AN ABSTRACT OF THE DISSERTATION OF

Jennifer Orme Zavaleta for the degree of Doctor of Philosophy in Wildlife Science presented on March 6, 2003.
Title: Integrative Risk Analysis of Vector-Borne Disease

Abstract approved

Redacted for Privacy

W. Daniel Edge

In this dissertation I explore the application of two novel modeling techniques for improving risk analysis of vector-borne disease and discuss their potential use in integrating environmental risk assessment that guides environmental and public health decisions. Techniques for analyzing risk have been considered inadequate due to a lack of understanding of the problem and an appropriate analytic-deliberative process clarifying the meaning of analytic findings and uncertainty (National Research Council (NRC), 1996; Peterman and Anderson, 1999). Thus, new integrative risk analysis tools are needed that are responsive to more complex environmental problems. In this work, I develop a qualitative community model that combines a conventional biomathematical model of vector-borne disease transmission with recent developments in community modeling. My procedure predicts the change in risk of vector-borne disease from press perturbations, a disturbance that results in a permanent change in a growth parameter. I also use a Relational Bayesian Modeling technique to exploit existing data to determine plausible mechanisms and geospatial and temporal patterns of disease spread. I apply these tools to Lyme disease and West Nile Encephalitis as examples of two different vector-borne diseases associated with complex ecological communities. Both the qualitative modeling and Bayesian methods provide an integrated risk analysis framework that identifies relationships important in the system and thus, guide the application of quantitative models or provide sufficient information for management decisions.

INTEGRATIVE RISK ANALYSIS OF VECTOR-BORNE DISEASE

By

Jennifer Orme Zavaleta

A Dissertation

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 6, 2003
Commencement June, 2003

Doctor of Philosophy dissertation of Jennifer Orme Zavaleta presented March 6, 2003

APPROVED:

Redacted for Privacy

Major Professor, representing Wildlife Science

Redacted for Privacy

Head of the Department of Fisheries and Wildlife

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for Privacy

Jennifer Orme Zavaleta, Author

ACKNOWLEDGEMENTS

I thank Drs. Lawrence Curtis, W. Daniel Edge, Annette MacKay Rossignol, Philippe Rossignol, Thomas Savage and Barbara Shields for their guidance and support as members of my graduate committee. I also wish to thank Drs. Lawrence Reiter, Gilman Veith, Steven Hedtke, Foster (Sonny) Mayer, Harold Kibby, Thomas Fontaine, and Ms. Patricia Jackson from the US Environmental Protection Agency, National Health and Environmental Effects Research Laboratory for their encouragement, support, and flexibility for me to seek this degree while balancing my work responsibilities. I am grateful for the input and critique from the OSU Loop Group: Jeffrey Dambacher, Selina Heppell, Jane Jorgensen, Hiram Li, Hans Luh, and Philippe Rossignol. I am especially indebted to Jane Jorgensen and Philippe Rossignol for their interest and collegiality in working with me. Their expertise and guidance were invaluable.

Drs. Stephen Guptill, USGS, Chester Moore, CDC, Fredrick (Rick) Kutz, and Elizabeth Hilborn, EPA, provided helpful advice and contact information supporting this research. Data on West Nile Virus were provided by the Maryland Departments of Agriculture, Health and Mental Hygiene, and Environment. I would also like to thank Drs. Paul Epstein, Wayne Munns, John Paul, and Robert Lackey for their critical and constructive reviews of this dissertation.

I am particularly thankful for the friendship and support of my colleagues, Drs. Nancy Chiu, Amal Mahfouz, Edward Ohanian and especially, Yogendra (Yogi) Patel, who passed away December, 2002. They encouraged me to pursue this degree many years ago. Last, but not least, I am most appreciative of my family: my parents William C. Orme II and Caroline Rae Secrest, and in particular, my husband David W. Zavaleta, and my daughters Sarah and Leah, for without their support and inspiration, none of this would be possible.

CONTRIBUTION OF AUTHORS

Dr. Philippe Rossignol was involved in the research design and writing of each manuscript. Drs. Jane Jorgensen and Bruce D'Ambrosio were involved with the design of the research, model development, and in the writing of Chapter 3. In addition, Drs. Hans Luh and Fredrick (Rick) Kutz contributed to data collection and interpretation for Chapter 3. Drs. Anne Fairbrother and W. Daniel Edge contributed to the development of Chapter 4. W. Daniel Edge provided guidance for the overall development of this dissertation.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

INTEGRATIVE RISK ANALYSIS OF VECTOR-BORNE DISEASE

CHAPTER 1

INTRODUCTION

Environmental protection decisions are often guided by risk assessments serving as tools to develop regulatory policy and other related guidance. Risk assessment reflects a process for estimating the likelihood of an adverse effect resulting from an anthropogenic stress (National Research Council [NRC], 1983). As such, it involves both qualitative and quantitative analyses relating exposure to a stressor and biological responses. A key component of risk assessment is risk characterization, which builds on an analysis of risk, providing decision makers with the overall evidence of a hazard. Techniques for analyzing risk have been considered inadequate due to a lack of understanding of the problem and an appropriate analytic-deliberative process clarifying the meaning of analytic findings and uncertainty (NRC, 1996; Peterman and Anderson, 1999). This dissertation presents two novel modeling techniques for improving risk analysis and discusses their potential for use in environmental risk assessment and public health. I apply these techniques to the ecological aspects of infectious disease, an emerging scientific research issue (NRC, 2000; DiGiulio and Benson, 2002) as an example of integrating ecological and human health risk analysis.

Background

Over the past thirty years, the use of risk assessment in environmental decisions has increased among the scientific and regulatory community and is now required by Federal, State, Tribal, and some local governments. Historically, risk assessments were developed to protect humans from the potential carcinogenic effects of chemical exposures. Risk assessments now address endpoints other than

cancer, extend to species other than humans, and consider non-chemical stressors. With the development of *Guidelines for Ecological Risk Assessment* (US Environmental Protection Agency [USEPA], 1998), the impact of risks in complex ecosystems can be assessed, including problems extending across temporal and spatial scales and different levels of biological organization.

The initial risk assessment paradigm popularized by the NRC's National Academy of Sciences (NAS) centers on four primary steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NRC, 1983). Hazard identification involves a qualitative description of possible adverse effects. Dose-response assessment provides a quantitative estimate of the relationship between exposure and the biological response. A description of exposure from source to receptors, including environmental fate, relevant pathways, magnitude and duration are encompassed in the exposure assessment step. The final step, risk characterization then provides a description of the weight of the evidence concerning the hazard and the uncertainties, variability and assumptions used in the quantitative assessment.

