with specific rodent hosts and geographic domination. American Society for Virology 15th annual meeting, July 13-17, 1996, London Ontario.
- Naveh Z. and Lieberman A. S. 1994. *Landscape ecology*. 2nd ed. Springer-Verlag, New York. 360 p.
- Nee S. 1994. How populations persist. *Nature* 367:123-124.
- Nichol S.T. Spiropoulou C.F. Morzunov S. Rollin P.E. Ksiazek T.G. Feldmann H. Sanchez A. Childs J. Zaki S. Peters C.J. 1993. Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* 262: 914-917.
- Nicholson A.J., 1941, The Homes and Social Habits of the Wood-Mouse (*Peromyscus leucopus noveboracensis*) in Southern Michigan. *The American midland Naturalist*, 25:196-223.
- O'Neill R.V., Krummel J.R., Gardner R.H., Sugihara G., Jackson B., DeAngelis D.L., Milne B.T., Turner M.G., Zygmunt B., Christensen S.W., Dale V.H., Graham R.L., 1988, Indices of landscape pattern, *Landscape Ecology*

1(3):153-162.

- Olfert E.D., Cross B.M., McWilliam A.A., 1993, Guide to the use and care of experimental animals, volume I, 2nd edition, Canadian Council on Animal Care.
- Parren S.G. and Capen D.E. 1985. Local distribution and coexistence of two species of *Peromyscus* in Vermont. *Journal of Mammalogy* 66:36-44
- Pielou E.C., 1975, *Ecological Diversity*, 165p., Wiley-Interscience publications, New York
- Pielou E.C., 1977, *Mathematical Ecology*, 385p., Wiley-Interscience publications, New York
- Pulliam H.R. 1988. Sources, sinks, and population regulation. *American Naturalist* 132:652-661.
- Redfield J.A., Krebs C.J., Taitt M.J. 1977. Competition between *Peromyscus maniculatus* and *Microtus townsendii* in grasslands of coastal British Columbia. *Journal of Animal Ecology* 46: 607-616.
- Reynolds H. unpublished data. 1994 and 1995 trapping records for deer mice from the Suffield proposed National Wildlife Area.
- Rowe J.E. St Jeor S.C. Riolo J. Otteson E.W. Monroe M.C. Henderson W.W. Ksiazek T.G. Rollin P.E. Nichol S.T. 1995. Coexistence of several novel Hantaviruses in rodents indigenous to North America. *Virology* 213: 122-130.
- Sekgororoane G.B. 1995. Relative abundance, richness, and diversity of small mammals at induced forest edges. *Canadian Journal of Zoology - Revue Canadienne de Zoologie*. 73(8): 1432-1437.
- Smith, D.A., 1993, Identifying White-footed and Deer Mice in Ontario. *Mammal*

Atlas News, 3(1):2-7

- Song J.W. Baek L.J. Gajdusek D.C. Yanagihara R. Gavrilovskaya I. Luft B.J. MacKow E.R. Hjelle B. 1994. Isolation of pathogenic hantavirus from white white-footed mouse. *Lancet* 334: 1637.
- Stah C.D. 1980. Vertical nesting distribution of two species of *Peromyscus* under experimental conditions. *Journal of Mammalogy* 61:141-143.
- Stickel L.F. 1968. Home range and travels. in King J.A. (editor) *Biology of Peromyscus (RODENTIA)*. Special publication no.2. The American Society of Mammalogists. 593p.
- Strickland D. and Rutter R.J. 1994. Mammals of Algonquin provincial park. Ontario Ministry Of Natural Resources. 46p.
- Taylor P.D. Fahrig L. Henein K. Merriam G. 1993. Connectivity is a vital element of landscape structure. *Oikos* 68(3): 571-573.
- Teferi T. and Millar J.S. 1993. Long distance homing by the deer mouse, *Peromyscus maniculatus*. *Canadian Field Naturalist* 107(1): 109-111.
- Terman C.R. 1968. Population Dynamics, in King J.A. (editor) *Biology of Peromyscus (RODENTIA)*. Special publication no.2. The American Society of Mammalogists. 593p.
- Wecker S.C. 1963. The role of early experience in habitat selection by the prairie deer mouse, *Peromyscus maniculatus bairdi*. *Ecological Monographs* 33:307-325.
- Wegner J. and Merriam G. 1990. Use of spatial elements in a farmland mosaic by a woodland rodent. *Biological Conservation* 54: 263-276.
- Wegner J. 1994. Ecological landscape variables for monitoring and management of forest biodiversity in Canada. Report to the Canadian

Ministry of Natural Resources by GM Group, Ecological Land
Management, Manotick Ontario.

Wolff J.O. and Hurlbutt B. 1982. Day refuges of *Peromyscus leucopus* and
Peromyscus maniculatus. *Journal of Mammalogy* 63:666-668.

Wolff J.O. 1989. Social Behaviour, in Kirkland G.L. Jr. and Layne J.N. (editors)
Advances in the study of *Peromyscus* (RODENTIA). Texas Tech
University Press. Lubbock Texas.

Appendix 1. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape composition index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Comp (1Km) is landscape composition at the 1 Kilometre scale.

Site No	y	n	y/n	Comp (1Km)
26	2	36	0.06	0.00
28	0	8	0.00	0.00
29	1	15	0.07	0.00
30	0	3	0.00	0.00
31	1	64	0.02	0.00
32	0	80	0.00	0.00
33	0	18	0.00	0.00
34	2	30	0.07	0.00
36	1	17	0.06	0.00
37	1	25	0.04	0.00
38	1	23	0.04	0.00
39	0	21	0.00	0.00
40	0	21	0.00	0.00
41	0	24	0.00	0.00
42	3	42	0.07	0.00
43	0	23	0.00	0.00
44	0	25	0.00	0.00
46	3	24	0.12	0.00
66	8	19	0.42	0.04
89.1	0	4	0.00	0.06
45	3	60	0.05	0.08
80	0	2	0.00	0.15
25	1	41	0.02	0.16
35	1	17	0.06	0.16
68	0	55	0.00	0.26
17	0	11	0.00	0.29
81	0	2	0.00	0.39
101	0	23	0.00	0.39
105	0	19	0.00	0.39
102	0	7	0.00	0.41
50	0	3	0.00	0.41
2	0	96	0.00	0.44
51	0	11	0.00	0.46
52	0	2	0.00	0.49
85	0	80	0.00	0.50
91	0	5	0.00	0.54
70	0	8	0.00	0.56
67	0	2	0.00	0.57

Appendix 1. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape composition index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Comp (1Km) is landscape composition at the 1 Kilometre scale.

8	2	24	0.08	0.58
78	1	64	0.02	0.59
103	0	36	0.00	0.59
106	0	17	0.00	0.59
90	0	3	0.00	0.60
104	0	6	0.00	0.61
48	0	55	0.00	0.62
14	2	43	0.05	0.71
61	0	3	0.00	0.71
73	0	2	0.00	0.71
83	0	6	0.00	0.71
89.2	2	9	0.22	0.74
49	12	54	0.22	0.75
84	0	160	0.00	0.75
23	0	48	0.00	0.77
13	0	2	0.00	0.79
18	3	26	0.12	0.79
47	0	151	0.00	0.79
95	0	20	0.00	0.80
53	0	3	0.00	0.81
79	1	13	0.08	0.81
75	0	6	0.00	0.85
60.1	14	38	0.37	0.86
24	0	4	0.00	0.88
89.3	0	5	0.00	0.88
86	0	4	0.00	0.89
87	0	3	0.00	0.89
7	0	12	0.00	0.90
94	0	16	0.00	0.90
6	0	1	0.00	0.91
22	0	18	0.00	0.91
77	0	8	0.00	0.91
11	3	103	0.03	0.93
82	1	66	0.02	0.93
54	8	111	0.07	0.94
58	0	3	0.00	0.94
71	0	2	0.00	0.94
56	0	2	0.00	0.95
16	0	36	0.00	0.96
20	0	6	0.00	0.96

Appendix 1. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape composition index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Comp (1Km) is landscape composition at the 1 Kilometre scale.

21	0	27	0.00	0.96
55	0	1	0.00	0.96
65	4	13	0.31	0.96
3	0	9	0.00	0.98
9	0	9	0.00	0.98
12	1	31	0.03	0.98
69	0	34	0.00	0.98
76	0	3	0.00	0.98
1	0	105	0.00	0.99
15	12	97	0.12	0.99
4	2	36	0.06	1.00
5	0	14	0.00	1.00
10	4	64	0.06	1.00
19	0	6	0.00	1.00
27	2	58	0.03	1.00
62	3	3	1.00	1.00
64	6	43	0.14	1.00
72	0	5	0.00	1.00
74	0	1	0.00	1.00
88	0	1	0.00	1.00
92	0	20	0.00	1.00
93	9	72	0.12	1.00
96	0	20	0.00	1.00

Appendix 2. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape configuration index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Conf (1Km) is landscape configuration at the 1 Kilometre scale.

Site No	y	n	y/n	Conf (1Km)
1	0	105	0.00	0.00
2	0	96	0.00	0.00
4	2	36	0.06	0.00
5	0	14	0.00	0.00
10	4	64	0.06	0.00
14	2	43	0.05	0.00
18	3	26	0.12	0.00
19	0	6	0.00	0.00
23	0	48	0.00	0.00
26	2	36	0.06	0.00
27	2	58	0.03	0.00
28	0	8	0.00	0.00
29	1	15	0.07	0.00
30	0	3	0.00	0.00
31	1	64	0.02	0.00
32	0	80	0.00	0.00
33	0	18	0.00	0.00
34	2	30	0.07	0.00
36	1	17	0.06	0.00
37	1	25	0.04	0.00
38	1	23	0.04	0.00
39	0	21	0.00	0.00
40	0	21	0.00	0.00
41	0	24	0.00	0.00
42	3	42	0.07	0.00
43	0	23	0.00	0.00
44	0	25	0.00	0.00
45	3	60	0.05	0.00
46	3	24	0.12	0.00
56	0	2	0.00	0.00
62	3	3	1.00	0.00
64	6	43	0.14	0.00
66	8	19	0.42	0.00
72	0	5	0.00	0.00
74	0	1	0.00	0.00
75	0	6	0.00	0.00
77	0	8	0.00	0.00
80	0	2	0.00	0.00
84	0	160	0.00	0.00

Appendix 2. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape configuration index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Conf (1Km) is landscape configuration at the 1 Kilometre scale.

85	0	80	0.00	0.00
88	0	1	0.00	0.00
89.1	0	4	0.00	0.00
92	0	20	0.00	0.00
93	9	72	0.12	0.00
95	0	20	0.00	0.00
96	0	20	0.00	0.00
68	0	55	0.00	0.03
8	2	24	0.08	0.03
17	0	11	0.00	0.04
70	0	8	0.00	0.04
90	0	3	0.00	0.04
60.1	14	38	0.37	0.05
89.3	0	5	0.00	0.05
83	0	6	0.00	0.06
61	0	3	0.00	0.06
81	0	2	0.00	0.06
94	0	16	0.00	0.06
25	1	41	0.02	0.07
67	0	2	0.00	0.07
52	0	2	0.00	0.08
11	3	103	0.03	0.08
82	1	66	0.02	0.08
78	1	64	0.02	0.09
51	0	11	0.00	0.09
50	0	3	0.00	0.10
54	8	111	0.07	0.10
58	0	3	0.00	0.10
101	0	23	0.00	0.10
91	0	5	0.00	0.10
104	0	6	0.00	0.10
103	0	36	0.00	0.12
47	0	151	0.00	0.13
73	0	2	0.00	0.13
35	1	17	0.06	0.14
53	0	3	0.00	0.14
102	0	7	0.00	0.14
48	0	55	0.00	0.14
89.2	2	9	0.22	0.15
16	0	36	0.00	0.17

Appendix 2. Frequencies of seropositive mice from 101 sites across Canada sorted by landscape configuration index. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Conf (1Km) is landscape configuration at the 1 Kilometre scale.

21	0	27	0.00	0.17
55	0	1	0.00	0.17
65	4	13	0.31	0.17
79	1	13	0.08	0.22
87	0	3	0.00	0.22
105	0	19	0.00	0.23
3	0	9	0.00	0.25
9	0	9	0.00	0.25
12	1	31	0.03	0.25
13	0	2	0.00	0.25
22	0	18	0.00	0.25
69	0	34	0.00	0.25
76	0	3	0.00	0.25
49	12	54	0.22	0.26
106	0	17	0.00	0.29
86	0	4	0.00	0.32
7	0	12	0.00	0.33
71	0	2	0.00	0.42
15	12	97	0.12	0.50
24	0	4	0.00	0.50
6	0	1	0.00	0.58
20	0	6	0.00	0.75

Appendix 3. Frequencies of seropositive mice from 101 sites across Canada sorted by julian date. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Date is the mean trapping date for all mice trapped at each site.

Site No	y	n	y/n	Date
61	0	3	0.00	56
58	0	3	0.00	59
66	8	19	0.42	76
55	0	1	0.00	79
56	0	2	0.00	82
70	0	8	0.00	105
72	0	5	0.00	105
76	0	3	0.00	105
87	0	3	0.00	105
65	4	13	0.31	128
48	0	55	0.00	133
49	12	54	0.22	147
47	0	151	0.00	150
8	2	24	0.08	154
5	0	14	0.00	160
12	1	31	0.03	160
31	1	64	0.02	161
36	1	17	0.06	162
60.1	14	38	0.37	163
3	0	9	0.00	167
37	1	25	0.04	168
32	0	80	0.00	170
38	1	23	0.04	171
41	0	24	0.00	171
43	0	23	0.00	171
84	0	160	0.00	176
42	3	42	0.07	178
39	0	21	0.00	182
40	0	21	0.00	183
34	2	30	0.07	184
79	1	13	0.08	188
73	0	2	0.00	197
80	0	2	0.00	197
88	0	1	0.00	197
90	0	3	0.00	197
91	0	5	0.00	197
44	0	25	0.00	197
46	3	24	0.12	197
64	6	43	0.14	202
27	2	58	0.03	202

Appendix 3. Frequencies of seropositive mice from 101 sites across Canada sorted by julian date. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Date is the mean trapping date for all mice trapped at each site.

89.1	0	4	0.00	202
35	1	17	0.06	211
93	9	72	0.12	213
33	0	18	0.00	214
78	1	64	0.02	215
94	0	16	0.00	215
45	3	60	0.05	220
89.3	0	5	0.00	225
28	0	8	0.00	226
29	1	15	0.07	227
30	0	3	0.00	227
74	0	1	0.00	228
102	0	7	0.00	229
75	0	6	0.00	230
4	2	36	0.06	232
82	1	66	0.02	233
26	2	36	0.06	234
92	0	20	0.00	235
95	0	20	0.00	235
96	0	20	0.00	235
25	1	41	0.02	242
85	0	80	0.00	245
22	0	18	0.00	245
89.2	2	9	0.22	249
105	0	19	0.00	249
53	0	3	0.00	251
62	3	3	1.00	251
81	0	2	0.00	252
18	3	26	0.12	254
6	0	1	0.00	254
86	0	4	0.00	254
9	0	9	0.00	256
7	0	12	0.00	256
101	0	23	0.00	258
106	0	17	0.00	260
103	0	36	0.00	261
15	12	97	0.12	263
69	0	34	0.00	263
83	0	6	0.00	265
1	0	105	0.00	266
52	0	2	0.00	266

Appendix 3. Frequencies of seropositive mice from 101 sites across Canada sorted by julian date. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Date is the mean trapping date for all mice trapped at each site.