The EPA's ecological guidelines (Figure 1.1) use a similar process beginning with a problem formulation step, which is a conceptualization of the problem and includes the development of an assessment plan (USEPA, 1992, 1998). The steps of exposure assessment and dose-response assessment, recast as characterization of exposure and effects, respectively, are encompassed in an overall analysis step. Included as part of the analysis step, are the development of exposure and stressor response profiles. Risk characterization makes up the final phase of the assessment and follows the same approach as under the NAS paradigm.

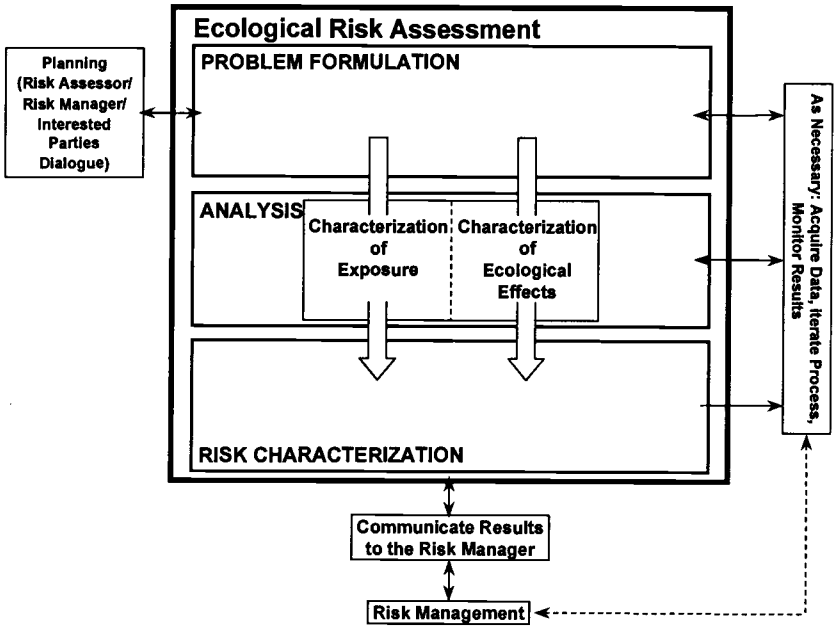


Figure 1.1. Ecological risk assessment framework (USEPA, 1992, 1998).

Risk characterization is an integral component to risk assessment. In their reexamination of risk characterization, the NAS suggested that risk characterization should be conducted at the onset of a risk assessment rather than as a concluding step, adding that it should be decision driven, involve the users of the information and reflect both analytic and deliberative processes (NRC, 1996). The analytic and deliberative processes are iterative--one influences the other. Analysis involves rigorous, replicable methods while deliberation involves a discussion of the issues that help frame further analysis.

Risk analysis can be quantitative and qualitative depending on the risk problem and available data. It involves the application of analytical techniques to understand risk, and weigh the impacts of different decision scenarios. Current analysis techniques are criticized as being inadequate and irrelevant, and have the potential to be misinterpreted due to a lack of understanding of the problem, and an inability to deal with uncertainty (NRC, 1996; Peterman and Anderson, 1999).

The paradigm for conducting distinct risk assessments for human health or ecological effects is now shifting toward the integration of these processes

(DiGiulio and Benson, 2002). The idea of integrating risk assessment approaches has been the topic of extensive discussion over the past decade. A recent forum sponsored by International Programme on Chemical Safety (IPCS) outlined an integrated process combining elements of both human health and ecological processes (Suter et al., 2003). Hazard identification becomes an element of problem formulation, and dose response assessment occurs as part of the characterization of effects. Stakeholders and risk managers are involved throughout the process to ensure buy-in and responsiveness of the assessment to the specific problem (Suter et al., 2003). Thus, integration combines the process of risk estimation for humans, biota, and natural resources into one assessment for the purpose of improving the information used in environmental decisions, resulting in more effective protection of resources that society values (Miranda et al., 2002; Suter et al., 2003). This approach would benefit from the consideration of interactions between stressors, receptors such as wildlife or humans, and the environment. Employing community ecology concepts in an integrated risk analysis approach may rectify the inadequacies of traditional analytical techniques. The basis for such an integrated approach is the perspective that ecosystems serve as part of the foundation for defining human well-being.

In this dissertation, we discuss two novel community-level models as new tools to be used in risk analysis. We apply these tools to the issue of emerging infectious disease, focusing on two different vector-borne diseases that are associated with complex ecological communities. The NRC Committee on Grand Challenges in Environmental Sciences suggested that an integrated risk assessment approach would be useful in addressing what they perceived as important environmental research challenges for the next generation (NRC, 2000). Emerging infectious disease and the environment were identified as one of four priority areas for research with a goal of improving our understanding of the interactions among pathogens, hosts/receptors, and the environment (NRC, 2000).

Emerging Infectious Disease

For centuries, the environment was considered a nidus, or hidden source for transmissible diseases (Pavlosky, 1966). Goodwin (1958) noted that even before etiologic agents had been identified, diseases such as malaria and the plague had been associated with specific habitats. Disease may be endemic to a particular region or habitat, or result from habitat disruption. Pavlovsky (1966) noted that although diseases appeared as new to physicians, they had been in the land, undiscovered, for a long time. This relationship becomes more apparent with increased population and globalization of human society. An increase in population density results in changes in human behavior and habitat alteration, leading to disease emergence or reemergence; potentially increasing human contact with disease carrying organisms (NRC, 1992; Wilson, 1995). Changes in climate and habitat may further result in adaptation or changes within organisms also leading to disease emergence or re-emergence (NRC, 1992; CDC, 1994; Patz et al., 1996).

The concept of disease emergence and reemergence was discussed by the NRC's Institute of Medicine (NRC, 1992) and further promoted by the Centers for Disease Control and Prevention (CDC, 1994) and Levins et al., (1994). The NRC (1992) characterized emerging infectious diseases as an increased incidence of clinically distinct conditions in humans. This definition has been expanded to include infectious diseases whose geographic range, host range, and prevalence have also been increasing in wildlife and plant populations (Daszak et al., 2000; Dobson and Foufopoulos, 2001; Friend et al., 2001).

Levins et al. (1994) characterized infectious disease as that which is brought about by a parasite, also referred to as a pathogen, invading a susceptible animal. The type of parasite could include microorganisms such as bacteria and viruses or multicellular organisms such as protozoa and helminthes. They are dependent on "host" animals for completing a part of their lifecycle. The lifecycle activity within the host results in the disease (Levins et al., 1994).