2	0	96	0.00	268
51	0	11	0.00	270
13	0	2	0.00	272
17	0	11	0.00	274
77	0	8	0.00	274
104	0	6	0.00	276
67	0	2	0.00	279
50	0	3	0.00	279
11	3	103	0.03	281
21	0	27	0.00	283
10	4	64	0.06	285
24	0	4	0.00	286
14	2	43	0.05	289
16	0	36	0.00	289
23	0	48	0.00	289
20	0	6	0.00	291
19	0	6	0.00	292
68	0	55	0.00	295
71	0	2	0.00	300
54	8	111	0.07	302

Appendix 4. Frequencies of seropositive mice from 101 sites across Canada sorted by mean annual temperature. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Temp is mean annual temperature for each site.

Site No	y	n	y/n	Temp
93	9	72	0.12	-3.3
94	0	16	0.00	-3.3
91	0	5	0.00	-2.3
68	0	55	0.00	-0.6
69	0	34	0.00	-0.6
78	1	64	0.02	-0.6
82	1	66	0.02	-0.6
84	0	160	0.00	-0.6
89.1	0	4	0.00	-0.6
89.2	2	9	0.22	-0.6
89.3	0	5	0.00	-0.6
92	0	20	0.00	-0.6
95	0	20	0.00	-0.6
96	0	20	0.00	-0.6
47	0	151	0.00	0.0
48	0	55	0.00	0.0
49	12	54	0.22	0.0
73	0	2	0.00	0.1
74	0	1	0.00	0.1
75	0	6	0.00	0.1
76	0	3	0.00	0.1
77	0	8	0.00	0.1
79	1	13	0.08	0.1
80	0	2	0.00	0.1
81	0	2	0.00	0.1
83	0	6	0.00	0.1
86	0	4	0.00	0.1
88	0	1	0.00	0.1
90	0	3	0.00	0.1
24	0	4	0.00	0.6
53	0	3	0.00	1.6
54	8	111	0.07	1.6
55	0	1	0.00	1.6
56	0	2	0.00	1.6
58	0	3	0.00	1.6
60.1	14	38	0.37	1.6
61	0	3	0.00	1.6
62	3	3	1.00	1.6
64	6	43	0.14	1.6

Appendix 4. Frequencies of seropositive mice from 101 sites across Canada sorted by mean annual temperature. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Temp is mean annual temperature for each site.

65	4	13	0.31	1.6
66	8	19	0.42	1.6
23	0	48	0.00	1.7
50	0	3	0.00	1.8
15	12	97	0.12	2.3
18	3	26	0.12	2.3
19	0	6	0.00	2.3
20	0	6	0.00	2.3
21	0	27	0.00	2.3
22	0	18	0.00	2.3
25	1	41	0.02	2.4
26	2	36	0.06	2.4
28	0	8	0.00	2.4
29	1	15	0.07	2.4
30	0	3	0.00	2.4
31	1	64	0.02	2.4
32	0	80	0.00	2.4
33	0	18	0.00	2.4
34	2	30	0.07	2.4
35	1	17	0.06	2.4
36	1	17	0.06	2.4
37	1	25	0.04	2.4
38	1	23	0.04	2.4
39	0	21	0.00	2.4
40	0	21	0.00	2.4
41	0	24	0.00	2.4
42	3	42	0.07	2.4
43	0	23	0.00	2.4
44	0	25	0.00	2.4
46	3	24	0.12	2.4
87	0	3	0.00	2.6
3	0	9	0.00	4.2
4	2	36	0.06	4.2
5	0	14	0.00	4.2
8	2	24	0.08	4.2
10	4	64	0.06	4.2
12	1	31	0.03	4.2
14	2	43	0.05	4.2
16	0	36	0.00	4.2
101	0	23	0.00	4.2
102	0	7	0.00	4.2

Appendix 4. Frequencies of seropositive mice from 101 sites across Canada sorted by mean annual temperature. n is the number of mice tested from each site, y is the number of these which tested positive for hantavirus antibodies, Temp is mean annual temperature for each site.

103	0	36	0.00	4.2
104	0	6	0.00	4.2
105	0	19	0.00	4.2
106	0	17	0.00	4.2
51	0	11	0.00	4.5
52	0	2	0.00	4.5
67	0	2	0.00	4.5
70	0	8	0.00	4.5
71	0	2	0.00	4.5
72	0	5	0.00	4.5
1	0	105	0.00	5.4
2	0	96	0.00	5.4
6	0	1	0.00	5.4
7	0	12	0.00	5.4
9	0	9	0.00	5.4
11	3	103	0.03	5.4
13	0	2	0.00	5.4
17	0	11	0.00	5.4
85	0	80	0.00	6.2
45	3	60	0.05	8.3
27	2	58	0.03	10.0

Appendix 5. Complete dataset for the spatial regression models.

Site #	Y	N	Y/N	Date	Comp(4Km)	Conf(4Km)	Comp(2Km)	Conf(2Km)
1	0	105	0.000	265.8	0.6843	1.0000	0.8313	1.0000
2	0	96	0.000	268.1	0.1812	1.4375	0.2062	1.7500
3	0	9	0.000	167	0.9156	1.0000	0.9688	1.0000
4	2	36	0.056	232.5	1.0000	1.0000	1.0000	1.0000
5	0	14	0.000	159.7	0.8781	1.0000	0.9438	1.0000
6	0	1	0.000	254	0.9187	1.0000	0.8438	1.5000
7	0	12	0.000	256.4	0.9062	1.0000	0.9000	1.0000
8	2	24	0.083	153.9	0.8375	1.0000	0.6938	1.0000
9	0	9	0.000	256.2	0.8718	1.0000	0.9688	1.0000
10	4	64	0.062	284.6	0.9718	1.0000	0.9688	1.0000
11	3	103	0.029	281.2	0.8938	1.0000	0.9375	1.0000
12	1	31	0.032	160.2	0.9937	1.0000	0.9875	1.0000
13	0	2	0.000	272	0.7937	1.3750	0.7875	1.0000
14	2	43	0.047	289	0.7718	1.0000	0.6625	1.1250
15	12	97	0.124	262.8	0.9937	1.0000	0.9875	1.0000
16	0	36	0.000	289	0.9093	1.0000	0.9438	1.0000
17	0	11	0.000	273.5	0.5468	1.1875	0.5312	2.0000
18	3	26	0.115	253.5	0.9000	1.0000	0.8000	1.0000
19	0	6	0.000	292.2	0.8968	1.1250	0.9062	1.0000
20	0	6	0.000	291	0.9406	1.0000	0.9688	1.0000
21	0	27	0.000	283.3	0.9250	1.0000	0.9438	1.0000
22	0	18	0.000	245.3	0.8250	1.0000	0.6875	1.2500
23	0	48	0.000	289	0.6875	1.0000	0.7625	1.0000
24	0	4	0.000	286.3	0.9218	1.0000	0.8687	1.2500
25	1	41	0.024	241.5	0.1406	1.8750	0.1500	2.5000
26	2	36	0.056	234.1	0.0063	1.0000	0.0125	1.0000
27	2	58	0.034	202	0.8813	1.0000	0.9625	1.0000
28	0	8	0.000	226.4	0.0000	1.0000	0.0000	1.0000
29	1	15	0.067	226.6	0.0094	1.0000	0.0188	1.0000
30	0	3	0.000	227	0.0000	1.0000	0.0000	1.0000
31	1	64	0.016	160.7	0.0094	1.0000	0.0000	1.0000
32	0	80	0.000	170.2	0.0000	1.0000	0.0000	1.0000
33	0	18	0.000	214	0.0281	1.0000	0.0312	1.0000
34	2	30	0.067	184.4	0.0094	1.0000	0.0000	1.0000
35	1	17	0.059	211.2	0.1312	1.2500	0.2250	1.7500
36	1	17	0.059	161.8	0.2719	1.0000	0.2937	1.0000
37	1	25	0.040	167.8	0.0469	1.0000	0.0437	1.0000
38	1	23	0.043	171	0.0281	1.0000	0.0000	1.0000
39	0	21	0.000	182	0.0000	1.0000	0.0000	1.0000
40	0	21	0.000	182.6	0.0625	1.0000	0.0000	1.0000
41	0	24	0.000	171	0.1469	1.0000	0.0750	1.0000
42	3	42	0.071	178.4	0.0281	1.0000	0.0000	1.0000
43	0	23	0.000	171	0.0656	1.0000	0.0250	1.0000

Appendix 5. Complete dataset for the spatial regression models.

Site #	Comp(1Km)	Conf(1Km)	Strain	Precip	Temp	Drought	HPS	Urban	Year
1	0.9875	1.0000	0	964	5.4	0	0	0	1995
2	0.1000	1.0000	0	964	5.4	0	0	0	1995
3	0.9750	1.0000	0	796	4.2	0	0	0	1994
4	1.0000	1.0000	0	796	4.2	0	0	0	1995
5	1.0000	1.0000	0	796	4.2	0	0	0	94/95
6	0.9125	1.5000	0	964	5.4	0	0	0	1995
7	0.9000	1.5000	0	964	5.4	0	0	0	1995
8	0.5750	1.2500	0	796	4.2	0	0	0	94/95
9	0.9750	1.0000	0	964	5.4	0	0	0	1995
10	1.0000	1.0000	0	796	4.2	0	0	0	1995
11	0.9250	1.0000	0	1063	5.4	0	0	0	1995
12	0.9750	1.0000	0	796	4.2	0	0	0	94/95
13	0.7875	1.2500	0	964	5.4	0	0	0	1995
14	0.7125	1.0000	0	796	4.2	0	0	0	1994
15	0.9875	1.0000	0	922	2.3	0	0	0	1995
16	0.9625	1.0000	0	796	4.2	0	0	0	1994
17	0.2875	1.7500	0	964	5.4	0	0	0	1995
18	0.7875	1.0000	0	922	2.3	0	0	0	1995
19	1.0000	1.0000	0	922	2.3	0	0	0	1995
20	0.9625	1.5000	0	922	2.3	0	0	0	1995
21	0.9625	1.0000	0	922	2.3	0	0	0	1995
22	0.9125	1.5000	0	922	2.3	0	0	0	1995
23	0.7750	1.0000	0	793	1.7	0	0	0	1994
24	0.8750	1.2500	0	874	0.6	0	0	0	1995
25	0.1625	2.0000	1	317	2.4	0	0	1	1994
26	0.0000	1.0000	1	317	2.4	0	0	0	1994
27	1.0000	1.0000	1	304	10	5	1	1	1994
28	0.0000	1.0000	1	317	2.4	0	0	0	1994
29	0.0000	1.0000	1	317	2.4	0	0	0	1994
30	0.0000	1.0000	1	317	2.4	0	0	0	1994
31	0.0000	1.0000	1	317	2.4	0	0	0	1994
32	0.0000	1.0000	1	317	2.4	0	0	0	1994
33	0.0000	1.0000	1	317	2.4	0	0	0	1994
34	0.0000	1.0000	1	317	2.4	0	0	0	1994
35	0.1625	2.2500	1	317	2.4	0	0	0	1994
36	0.1750	1.0000	1	317	2.4	0	0	0	1994
37	0.0000	1.0000	1	317	2.4	0	0	0	1994
38	0.0000	1.0000	1	317	2.4	0	0	0	1994
39	0.0000	1.0000	1	317	2.4	0	0	0	1994
40	0.0000	1.0000	1	317	2.4	0	0	0	1994
41	0.0000	1.0000	1	317	2.4	0	0	0	1994
42	0.0000	1.0000	1	317	2.4	0	0	0	1994
43	0.0000	1.0000	1	317	2.4	0	0	0	1994

Appendix 5. Complete dataset for the spatial regression models.

44	0	25	0.000	197.2	0.0000	1.0000	0.0000	1.0000
45	3	60	0.050	219.8	0.5718	1.0000	0.2000	1.1250
46	3	24	0.125	197.2	0.0000	1.0000	0.0000	1.0000
47	0	151	0.000	150.2	0.8312	1.0000	0.8938	1.0000
48	0	55	0.000	132.8	0.6687	1.5625	0.6062	2.1250
49	12	54	0.222	146.9	0.5063	1.5000	0.6687	2.1250
50	0	3	0.000	279.3	0.7031	1.0000	0.5825	1.3750
51	0	11	0.000	270.5	0.7312	1.0000	0.6062	1.1250
52	0	2	0.000	266	0.8719	1.0000	0.7437	1.0000
53	0	3	0.000	251	0.7188	1.0000	0.6687	1.0000
54	8	111	0.072	301.5	0.8625	1.0000	0.8938	1.1250
55	0	1	0.000	79	0.9625	1.0000	0.9250	1.0000
56	0	2	0.000	82	0.9094	1.0000	0.9187	1.0000
58	0	3	0.000	59.3	0.8562	1.0000	0.9688	1.0000
60.1	14	38	0.368	163.2	0.8313	1.0000	0.8438	1.1250
61	0	3	0.000	55.7	0.7812	1.0000	0.7312	1.0000
62	3	3	1.000	251	0.9812	1.0000	0.9750	1.0000
64	6	43	0.140	201.9	1.0000	1.2500	1.0000	1.0000
65	4	13	0.308	128.5	0.9219	1.0000	0.9812	1.0000
66	8	19	0.421	76.4	0.3656	1.2500	0.1875	1.0000
67	0	2	0.000	279	0.8438	1.1250	0.7375	1.3750
68	0	55	0.000	295	0.4437	1.2500	0.3938	1.8750
69	0	34	0.000	263	0.8625	1.0000	0.9375	1.0000
70	0	8	0.000	105	0.6656	1.0000	0.6375	1.0000
71	0	2	0.000	300	0.8781	1.0000	0.9125	1.0000
72	0	5	0.000	105	0.9781	1.0000	1.0000	1.0000
73	0	2	0.000	197	0.7906	1.0625	0.7188	1.3750
74	0	1	0.000	228	0.9625	1.0000	1.0000	1.0000
75	0	6	0.000	230.5	0.7656	1.0000	0.7875	1.0000
76	0	3	0.000	105	0.8375	1.0000	0.9875	1.0000
77	0	8	0.000	274.4	0.7781	1.0000	0.8375	1.2500
78	1	64	0.016	215	0.6812	1.0625	0.6562	1.1250
79	1	13	0.077	188.5	0.6750	1.0000	0.8188	1.3750
80	0	2	0.000	197	0.3219	1.0000	0.1813	1.0000
81	0	2	0.000	252.5	0.5656	1.3125	0.3688	1.7500
82	1	66	0.015	233	0.7468	1.0625	0.8938	1.0000
83	0	6	0.000	264.7	0.7250	1.0000	0.5688	1.1250
84	0	160	0.000	176	0.6062	1.2500	0.6625	1.0000
85	0	80	0.000	245	0.4500	1.0000	0.5000	1.0000
86	0	4	0.000	254.3	0.9563	1.0000	0.9438	1.0000
87	0	3	0.000	105	0.9594	1.0000	0.9187	1.0000
88	0	1	0.000	197	0.7781	1.0000	0.9312	1.0000
89.1	0	4	0.000	202	0.8125	1.1250	0.8875	1.1250
89.2	2	9	0.222	248.6	0.5312	1.0000	0.4250	1.1250

Appendix 5. Complete dataset for the spatial regression models.