Microparasites, such as bacteria and viruses, can be introduced to a host either directly or indirectly through a vector. Macroparasites, such as helminthes, have more complex lifecycles and are largely dependent on vectors for disease transmission (Levins et al., 1994). Vectors may include insects such as mosquitoes, flies, ticks, and fleas, as well as rodents and other mammals.

Modeling Disease Emergence

Much of what is known about diseases transmitted through a vector, hence vector-borne disease, has been learned from modeling. Perhaps the first ecological model of disease was Koch's Germ Theory (VanLeeuwen et al., 1999). Koch's model conceptually depicted a stable equilibrium between the environment, host and agent. Disruption to any of these three elements could positively or negatively affect the health status of the host or the disease agent (VanLeeuwen et al., 1999). Ross (1908, 1910) developed the first biomathematical model for vector-borne disease (Bailey, 1982). His pioneering model of malaria later refined by Macdonald (1952), characterized the number of infections that could be distributed by a vector within a community from a single case, also known as the basic reproduction rate.

Although these types of models revolutionized the public health community, they described simple systems containing only two to three variables and assumed that all parasites, or agents, are infectious, and cause one type of disease (VanLeeuwen et al., 1999). Many disease systems, however, are more complex, potentially having both vector and zoonotic components in their transmission (Levins et al., 1994; Real, 1996). Thus, models describing these infectious diseases need to involve more than three variables. In addition, the etiology of many diseases indicates multiple causes with many agents capable of causing more than one disease (Levins et al., 1994).

Since Ross and Macdonald's malaria model, a myriad of infectious disease models have been developed. The 'Susceptible, Infected, and Recovered (SIR)' epidemiologic model was developed to understand the dynamics of epidemics

(Kermack and McKendrick, 1927). This model explores the growth of infection among individuals who are susceptible, infected, or recovered (Sattenspiel, 1990). Hethcote (1976) expanded the SIR model and developed a deterministic, communicable disease model where birth and death rates were evaluated for different age classes in a population. Age classes were further categorized as susceptible, infected, recovered and immune, or recovered and not immune. Using a similar concept, Anderson and May (1979) developed a simple dynamic model explaining disease behavior in populations of laboratory mice with an interest in the consequences of acquired immunity within the host population.

Post et al. (1983) developed a different mathematical model to understand epidemic processes. Post and coworkers used their model to understand the concept of population threshold and spatial arrangement in sustaining disease. They demonstrated that spatial heterogeneity of host populations has an effect on disease thresholds that is dependent on the interaction between the infected and susceptible populations.

Many of the quantitative models of vector-borne disease have built on the original concept of the Ross-Macdonald model estimating the basic reproduction rate. These models were used to demonstrate potential regulatory roles that parasites had on wildlife populations (May, 1993; Dobson and Hudson, 1994; Hudson et al., 1998; Tompkins and Begon, 1999). Understanding the role diseases play in population regulation as well as community dynamics is integral to the development of conservation strategies (Dobson and May, 1986; Hess, 1996).

Macdonald (1980) and Plowright (1982) provided qualitative descriptions of wildlife disease. Macdonald (1980) presented a qualitative argument for the control of rabies by considering the relationship between vector ecology and behavior, which in this case was the fox (*Vulpes vulpes*), and its role in the community. Community dynamics were also suggested to be important in considering control options for the rinderpest virus in Africa (Plowright, 1982). Plowright (1982) observed that contact with cattle (*Bos taurus*), the reservoir for the virus, during herd migration increased juvenile mortality in wildebeest (*Connochaetes taurinus*) and buffalo (*Syncerus caffer*). Plowright (1982) further

noted the importance of host population dynamics and behavior in the epidemic episodes of the rinderpest virus.

Many of the vector-borne disease models are age-classified, Leslie-type models aimed at assessing species fitness, or lambda (λ). Pathogens, however, are closely enmeshed in the environment and animal communities. While there are models that describe community level interactions (Roundy, 1978; Lotz et al., 1995; Miller et al., 2002), their general use is made difficult, largely due to a lack of quantitative knowledge.

Qualitative models are used to understand important relationships and interactions among variables of a complex community system. In particular, qualitative models are useful when variables are difficult to measure (Puccia and Levins, 1985). These types of models have been used in ecology to generate hypotheses or predictions of system behavior in response to perturbations. Puccia and Levins (1991) noted that qualitative models could be used to evaluate the direction of change, resilience, and stability of ecological systems.

Loop and matrix analyses are examples of qualitative models (Puccia and Levins, 1991). Loop analysis has been used to characterize simple predator-prey relationships (Dambacher et al., 1999) and more complex transitions in community composition over time (Ortiz and Wolff, 2002). Based on differential equations characterizing a change in a particular variable over time, loop analysis, a type of signed digraph, provides a pictorial display of a complex (having more than two variables) community that is at or near equilibrium (Levins, 1975; Puccia and Levins, 1985, 1991). From the loop model, a community matrix can be developed.

Qualitative predictions can be developed through an analysis of pulse or press perturbations. Pulse perturbations result in a temporary change in one variable but then returns to its original state. A press perturbation results in a permanent change in a growth parameter of a variable. The direction of change can be predicted from the community matrix (Dambacher et al., 2002).

In Chapter 2, I present and validate from the literature a new procedure to predict changes in risk through a qualitative prediction of vector-borne disease behavior within an ecological community. This procedure builds on the foundation

of the Ross-Macdonald model for vector-borne disease and recent mathematical developments in community ecology. The procedure uses a qualitative modeling approach that can simulate a systems behavior without quantitative parameterization. Results of this approach can generate more focused hypotheses to guide quantitative models.

To illustrate this approach, I constructed a community model for Lyme disease, which is representative of a disease where the ecological relationships are documented. Lyme disease is found in temperate, forested landscapes and is the result of long-term ecological disturbance related to the presence of deer (*Odocoileus virginianus*). It is caused by a spirochete bacterium *Borrelia burgdorferi* that infects ticks, wildlife, and humans (Ostfeld, 1997). Although Lyme disease has likely been present in North America for decades, it reached public attention in the 1970s following the discovery of a cluster of childhood arthritis cases in Lyme, Connecticut (Ostfeld, 1997). The disease is carried by a tick, *Ixodes dammini* (aka *I. scapularis*) found in the northeast or mid-western U.S. and *I. pacificus* in the western U.S. In humans, the disease is first exhibited as a skin rash; neurological problems and arthritis in the knee, hip or other joints can follow in chronic cases.