44	0.0000	1.0000	1	317	2.4	0	0	0	1994
45	0.0750	1.0000	1	256	8.3	7	1	1	1994
46	0.0000	1.0000	1	317	2.4	0	0	0	1994
47	0.7925	1.7500	1	695	0	0	0	0	94/95
48	0.6250	3.0000	1	695	0	0	0	0	1995
49	0.7500	3.2500	1	695	0	0	0	0	1995
50	0.4125	2.0000	1	1524	1.8	0	0	1	1995
51	0.4625	2.2500	1	547	4.5	0	0	1	1995
52	0.4875	1.5000	1	547	4.5	0	0	1	1995
53	0.8125	1.7500	1	467	1.6	0	0	1	1995
54	0.9375	1.0000	1	467	1.6	0	0	0	1995
55	0.9625	1.0000	1	467	1.6	0	0	1	1995
56	0.9500	1.0000	1	467	1.6	0	0	1	1995
58	0.9375	1.0000	1	467	1.6	0	0	1	1995
60.1	0.8625	1.2500	1	467	1.6	0	0	1	1995
61	0.7125	1.2500	1	467	1.6	0	0	1	1995
62	1.0000	1.0000	1	467	1.6	0	0	1	1995
64	1.0000	1.0000	1	467	1.6	0	0	1	1995
65	0.9625	1.0000	1	467	1.6	0	0	1	1995
66	0.0875	1.0000	1	467	1.6	0	0	1	1995
67	0.5750	1.7500	1	547	4.5	0	0	1	1995
68	0.2625	1.7500	1	454	-0.6	0	0	0	1995
69	0.9500	1.0000	1	454	-0.6	0	0	0	1994
70	0.5625	1.2500	1	547	4.5	0	0	1	1995
71	0.9375	1.5000	1	547	4.5	0	0	1	1995
72	1.0000	1.0000	1	547	4.5	0	0	1	1995
73	0.7125	2.2500	1	723	0.1	0	0	1	1995
74	1.0000	1.0000	1	723	0.1	0	0	1	1995
75	0.8500	1.0000	1	723	0.1	0	0	1	1995
76	0.9750	1.0000	1	723	0.1	0	0	1	1995
77	0.9125	1.0000	1	723	0.1	0	0	1	1995
78	0.5875	1.7500	1	454	-0.6	0	0	0	1994
79	0.8125	2.0000	1	723	0.1	0	0	1	1995
80	0.1500	1.5000	1	723	0.1	0	0	1	1995
81	0.3875	2.2500	1	723	0.1	0	0	1	1995
82	0.9250	1.0000	1	454	-0.6	0	0	0	1994
83	0.7125	1.2500	1	723	0.1	0	0	1	1995
84	0.7500	1.0000	1	454	-0.6	0	0	0	93/95
85	0.5000	1.0000	1	815	6.2	0	0	0	1994
86	0.8875	2.0000	1	723	0.1	0	0	1	1995
87	0.8875	1.7500	1	556	2.6	0	0	1	1995
88	1.0000	1.0000	1	723	0.1	0	0	1	1995
89.1	0.8750	1.2500	1	454	-0.6	0	0	1	1994
89.2	0.0625	1.2500	1	454	-0.6	0	0	1	1994

Appendix 5. Complete dataset for the spatial regression models.

89.3	0	5	0.000	225	0.8499	1.0000	0.7875	1.5000
90	0	3	0.000	197	0.6719	1.0000	0.6750	1.0000
91	0	5	0.000	197	0.8531	1.0000	0.7375	1.2500
92	0	20	0.000	235	0.8343	1.0000	0.9563	1.0000
93	9	72	0.125	213.1	0.9687	1.0000	1.0000	1.0000
94	0	16	0.000	215	0.8781	1.0000	0.8375	1.1250
95	0	20	0.000	235	0.5844	1.0000	0.6750	1.0000
96	0	20	0.000	235	0.6562	1.0000	0.9125	1.0000
101	0	23	0	258.4	0.5844	1.4375	0.3688	2.1250
102	0	7	0	228.6	0.8375	1.5000	0.7812	2.5000
103	0	36	0	261.1	0.8062	1.3750	0.7375	1.8750
104	0	6	0	276.3	0.7531	1.2500	0.7875	1.5000
105	0	19	0	249.2	0.6625	1.4375	0.4750	2.1250
106	0	17	0	260	0.8468	1.2500	0.7937	1.2500

Appendix 5. Complete dataset for the spatial regression models.

89.3	0.7375	2.2500	1	454	-0.6	0	0	1	1994
90	0.6000	1.2500	1	723	0.1	0	0	1	1995
91	0.5375	1.5000	1	416	-2.3	0	0	1	1995
92	1.0000	1.0000	1	454	-0.6	0	0	0	1994
93	1.0000	1.0000	1	425	-3.3	0	0	0	1994
94	0.9000	1.2500	1	425	-3.3	0	0	1	1994
95	0.8000	1.0000	1	454	-0.6	0	0	0	1994
96	1.0000	1.0000	1	454	-0.6	0	0	0	1994
101	0.3875	3.0000	0	796	4.2	0	0	0	1995
102	0.7250	2.5000	0	796	4.2	0	0	0	1995
103	0.5875	2.7500	0	796	4.2	0	0	0	1995
104	0.6125	2.5000	0	796	4.2	0	0	0	1995
105	0.3875	2.7500	0	796	4.2	0	0	0	1995
106	0.5875	2.5000	0	796	4.2	0	0	0	1995

Appendix 6. Complete dataset for the temporal regression model.

Site #	Date	Y	N	Y/N	ds1	ds2	ds3	ds4	ds5	ds6	ds7
1	262	0	13	0.000	0	0	0	0	0	0	1
1	263	0	28	0.000	0	0	0	0	0	0	1
1	264	0	21	0.000	0	0	0	0	0	0	1
1	269	0	12	0.000	0	0	0	0	0	0	1
1	270	0	31	0.000	0	0	0	0	0	0	1
2	262	0	4	0.000	0	0	0	0	0	1	0
2	263	0	7	0.000	0	0	0	0	0	1	0
2	264	0	19	0.000	0	0	0	0	0	1	0
2	269	0	17	0.000	0	0	0	0	0	1	0
2	270	0	26	0.000	0	0	0	0	0	1	0
2	271	0	20	0.000	0	0	0	0	0	1	0
2	272	0	3	0.000	0	0	0	0	0	1	0
3	159	0	2	0.000	0	0	0	0	0	1	1
3	160	0	1	0.000	0	0	0	0	0	1	1
3	161	0	1	0.000	0	0	0	0	0	1	1
3	168	0	1	0.000	0	0	0	0	0	1	1
3	173	0	1	0.000	0	0	0	0	0	1	1
3	174	0	2	0.000	0	0	0	0	0	1	1
3	175	0	1	0.000	0	0	0	0	0	1	1
4	218	0	5	0.000	0	0	0	0	1	0	0
4	219	1	5	0.200	0	0	0	0	1	0	0
4	220	0	4	0.000	0	0	0	0	1	0	0
4	222	1	4	0.250	0	0	0	0	1	0	0
4	224	0	1	0.000	0	0	0	0	1	0	0
4	233	0	2	0.000	0	0	0	0	1	0	0
4	234	0	3	0.000	0	0	0	0	1	0	0
4	235	0	1	0.000	0	0	0	0	1	0	0
4	239	0	1	0.000	0	0	0	0	1	0	0
4	242	0	2	0.000	0	0	0	0	1	0	0
4	243	0	2	0.000	0	0	0	0	1	0	0
4	247	0	1	0.000	0	0	0	0	1	0	0
4	267	0	5	0.000	0	0	0	0	1	0	0
5	150	0	2	0.000	0	0	0	0	1	0	1
5	151	0	1	0.000	0	0	0	0	1	0	1
5	152	0	4	0.000	0	0	0	0	1	0	1
5	153	0	1	0.000	0	0	0	0	1	0	1
5	154	0	1	0.000	0	0	0	0	1	0	1
5	166	0	1	0.000	0	0	0	0	1	0	1
5	168	0	1	0.000	0	0	0	0	1	0	1
5	179	0	2	0.000	0	0	0	0	1	0	1
5	180	0	1	0.000	0	0	0	0	1	0	1
6	254	0	1	0.000	0	0	0	0	1	1	0
7	255	0	3	0.000	0	0	0	0	1	1	1

Appendix 6. Complete dataset for the temporal regression model.

7	256	0	4	0.000	0	0	0	0	1	1	1
7	257	0	2	0.000	0	0	0	0	1	1	1
7	258	0	3	0.000	0	0	0	0	1	1	1
8	150	0	10	0.000	0	0	0	1	0	0	0
8	151	0	4	0.000	0	0	0	1	0	0	0
8	152	1	4	0.250	0	0	0	1	0	0	0
8	152	0	2	0.000	0	0	0	1	0	0	0
8	153	0	1	0.000	0	0	0	1	0	0	0
8	167	0	1	0.000	0	0	0	1	0	0	0
8	174	1	1	1.000	0	0	0	1	0	0	0
8	175	0	1	0.000	0	0	0	1	0	0	0
9	255	0	1	0.000	0	0	0	1	0	0	1
9	256	0	7	0.000	0	0	0	1	0	0	1
9	259	0	1	0.000	0	0	0	1	0	0	1
10	276	0	15	0.000	0	0	0	1	0	1	0
10	277	0	7	0.000	0	0	0	1	0	1	0
10	278	0	1	0.000	0	0	0	1	0	1	0
10	284	1	5	0.200	0	0	0	1	0	1	0
10	285	2	14	0.143	0	0	0	1	0	1	0
10	286	0	3	0.000	0	0	0	1	0	1	0
10	287	0	3	0.000	0	0	0	1	0	1	0
10	291	0	1	0.000	0	0	0	1	0	1	0
10	292	0	2	0.000	0	0	0	1	0	1	0
10	296	1	6	0.167	0	0	0	1	0	1	0
10	297	0	7	0.000	0	0	0	1	0	1	0
11	279	1	21	0.048	0	0	0	1	0	1	1
11	280	0	22	0.000	0	0	0	1	0	1	1
11	281	0	20	0.000	0	0	0	1	0	1	1
11	282	1	9	0.111	0	0	0	1	0	1	1
11	283	0	17	0.000	0	0	0	1	0	1	1
11	284	1	14	0.071	0	0	0	1	0	1	1
12	150	1	4	0.250	0	0	0	1	1	0	0
12	151	0	5	0.000	0	0	0	1	1	0	0
12	152	0	1	0.000	0	0	0	1	1	0	0
12	152	0	5	0.000	0	0	0	1	1	0	0
12	153	0	1	0.000	0	0	0	1	1	0	0
12	159	0	5	0.000	0	0	0	1	1	0	0
12	166	0	1	0.000	0	0	0	1	1	0	0
12	167	0	1	0.000	0	0	0	1	1	0	0
12	168	0	1	0.000	0	0	0	1	1	0	0
12	173	0	1	0.000	0	0	0	1	1	0	0
12	179	0	2	0.000	0	0	0	1	1	0	0
12	180	0	3	0.000	0	0	0	1	1	0	0
12	181	0	1	0.000	0	0	0	1	1	0	0

Appendix 6. Complete dataset for the temporal regression model.

13	272	0	2	0.000	0	0	0	1	1	0	1
14	289	2	43	0.047	0	0	0	1	1	1	0
15	256	0	3	0.000	0	0	0	1	1	1	1
15	257	1	4	0.250	0	0	0	1	1	1	1
15	258	0	1	0.000	0	0	0	1	1	1	1
15	262	4	24	0.167	0	0	0	1	1	1	1
15	263	4	28	0.143	0	0	0	1	1	1	1
15	264	2	24	0.083	0	0	0	1	1	1	1
15	265	1	13	0.077	0	0	0	1	1	1	1
16	289	0	36	0.000	0	0	1	0	0	0	0
17	267	0	3	0.000	0	0	1	0	0	0	1
17	276	0	8	0.000	0	0	1	0	0	0	1
18	238	0	1	0.000	0	0	1	0	0	1	0
18	239	0	1	0.000	0	0	1	0	0	1	0
18	246	0	10	0.000	0	0	1	0	0	1	0
18	247	2	7	0.286	0	0	1	0	0	1	0
18	274	1	6	0.167	0	0	1	0	0	1	0
18	281	0	1	0.000	0	0	1	0	0	1	0
19	291	0	1	0.000	0	0	1	0	0	1	1
19	292	0	3	0.000	0	0	1	0	0	1	1
19	293	0	2	0.000	0	0	1	0	0	1	1
20	291	0	6	0.000	0	0	1	0	1	0	0
21	249	0	1	0.000	0	0	1	0	1	0	1
21	253	0	1	0.000	0	0	1	0	1	0	1
21	256	0	1	0.000	0	0	1	0	1	0	1
21	258	0	1	0.000	0	0	1	0	1	0	1
21	261	0	1	0.000	0	0	1	0	1	0	1
21	265	0	1	0.000	0	0	1	0	1	0	1
21	268	0	1	0.000	0	0	1	0	1	0	1
21	269	0	2	0.000	0	0	1	0	1	0	1
21	275	0	1	0.000	0	0	1	0	1	0	1
21	276	0	1	0.000	0	0	1	0	1	0	1
21	282	0	1	0.000	0	0	1	0	1	0	1
21	291	0	1	0.000	0	0	1	0	1	0	1
21	295	0	1	0.000	0	0	1	0	1	0	1
21	296	0	2	0.000	0	0	1	0	1	0	1
21	297	0	1	0.000	0	0	1	0	1	0	1
21	298	0	4	0.000	0	0	1	0	1	0	1
21	299	0	2	0.000	0	0	1	0	1	0	1
21	300	0	2	0.000	0	0	1	0	1	0	1
21	301	0	1	0.000	0	0	1	0	1	0	1
21	302	0	1	0.000	0	0	1	0	1	0	1
22	236	0	2	0.000	0	0	1	0	1	1	0
22	237	0	1	0.000	0	0	1	0	1	1	0

Appendix 6. Complete dataset for the temporal regression model.

22	241	0	2	0.000	0	0	1	0	1	1	0
22	242	0	3	0.000	0	0	1	0	1	1	0
22	249	0	4	0.000	0	0	1	0	1	1	0
22	250	0	4	0.000	0	0	1	0	1	1	0
22	251	0	2	0.000	0	0	1	0	1	1	0
23	289	0	48	0.000	0	0	1	0	1	1	1
24	285	0	1	0.000	0	0	1	1	0	0	0
24	286	0	1	0.000	0	0	1	1	0	0	0
24	287	0	2	0.000	0	0	1	1	0	0	0
25	236	1	20	0.050	0	0	1	1	0	0	1
25	241	0	10	0.000	0	0	1	1	0	0	1
25	245	0	2	0.000	0	0	1	1	0	0	1
25	246	0	5	0.000	0	0	1	1	0	0	1
25	263	0	4	0.000	0	0	1	1	0	0	1
26	217	0	16	0.000	0	0	1	1	0	1	0
26	219	2	5	0.400	0	0	1	1	0	1	0
26	233	0	3	0.000	0	0	1	1	0	1	0
26	235	0	1	0.000	0	0	1	1	0	1	0
26	266	0	11	0.000	0	0	1	1	0	1	0
27	202	2	58	0.034	0	0	1	1	0	1	1
28	226	0	5	0.000	0	0	1	1	1	0	0
28	227	0	3	0.000	0	0	1	1	1	0	0
29	226	1	7	0.143	0	0	1	1	1	0	1
29	227	0	7	0.000	0	0	1	1	1	0	1
29	228	0	1	0.000	0	0	1	1	1	0	1
30	227	0	3	0.000	0	0	1	1	1	1	0
31	171	0	22	0.000	0	0	1	1	1	1	1
31	172	0	4	0.000	0	0	1	1	1	1	1
31	200	1	22	0.045	0	0	1	1	1	1	1
31	201	0	15	0.000	0	0	1	1	1	1	1
31	202	0	6	0.000	0	0	1	1	1	1	1
31	203	0	5	0.000	0	0	1	1	1	1	1
32	79	0	1	0.000	0	1	0	0	0	0	0
32	171	0	54	0.000	0	1	0	0	0	0	0
32	172	0	21	0.000	0	1	0	0	0	0	0
32	173	0	4	0.000	0	1	0	0	0	0	0
33	182	0	1	0.000	0	1	0	0	0	0	1
33	214	0	1	0.000	0	1	0	0	0	0	1
33	215	0	3	0.000	0	1	0	0	0	0	1
33	216	0	10	0.000	0	1	0	0	0	0	1
33	217	0	3	0.000	0	1	0	0	0	0	1
34	171	0	20	0.000	0	1	0	0	0	1	0
34	211	2	9	0.222	0	1	0	0	0	1	0
34	212	0	1	0.000	0	1	0	0	0	1	0