The second risk analysis tool I developed is presented in Chapter 3. Here I demonstrate the use of Relational Bayesian Modeling (RBM), a model discovery technique using machine-learning technology, to construct quantitative, biologically-consistent models from sparse survey data of the spread of West Nile Virus (WNV). Relational Bayesian Modeling is a method for building models using relational data. It encourages the modeler to interact with the data and develop multiple hypotheses concerning the incidence and spread of the disease as a way of exploring the combined data residing in multiple data tables. The models constructed may be updated as new information becomes available in the form of additional data or expert knowledge contributed by experts.

West Nile Virus was selected as an example of an emerging infectious disease whose ecology is less well known. The WNV produces West Nile Encephalitis. It has recently been found in temperate regions such as Europe and

North America (Komar, 2000). Known as an arboviral disease of birds in particular, it poses a risk to other wildlife, domestic animals, including horses, and humans. The disease was first described in humans from a case in Uganda in 1937 and was characterized as a mosquito-borne virus in Egypt in the 1950s (Komar, 2000). West Nile Virus is endemic in Egypt; over the past 40 years it has spread to several countries in Europe, Africa, the Middle East, Asia, and now North America.

The WNV is a bird virus that is spread by mosquitoes (Komar, 2000). The distribution of WNV suggests that the spread of the virus is related to bird migration, and perhaps commerce (Lundstrom, 1999; Rappole et al., 2000). Of particular concern is the risk of fatal encephalitis in horses, birds, and humans. In North America, the primary vector species is thought to be *Culex pipiens*. *C. pipiens* was first implicated in the transmission of the virus in the New York City outbreak of 1999, but may be only a moderately effective vector (Komar, 2000); other mosquito species were found to be highly susceptible (Enserink, 2000). The primary vertebrate host species appear to be passerine birds. At present, WNV has been detected in over 160 species of birds, and numerous species of mammals including bats (Enserink, 2000; CDC, 2002a). In addition, more than 25 different species of mosquitoes have tested positive for WNV including those active in the morning, daytime, and evening (CDC, 2002a).

The emergence of WNV, typically an "Old World" pathogen, in the "New World" raises the consciousness that vector-borne disease has the potential to spread anywhere environmental conditions are favorable. The existence of a more global economy and increased air travel, enhance this potential through the inadvertent introduction of nonnative organisms, including pathogens.

Use of community models are discussed in Chapter 4 as a means of integrating risk analysis for human and ecological endpoints. The iterative and heuristic nature of these models, improve our ability to evaluate the impacts of human and natural activity on complex ecosystems, including humans. They provide a general, but realistic and practical approach for developing hypotheses concerning the interacting relationships of community members. Issues of uncertainty are accounted for through the analysis of probability distributions. The

value of the use of these models in an integrated risk analysis framework will be to better inform environmental and public health decisions.

CHAPTER 2
COMMUNITY-LEVEL ANALYSIS OF RISK OF VECTOR-BORNE DISEASE

Jennifer Orme Zavaleta and Philippe A. Rossignol

Abstract

Ecological community structure can be a key factor in understanding the risk to public health of communicable disease emergence, the mode of transmission, and control options (Forget and Lebel, 2001). Community structure is particularly important in vector-borne zoonotic diseases with complex life cycles. Population models, such as the Ross-Macdonald model (Bailey, 1982), have been important in developing and characterizing our current understanding of human vector-borne disease. However, these models often by-pass or minimize community-level interactions. In diseases restricted to human hosts, this focus may be of benefit in understanding transmission, but in zoonotic diseases in particular, important community-level considerations may be lost (LoGuidice et al., 2003). Another limitation is that the level of quantification possible in population models may not be achievable in community models. Qualitative community model analysis (Puccia and Levins, 1991) may provide a meaningful alternative to modeling vector-borne disease. We built on recent mathematical developments in qualitative community modeling (Dambacher et al., 2002) coupled with conventional biomathematical models of vector-borne disease transmission, to provide new procedures to analyze risk. Our procedure predicts the change in risk of vector-borne disease from press perturbations, such as control measures, habitat alteration or global warming. We demonstrate the application of this procedure to an oak forest community to predict the risk of Lyme disease. Our predictions of the community dynamics of Lyme disease are consistent with observations observed in the literature

Introduction

Ecological community structure is a key factor in understanding the public health risk of communicable disease emergence, mode of transmission, and control options (Forget and Lebel, 2001). Community structure is particularly important in vector-borne parasitic diseases, where a minimum of three species, namely, host, vector, and pathogen, is involved. In the case of human diseases such as malaria and dengue fever, further zoonotic components are irrelevant or negligible. In the case of zoonotic diseases, disease systems often involve numerous and complex vector and zoonotic components in their transmission, and perhaps more than one host. The number of parameters and variables needed to characterize such vector-borne disease dynamics is greater than those typically used in public health models.

Deterministic and stochastic population models are important in characterizing our understanding of the ecological relationship with vector-borne disease (Dobson and Hudson, 1994). These models stem from the landmark concept of basic reproduction rate developed by Ross (1908, 1910) and Lotka (1923), and later popularized by Macdonald (1952). Bailey (1982) provides a more formal and useful presentation of these concepts. The models were used to demonstrate potential regulatory roles parasites have on animal populations (May, 1993; Dobson and Hudson, 1994; Hudson et al., 1998; Tompkins and Begon, 1999).

In addition to epidemiologic considerations, understanding the role that disease and parasitism plays in population regulation as well as in community dynamics is integral to the development of wildlife conservation strategies (Dobson and May, 1986; Hess, 1996). However, because current disease models focus on population dynamics, they bypass direct consideration of community-level interactions. This omission is due in no small part to insufficient quantitative information needed to model community interactions as well as the lack of appropriate models.

Qualitative community models can provide a practical and rigorous alternative to modeling transmission of vector-borne disease. One form of

qualitative modeling, Loop analysis, involves both signed digraphs and matrix analysis (Puccia and Levins, 1985, 1991). From the signed digraph model, a community matrix can be developed and used to assess stability conditions and to make qualitative predictions of population response to press perturbations in community structure. A press perturbation is a permanent change in a growth parameter such as a birth or death rate. These models are particularly useful in predicting responses to anthropogenic disturbances. Recent mathematical developments have provided a degree of flexibility and reliability that was previously lacking in the approach (Dambacher et al., 2002, 2003a, and b). We present a new procedure predicting changes in risk of vector-borne disease. This procedure predicts system behavior with minimal quantitative parameterization, and evaluates changes in risk of vector-borne disease from an ecological community perspective arising from perturbations, such as habitat alteration or global warming.

Models and Methods

Here we summarize models used in public health and community ecology that we considered in developing a new procedure for qualitatively predicting community-level response to stress and vector-borne disease risk. We apply this procedure using Lyme disease as an example of a vector-borne disease where the disease ecology is well known.

Basic Reproduction Rate

Ross, (1908, 1910) first developed a biomathematical model characterizing the disease status between host and vector populations, later formalized by Lotka (1923). The Ross model, as popularized for malaria by Macdonald (1952), provides a basic model of disease transmission that can apply to vector-borne diseases. The model, often called the Ross-Macdonald model, focuses on the basic reproductive

rate (R_o), which is the number of secondary infections that can arise from a single primary case.

To control malaria during the World Health Organization's campaign, Garrett-Jones (1964) proposed a simplification that focused on the vector. This simplification, referred to as vectorial capacity (VC), is defined as the maximal average daily (at least in malariology; units are otherwise system specific) number of infective contacts possible between vector population and its host (Garrett-Jones, 1964; Bailey, 1982). In the case of malaria where rates are daily by convention, VC is directly proportional to basic reproduction rate. A major practical advantage is that VC is determined solely from the entomological parameters of the Ross-Macdonald formulation of basic reproduction rate. Another benefit of the VC equation is that the impact of an infected vector population on the epidemiology of a disease can be evaluated even in the absence of the parasite (Bailey, 1982). Thus, there are fewer parameters to determine compared with the Ross-Macdonald equation for R_0.

The parameters of VC (*Equation 1*) are: (1) the biting rate (ma), where (m) is the relative number of vectors with respect to host and (a) denotes the biting habit of the vector; (2) the probability of vectors surviving to become infective (p^n), where p is the probability of daily survival and n is the duration of the extrinsic incubation period (a constant under most conditions); and (3) the life expectancy of the vector ($1/-\log_e[p]$). Parameter, a, is the product of the host preference (proportion of competent to non-competent hosts fed upon) to frequency of feeding, which is equal to the inverse of the oogonic cycle in the case of mosquitoes. The biting habit (a) is factored into the equation twice, once to account for the initial bite, then a second time to account for bites that infect a host. The derivation is as follows: A relative number (m) of vectors bite an infected host at a specific rate (a); a proportion (p) of which survive each day of the extrinsic incubation period (n). These infective vectors live for a period ($1/-\log_e[p]^{-1}$) and bite at a rate (a).

$$VC = \frac{ma^2 p^n}{-\log_e(p)} \qquad \textit{Equation 1}$$

In both the basic and daily reproduction rate models, a key variable is probability of daily survival of the vector, which is represented in ($-p^n/log_e[p]$). In the basic reproduction rate model, Ross (1908, 1910) determined that this term, being exponential, could be the most important parameter in malaria transmission, rather than the intuitive, but linear, relative density (m), in considering control options of vector-borne diseases. Garrett-Jones' (1964) concept of VC reinforced this counterintuitive finding. Once the relative abundance, or any other parameter, of vectors falls below a certain threshold, disease will decline to extinction.

Qualitative Community Models

We demonstrate that direction of change following input in the form of a press perturbation in the important parameters of the generalized Ross-Macdonald model, namely, relative density (m), frequency of contact, or the biting habit (a), and vector survival (p) can be evaluated from community models. Community models, in the form of signed digraphs and the corresponding community matrix, are used to describe direct and indirect interactions between populations in a community (Levins, 1975; Puccia and Levins, 1985). Loop analysis has been used to characterize predator-prey systems (Dambacher et al., 1999), and changes in abundance (Dambacher et al., 2002), to predict the impact of species introductions (Li et al., 1999; Castillo et al., 2000), and to explain complex transitions in community composition over time (Bodini, 1998; Ortiz and Wolff, 2002). Experimental comparison of various community modeling approaches suggests that loop analysis was the theoretical approach best suited for predicting the behaviour of complex community structures following a perturbation (Hulot et al., 2000).

Density-dependent interactions, within and between biological variables of a community, form the structure of the community matrix (**A**) (Levins, 1968, 1975). The negative of the inverse of the community matrix ($-\mathbf{A}^{-1}$), is a straightforward procedure that predicts direction of change in abundance of a population within a community following a press perturbation (Bender et al.,

1984). The negative of the inverse is equal to the classical 'adjoint' of the matrix divided by its determinant. Based on the adjoint, Dambacher et al. (2002) derived a 'weighted-predictions matrix' that assesses the indeterminacy of predictions.

Loop analysis allows for a qualitative estimate of interactions among community variables. While a perturbation may only affect one variable of a community directly, other variables are affected as a result of the interconnections within the community (Puccia and Levins, 1985). Perturbations may affect the abundance of organisms in a population and impact other population demographics such as age structure that lead to turnover of the population (Puccia and Levins, 1985). Turnover, the reciprocal of life expectancy of a population (Puccia and Levins, 1985), is determined from the adjoint, or inverse, of the community matrix. Dambacher et al. (submitted) developed an algorithm (see *www.jambrosi.com*) based on the Puccia and Levins (1985) effort for predicting change in life expectancy (e) following a perturbation.

Lyme Disease

Lyme disease is found in temperate, forested landscapes and is the result of long-term ecological disturbance related to the presence of deer. It is caused by a spirochete bacterium *Borrelia burgdorferi* that infects ticks, wildlife, and humans (Ostfeld, 1997). Although Lyme disease has likely been present in North America for decades, it reached public attention in the 1970's following the discovery of a cluster of childhood arthritis cases in Lyme, Connecticut (Ostfeld, 1997). The disease is carried by a tick - vector, *Ixodes dammini* (aka *I. scapularis*) found in the northeast and mid-western United States (US) and *I. pacificus* in the western US. In humans, the disease is first exhibited as a skin rash; neurological problems and arthritis in the knee, hip or other joints can follow in chronic cases.

Transmission and propagation of the disease involves an interrelationship between the tick-vector and three principal hosts: small mammals such as the deer mouse (*Peromyscus leucopus*), deer (*Odocoileus virginianus*) and humans. In this

relationship, deer mice serve as the main reservoir for the bacterium. As tick larva hatch, they become infected when they feed on infected deer mice (Ostfeld, 1997). The infected larva molt into nymphs, considered the principal agent for disease transmission because they are more difficult to detect than adult ticks. Nymphs will infect deer mice, deer and humans. Deer are important hosts in the tick life cycle because male ticks often mate with females while they are feeding on the deer.

Results

To evaluate vector-borne disease risk within the context of a community model, we integrate the parameters of the Ross-Macdonald model, and specifically VC, with loop analysis involving the community matrix. The integration of these concepts allows for predicting change in disease risk in a host population following a press perturbation to a remote variable. In our procedure, risk is defined as vectorial capacity. Changes in key parameters for VC: relative abundance, frequency of contact (host preference), and life expectancy are evaluated from manipulations of the community matrix.