Appendix 6. Complete dataset for the temporal regression model.

35	210	1	5	0.200	0	1	0	0	0	1	1
35	211	0	3	0.000	0	1	0	0	0	1	1
35	212	0	9	0.000	0	1	0	0	0	1	1
36	158	1	11	0.091	0	1	0	0	1	0	0
36	168	0	4	0.000	0	1	0	0	1	0	0
36	170	0	2	0.000	0	1	0	0	1	0	0
37	167	0	11	0.000	0	1	0	0	1	0	1
37	168	0	8	0.000	0	1	0	0	1	0	1
37	169	1	6	0.167	0	1	0	0	1	0	1
38	171	1	23	0.043	0	1	0	0	1	1	0
39	181	0	6	0.000	0	1	0	0	1	1	1
39	182	0	10	0.000	0	1	0	0	1	1	1
39	183	0	5	0.000	0	1	0	0	1	1	1
40	182	0	13	0.000	0	1	0	1	0	0	0
40	183	0	4	0.000	0	1	0	1	0	0	0
40	184	0	4	0.000	0	1	0	1	0	0	0
41	171	0	24	0.000	0	1	0	1	0	0	1
42	171	1	20	0.050	0	1	0	1	0	1	0
42	184	0	8	0.000	0	1	0	1	0	1	0
42	185	1	7	0.143	0	1	0	1	0	1	0
42	186	1	6	0.167	0	1	0	1	0	1	0
42	188	0	1	0.000	0	1	0	1	0	1	0
43	171	0	23	0.000	0	1	0	1	0	1	1
44	196	0	6	0.000	0	1	0	1	1	0	0
44	197	0	9	0.000	0	1	0	1	1	0	0
44	198	0	10	0.000	0	1	0	1	1	0	0
45	217	3	34	0.088	0	1	0	1	1	0	1
45	223	0	16	0.000	0	1	0	1	1	0	1
45	224	0	10	0.000	0	1	0	1	1	0	1
46	196	0	5	0.000	0	1	0	1	1	1	0
46	197	1	12	0.083	0	1	0	1	1	1	0
46	198	1	4	0.250	0	1	0	1	1	1	0
46	199	1	3	0.333	0	1	0	1	1	1	0
47	128	0	7	0.000	0	1	0	1	1	1	1
47	131	0	3	0.000	0	1	0	1	1	1	1
47	132	0	3	0.000	0	1	0	1	1	1	1
47	135	0	4	0.000	0	1	0	1	1	1	1
47	137	0	5	0.000	0	1	0	1	1	1	1
47	138	0	1	0.000	0	1	0	1	1	1	1
47	141	0	3	0.000	0	1	0	1	1	1	1
47	152	0	101	0.000	0	1	0	1	1	1	1
47	158	0	2	0.000	0	1	0	1	1	1	1
47	159	0	7	0.000	0	1	0	1	1	1	1
47	161	0	3	0.000	0	1	0	1	1	1	1

Appendix 6. Complete dataset for the temporal regression model.

47	162	0	3	0.000	0	1	0	1	1	1	1
47	163	0	9	0.000	0	1	0	1	1	1	1
48	129	0	25	0.000	0	1	1	0	0	0	0
48	130	0	6	0.000	0	1	1	0	0	0	0
48	137	0	13	0.000	0	1	1	0	0	0	0
48	138	0	11	0.000	0	1	1	0	0	0	0
49	146	7	26	0.269	0	1	1	0	0	0	1
49	147	1	6	0.167	0	1	1	0	0	0	1
49	148	4	22	0.182	0	1	1	0	0	0	1
50	276	0	1	0.000	0	1	1	0	0	1	0
50	280	0	1	0.000	0	1	1	0	0	1	0
50	282	0	1	0.000	0	1	1	0	0	1	0
51	244	0	1	0.000	0	1	1	0	0	1	1
51	247	0	1	0.000	0	1	1	0	0	1	1
51	261	0	1	0.000	0	1	1	0	0	1	1
51	262	0	1	0.000	0	1	1	0	0	1	1
51	266	0	1	0.000	0	1	1	0	0	1	1
51	267	0	1	0.000	0	1	1	0	0	1	1
51	274	0	2	0.000	0	1	1	0	0	1	1
51	286	0	1	0.000	0	1	1	0	0	1	1
51	291	0	1	0.000	0	1	1	0	0	1	1
51	304	0	1	0.000	0	1	1	0	0	1	1
52	244	0	1	0.000	0	1	1	0	1	0	0
52	268	0	1	0.000	0	1	1	0	1	0	0
53	251	0	3	0.000	0	1	1	0	1	0	1
54	300	2	38	0.053	0	1	1	0	1	1	0
54	301	1	24	0.042	0	1	1	0	1	1	0
54	302	1	22	0.045	0	1	1	0	1	1	0
54	303	1	4	0.250	0	1	1	0	1	1	0
54	304	3	23	0.130	0	1	1	0	1	1	0
55	79	0	1	0.000	0	1	1	0	1	1	1
56	82	0	1	0.000	0	1	1	1	0	0	0
56	86	0	1	0.000	0	1	1	1	0	0	0
58	59	0	2	0.000	0	1	1	1	0	1	0
58	60	0	1	0.000	0	1	1	1	0	1	0
60.1	60	0	2	0.000	0	1	1	1	1	0	0
60.1	65	1	4	0.250	0	1	1	1	0	0	1
60.1	68	0	1	0.000	0	1	1	1	0	0	1
60.1	69	0	2	0.000	0	1	1	1	0	0	1
60.1	75	2	3	0.667	0	1	1	1	0	0	1
60.1	79	1	2	0.500	0	1	1	1	0	0	1
60.1	86	1	5	0.200	0	1	1	1	0	0	1
60.1	90	0	1	0.000	0	1	1	1	1	0	0
60.1	260	5	9	0.556	0	1	1	1	0	0	1

Appendix 6. Complete dataset for the temporal regression model.