To illustrate our procedure, we constructed a model of a Lyme disease vector-host community (Figure 2.1) based on Ostfeld et al. (1996). The community and adjoint matrices are shown along with their interpretation in Table 2.1; the life expectancy matrix is presented in Table 2.2. Ostfeld and coworkers (1996) suggest that an increased acorn mast (i.e., increased acorn production) would attract deer, mice, and other animals, and result in an increase in ticks that potentially carry Lyme disease, thus increasing disease risk (Ostfeld, 1997). Qualitative predictions developed from a loop analysis lend support to Ostfelds' observations showing increases in population density of mice and ticks following an increase in acorn production (Table 2.1). Similarly, a positive press on gypsy moths would result in decreased acorn production because gypsy moths feed on oak leaves.

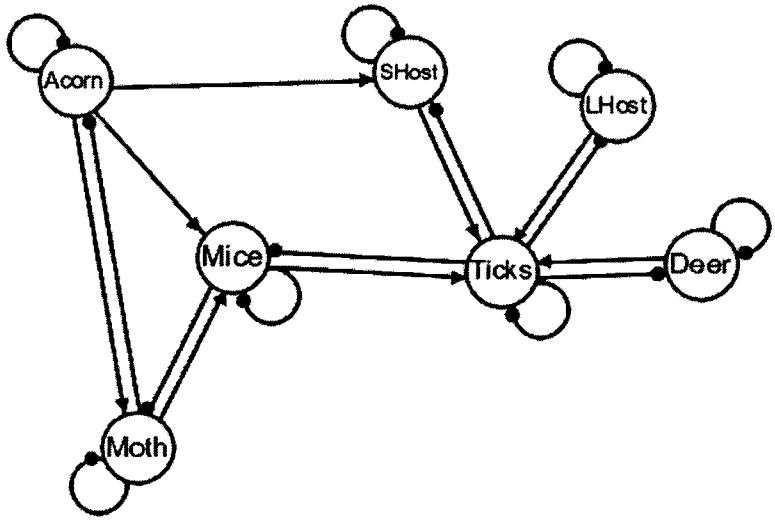


Figure 2.1 Signed digraph of the Lyme disease vector-host community. Circles represent variables; lines represent edges with arrows indicating positive effects and small, dark circles indicating negative effects. Curved lines with small, dark circles are self-regulating effects.

Predicting changes in population abundance, however, does not constitute a complete assessment of risk. Community structure itself also affects risk. Referring to the loop model (Figure 2.1) based on the Ostfeld et al. (1996) model of an oak forest community, we can develop a qualitative prediction of risk from the community matrix, which serves as the basis for determining changes in the parameters of VC (Tables 2.3 and 2.4). Responses depicted in the adjoint and life expectancy matrices serve as an index for the parameters in the VC equation. Thus, a change in relative abundance (m) is determined by a change in the ratio of vector to a competent host following a press perturbation to a variable such as deer (Table 2.3). The adjoint of the community matrix is also used to determine host preference

Table 2.1 Response of community variables following a positive press perturbation within a Lyme disease host-vector community (Figure 2.1). A = community matrix; adj_A = adjoint; W= weighted matrix. A positive press to mice results in an increased abundance in ticks (+), a decrease in deer (-), and an ambiguous (?) impact to small hosts. Responses are determined by comparing the sign of the response from the adjoint matrix for a variable with the weighted value for that variable. Weights <0.5 are deemed unreliable. 'Weight' is a mathematical term accounting for the ratio of positive to negative cycles, or loops, present in the response, that is, the element of the adjoint of the community matrix (see discussions in Dambacher et al., 2002); a weight of 0.5 or greater has been shown to be equivalent to 95% reliability based on simulation studies (Dambacher et al., 2002).

$$A := \begin{bmatrix} -1 & 1 & 0 & 0 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad adj\_A = \begin{bmatrix} 2 & 2 & 4 & -1 & 3 & 2 & 2 \\ -2 & 9 & -4 & 1 & -3 & -2 & -2 \\ -1 & -1 & 9 & -5 & 4 & -1 & -1 \\ 1 & 1 & 2 & 5 & -4 & 1 & 1 \\ -2 & -2 & 7 & 1 & 8 & -2 & -2 \\ -3 & -3 & 5 & -4 & 1 & 8 & -3 \\ -2 & -2 & -4 & 1 & -3 & -2 & 9 \end{bmatrix}$$

$$W = \begin{bmatrix} 0.50 & 0.50 & 1. & 0.33 & 1. & 0.50 & 0.50 \\ 0.50 & 0.60 & 1. & 0.33 & 1. & 0.50 & 0.50 \\ 1. & 1. & 1. & 1. & 1. & 1. & 1. \\ 1. & 1. & 0.20 & 1. & 1. & 1. & 1. \\ 1. & 1. & 0.78 & 0.11 & 1. & 1. & 1. \\ 0.60 & 0.60 & 0.56 & 0.67 & 0.20 & 0.57 & 0.60 \\ 0.50 & 0.50 & 1. & 0.33 & 1. & 0.50 & 0.60 \end{bmatrix}$$

	Ticks	Deer	Acorn	Gypsy Moths	Mice	Small NC*Host	Large NC Host
Ticks	+	+	+	?	+	+	+
Deer	-	+	-	?	-	-	-
Acorn	-	-	+	-	+	-	-
Gypsy Moths	+	+	?	+	-	?	+
Mice	-	-	+	?	+	-	-
Small NC Host	-	-	+	-	?	+	-
Large NC Host	-	-	-	?	-	-	+

Table 2.2 Life expectancy matrix for community variables following a positive press perturbation within a Lyme disease host-vector community (Figure 2.1). The diagonal elements reflect the results of a positive press perturbation through increased (+), decreased (-), or ambiguous (?) death or birth rates. Life expectancy responses are determined by comparing the sign of the response from the delta E death and birth matrices for a variable with the weighted value for that variable. Weights <0.5 are deemed unreliable. 'Weight' is a mathematical term accounting for the ratio of positive to negative cycles, or loops, present in the response, that is, the element of the adjoint of the community matrix (see discussions in Dambacher et al., 2002); a weight of 0.5 or greater has been shown to be equivalent to 95% reliability based on simulations studies (Dambacher et al., 2002). For example, if the site of input is to ticks, the response is an increased death rate/ambiguous response on the birth rate.

$$deltaE\_d := \begin{bmatrix} 11 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 9 & -4 & 1 & -3 & -2 & -2 \\ -1 & -1 & 9 & -5 & 4 & -1 & -1 \\ 2 & 2 & -7 & 10 & -8 & 2 & 2 \\ -2 & -2 & -4 & 1 & 8 & -2 & -2 \\ -2 & -2 & -4 & 1 & -3 & 9 & -2 \\ -2 & -2 & -4 & 1 & -3 & -2 & 9 \end{bmatrix}$$

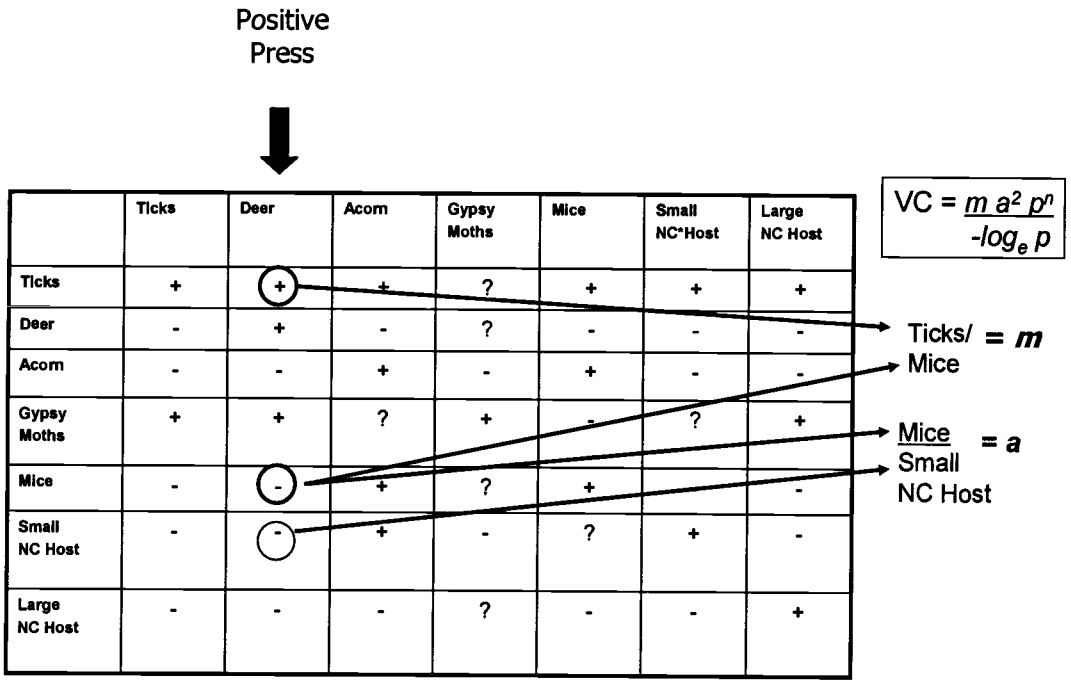
$$deltaE\_b := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & -2 & -4 & 1 & -3 & -2 & -2 \\ -1 & -1 & -2 & -5 & 4 & -1 & -1 \\ 2 & 2 & -7 & -1 & -8 & 2 & 2 \\ -2 & -2 & -4 & 1 & -3 & -2 & -2 \\ -2 & -2 & -4 & 1 & -3 & -2 & -2 \\ -2 & -2 & -4 & 1 & -3 & -2 & -2 \end{bmatrix}$$

$$Weighted\_deltaE\_d = \begin{bmatrix} 0.58 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.50 & 0.60 & 1. & 0.33 & 1. & 0.50 & 0.50 \\ 1. & 1. & 1. & 1. & 1. & 1. & 1. \\ 1. & 1. & 0.37 & 1. & 1. & 1. & 1. \\ 0.50 & 0.50 & 0.14 & 0.053 & 0.50 & 0.50 & 0.50 \\ 0.33 & 0.33 & 0.22 & 0.091 & 0.33 & 0.60 & 0.33 \\ 0.50 & 0.50 & 1. & 0.33 & 1. & 0.50 & 0.60 \end{bmatrix}$$

$$Weighted\_deltaE\_b = \begin{bmatrix} 0. & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.50 & 0.059 & 1. & 0.33 & 1. & 0.50 & 0.50 \\ 1. & 1. & 0.071 & 1. & 1. & 1. & 1. \\ 1. & 1. & 0.37 & 0.053 & 1. & 1. & 1. \\ 0.50 & 0.50 & 0.14 & 0.053 & 0.16 & 0.50 & 0.50 \\ 0.33 & 0.33 & 0.22 & 0.091 & 0.33 & 0.062 & 0.33 \\ 0.50 & 0.50 & 1. & 0.33 & 1. & 0.50 & 0.059 \end{bmatrix}$$

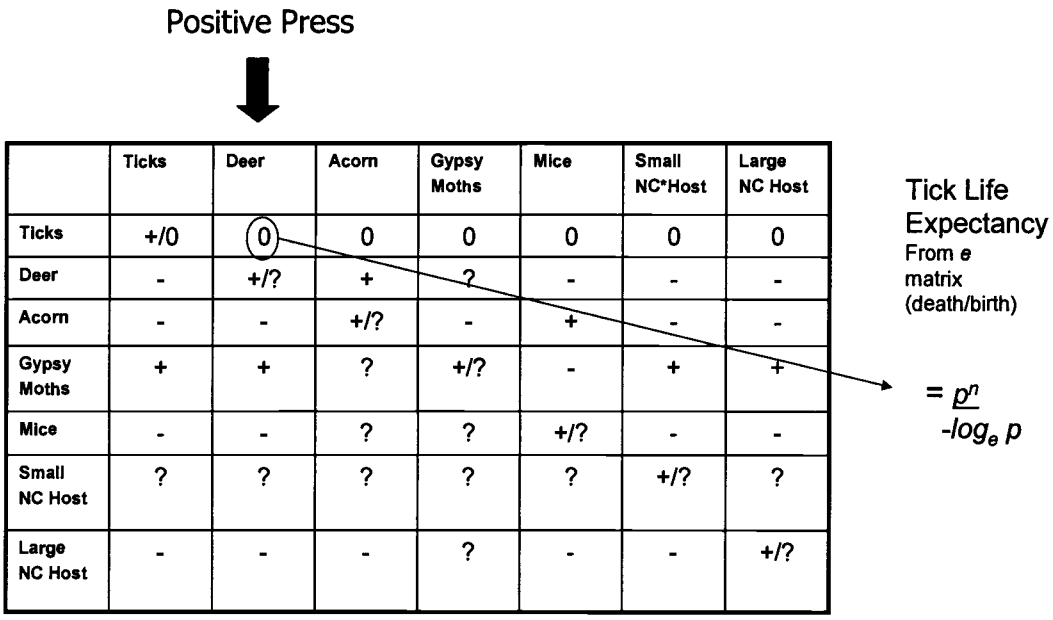
	Ticks	Deer	Acorn	Gypsy Moths	Mice	Small NC*Host	Large NC Host
Ticks	+/0	0	0	0	0	0	0
Deer	-	+/?	+	?	-	-	-
Acorn	-	-	+/?	-	+	-	-
Gypsy Moths	+	+	?	+/?	-	+	+
Mice	-	-	?	?	+/?	-	-
Small NC Host	?	?	?	?	?	+/?	?
Large NC Host	-	-	-	?	-	-	+/?