60.1	261	2	4	0.500	0	1	1	1	0	0	1
60.1	263	2	2	1.000	0	1	1	1	0	0	1
60.1	264	0	1	0.000	0	1	1	1	0	0	1
60.1	266	0	1	0.000	0	1	1	1	0	0	1
60.1	272	0	1	0.000	0	1	1	1	0	1	1
61	51	0	2	0.000	0	1	1	1	1	0	1
61	65	0	1	0.000	0	1	1	1	1	0	1
62	251	3	3	1.000	0	1	1	1	1	1	0
64	58	1	1	1.000	1	0	0	0	0	0	0
64	59	1	1	1.000	1	0	0	0	0	0	0
64	65	0	1	0.000	1	0	0	0	0	0	0
64	69	0	1	0.000	0	1	1	1	1	1	1
64	79	0	2	0.000	0	1	1	1	1	1	1
64	82	1	4	0.250	1	0	0	0	0	0	0
64	86	0	1	0.000	1	0	0	0	0	0	0
64	241	1	4	0.250	1	0	0	0	0	0	0
64	242	1	6	0.167	1	0	0	0	0	0	0
64	243	0	4	0.000	1	0	0	0	0	0	0
64	244	0	3	0.000	1	0	0	0	0	0	0
64	246	0	1	0.000	1	0	0	0	0	0	0
64	247	0	2	0.000	1	0	0	0	0	0	0
64	248	0	3	0.000	1	0	0	0	0	0	0
64	249	0	3	0.000	1	0	0	0	0	0	0
64	250	0	2	0.000	1	0	0	0	0	0	0
64	251	0	2	0.000	1	0	0	0	0	0	0
64	253	0	1	0.000	1	0	0	0	0	0	0
64	254	1	1	1.000	1	0	0	0	0	0	0
65	59	0	4	0.000	1	0	0	0	0	0	1
65	72	3	3	1.000	1	0	0	0	0	0	1
65	79	0	1	0.000	1	0	0	0	0	0	1
65	228	1	5	0.200	1	0	0	0	0	0	1
66	67	3	3	1.000	1	0	0	0	0	1	0
66	69	0	2	0.000	1	0	0	0	0	1	0
66	72	1	4	0.250	1	0	0	0	0	1	0
66	75	1	2	0.500	1	0	0	0	0	1	0
66	79	0	2	0.000	1	0	0	0	0	1	0
66	82	1	1	1.000	1	0	0	0	0	1	0
66	86	2	4	0.500	1	0	0	0	0	1	0
66	90	0	1	0.000	1	0	0	0	0	1	0
67	279	0	2	0.000	1	0	0	0	0	1	1
68	295	0	55	0.000	1	0	0	0	1	0	0
69	263	0	34	0.000	1	0	0	0	1	0	1
70	105	0	8	0.000	1	0	0	0	1	1	0
71	300	0	2	0.000	1	0	0	0	1	1	1

Appendix 6. Complete dataset for the temporal regression model.

72	105	0	5	0.000	1	0	0	1	0	0	0
73	197	0	2	0.000	1	0	0	1	0	0	1
74	228	0	1	0.000	1	0	0	1	0	1	0
75	105	0	2	0.000	1	0	0	1	0	1	1
75	290	0	1	0.000	1	0	0	1	0	1	1
75	293	0	2	0.000	1	0	0	1	0	1	1
75	297	0	1	0.000	1	0	0	1	0	1	1
76	105	0	3	0.000	1	0	0	1	1	0	0
77	261	0	2	0.000	1	0	0	1	1	0	1
77	276	0	3	0.000	1	0	0	1	1	0	1
77	278	0	2	0.000	1	0	0	1	1	0	1
77	289	0	1	0.000	1	0	0	1	1	0	1
78	215	1	64	0.016	1	0	0	1	1	1	0
79	105	1	4	0.250	1	0	0	1	1	1	1
79	197	0	6	0.000	1	0	0	1	1	1	1
79	282	0	2	0.000	1	0	0	1	1	1	1
79	284	0	1	0.000	1	0	0	1	1	1	1
80	197	0	2	0.000	1	0	1	0	0	0	0
81	252	0	1	0.000	1	0	1	0	0	0	1
81	253	0	1	0.000	1	0	1	0	0	0	1
82	233	1	66	0.015	1	0	1	0	0	1	0
83	248	0	2	0.000	1	0	1	0	0	1	1
83	273	0	4	0.000	1	0	1	0	0	1	1
84	136	0	55	0.000	1	0	1	0	1	0	0
84	197	0	105	0.000	1	0	1	0	1	0	0
85	245	0	80	0.000	1	0	1	0	1	0	1
86	228	0	1	0.000	1	0	1	0	1	1	0
86	253	0	2	0.000	1	0	1	0	1	1	0
86	283	0	1	0.000	1	0	1	0	1	1	0
87	105	0	3	0.000	1	0	1	0	1	1	1
88	197	0	1	0.000	1	0	1	1	0	0	0
89.1	201	0	2	0.000	1	0	1	1	0	0	1
89.1	202	0	2	0.000	1	0	1	1	0	0	1
89.2	207	0	2	0.000	1	0	1	1	0	0	1
89.2	243	1	2	0.500	1	0	1	1	0	0	1
89.2	255	1	2	0.500	1	0	1	1	0	0	1
89.2	275	0	2	0.000	1	0	1	1	0	0	1
89.2	277	0	1	0.000	1	0	1	1	0	0	1
89.3	224	0	4	0.000	1	0	1	1	0	0	1
89.3	229	0	1	0.000	1	0	1	1	0	0	1
90	197	0	3	0.000	1	0	1	1	0	1	0
91	197	0	5	0.000	1	0	1	1	0	1	1
92	235	0	20	0.000	1	0	1	1	1	0	0
93	208	1	5	0.200	1	0	1	1	1	0	1

Appendix 6. Complete dataset for the temporal regression model.

93	209	3	19	0.158	1	0	1	1	1	0	1
93	210	3	13	0.231	1	0	1	1	1	0	1
93	211	0	5	0.000	1	0	1	1	1	0	1
93	214	0	6	0.000	1	0	1	1	1	0	1
93	215	2	8	0.250	1	0	1	1	1	0	1
93	216	0	3	0.000	1	0	1	1	1	0	1
93	222	0	5	0.000	1	0	1	1	1	0	1
93	223	0	5	0.000	1	0	1	1	1	0	1
93	224	0	3	0.000	1	0	1	1	1	0	1
94	215	0	16	0.000	1	0	1	1	1	1	0
95	235	0	20	0.000	1	0	1	1	1	1	1
96	235	0	20	0.000	1	1	0	0	0	0	0
101	234	0	7	0.000	1	1	0	0	1	0	1
101	235	0	1	0.000	1	1	0	0	1	0	1
101	243	0	3	0.000	1	1	0	0	1	0	1
101	276	0	3	0.000	1	1	0	0	1	0	1
101	277	0	6	0.000	1	1	0	0	1	0	1
101	284	0	3	0.000	1	1	0	0	1	0	1
102	228	0	3	0.000	1	1	0	0	1	1	0
102	229	0	4	0.000	1	1	0	0	1	1	0
103	222	0	3	0.000	1	1	0	0	1	1	1
103	255	0	8	0.000	1	1	0	0	1	1	1
103	257	0	1	0.000	1	1	0	0	1	1	1
103	258	0	6	0.000	1	1	0	0	1	1	1
103	262	0	2	0.000	1	1	0	0	1	1	1
103	263	0	5	0.000	1	1	0	0	1	1	1
103	273	0	4	0.000	1	1	0	0	1	1	1
103	278	0	5	0.000	1	1	0	0	1	1	1
103	283	0	2	0.000	1	1	0	0	1	1	1
104	272	0	4	0.000	1	1	0	1	0	0	0
104	285	0	2	0.000	1	1	0	1	0	0	0
105	223	0	2	0.000	1	1	0	1	0	0	1
105	244	0	7	0.000	1	1	0	1	0	0	1
105	245	0	6	0.000	1	1	0	1	0	0	1
105	273	0	2	0.000	1	1	0	1	0	0	1
105	277	0	1	0.000	1	1	0	1	0	0	1
105	287	0	1	0.000	1	1	0	1	0	0	1
106	227	0	8	0.000	1	1	0	1	0	1	0
106	230	0	3	0.000	1	1	0	1	0	1	0
106	263	0	3	0.000	1	1	0	1	0	1	0
106	275	0	3	0.000	1	1	0	1	0	1	0