Table 2.3. Input from adjoint of the community matrix for determining Lyme disease risk. Responses serve as indices for parameters in the equation. Following a positive press to acorns, risk is determined from the ratio of responses for different variables. Parameter 'm' is determined from the ratio of ticks to mice. Parameter 'a' is the ratio of the response in mice to that of small non-competent hosts. The index for life expectancy is shown in Table 2.4.



	Ticks	Deer	Acorn	Gypsy Moths	Mice	Small NC*Host	Large NC Host
Ticks	+	+	+	?	+	+	+
Deer	-	+	-	?	-	-	-
Acorn	-	-	+	-	+	-	-
Gypsy Moths	+	+	?	+	-	?	+
Mice	-	-	+	?	+		-
Small NC Host	-	-	+	-	?	+	-
Large NC Host	-	-	-	?	-	-	+

Table 2.4. Input from the life expectancy matrix for determining Lyme disease risk. Input is determined from the diagonal of the matrix and reflects increase (+), decreased(-), or ambiguous death or birth rates (see Table 2.2).



	Ticks	Deer	Acorn	Gypsy Moths	Mice	Small NC*Host	Large NC Host
Ticks	+/0	0	0	0	0	0	0
Deer	-	+/?	+	?	-	-	-
Acorn	-	-	+/?	-	+	-	-
Gypsy Moths	+	+	?	+/?	-	+	+
Mice	-	-	?	?	+/?	-	-
Small NC Host	?	?	?	?	?	+/?	?
Large NC Host	-	-	-	?	-	-	+/?

(*a*). Assuming a constant contact frequency, change in host preference (a) is estimated from the ratio of the abundance of the competent host (mice) to that of a non-competent host within the community (small hosts). Finally, change in the vector survival parameter ($p^n/-\log_e p$) is determined from the response of the vector in the life expectancy matrix (Table 2.4).

Any change in one of the three parameters of VC might result in a predicted increased or decreased risk. For example, our model predicts that a positive press to deer would increase tick and gypsy moth abundance while decreasing the abundance of acorns, mice and other small hosts (Table 2.3). A positive press to deer results in no change to tick life expectancy (Table 2.4). As a result, risk for Lyme disease would increase due to the increased ratio of tick abundance to mice, parameter *m*. This result is supported by the observations of Wilson et al. (1983) and Lane et al. (1991). They summarize studies conducted on Nantucket and Great Island off the coast of Massachusetts where Lyme disease was endemic. Deer were

drastically reduced from Great Island resulting in a significant reduction of tick populations infesting rodents. Tick populations on Nantucket, where there was no deer intervention, remained stable. Thus, by altering the community structure, relative abundance (*m*) and hence, risk, was reduced.

In another example, Ostfeld (1997) documented that mice and tick populations increased after an increased oak mast, increasing the infection rate of nymphal ticks with the bacterium that causes Lyme disease, *Borrelia burgdorferi*, thus increasing the risk of Lyme disease. While our procedure supports Ostfeld's observations of changes in tick and mice abundance (Table 2.3), the epidemiologic implications are less clear. We can predict that a positive press perturbation increases the abundance of mice and tick populations. However, taking the ratio of these responses from the VC equation, the parameter (*m*) is unchanged. A positive press to acorns also increases the abundance of small, non-competent hosts, thus the ratio for (*a*) also remains unchanged. A positive press to acorns has no impact to tick life expectancy (Table 2.4) suggesting overall, little to no impact on disease risk. To increase risk, there would either need to be a decrease in mice abundance while tick abundance remained constant, or a decrease in the abundance of small non-competent hosts relative to mice. In fact, the risk of Lyme disease has been suggested to decrease with increased biodiversity of a community where additional non-competent hosts serve as a dilution factor (Mather et al., 1989; Ostfeld and Keesing, 2000; LoGuidice et al., 2003).

The application of our procedure provides a qualitative mechanism for evaluating vector-borne disease risk within a complex community. A perturbation to a variable such as acorn production is likely to reverse, whereas the removal of oaks from a forest or changing the population density of non-competent hosts such as deer would cause a more permanent change in community structure and thus affect disease risk. Similar to the quantitative use of VC to control vector borne disease, our qualitative procedure allows for predictions of community response following press perturbations with no quantitative parameterization.

Discussion

We present a novel procedure for analyzing vector-borne disease behavior within an ecological community. Our procedure integrates VC, a measure of disease transmission, with community variables and adds a new dimension to public health analysis of vector-borne disease behavior at a community level. A qualitative community analysis provides useful predictions of the impacts of anthropogenic change such as habitat availability, or that which impacts population density of vectors and hosts within the community.

Our approach differs from many of the vector-borne disease models that are age-classified, Leslie-type models aimed at assessing species fitness (Anderson and May, 1979; Hudson et al., 1998) or simulation models that are used to estimate spread of disease (Nicholson and Mather, 1996; LoGuidice et al., 2003). These types of models address impacts of disease on an individual species, and do not address the whole ecological community. It is important to consider community interactions where zoonotic pathogens are closely enmeshed in ecological communities. While there are models that describe community-level interactions (Roundy, 1978; Lotz et al. 1995; Miller et al. 2002), their general use is made difficult due to a lack of quantitative knowledge.

The advantage to our modeling approach is that it redefines a traditionally quantitative population-level model, VC, in the context of qualitative community interactions. Through the use of our procedure, reasonable and rigorous predictions of vector-borne disease risk can be generated from changes in community structure. Our procedure, however, is not as precise as population models for human diseases associated with impoverished conditions such as malaria, where community diversity is low, or where there is direct transmission that does not include a vector. However, qualitative analysis is better suited to address poorly specified complex systems. For those diseases that are vector-borne and zoonotic, our procedure can effectively predict an ecological community response to a perturbation, which in turn can generate focused hypotheses to guide data collection and control management strategies as interventions.

CHAPTER 3
DISCOVERY AND INTERACTIVE DEVELOPMENT OF A COMMUNITY-LEVEL MODEL OF DISEASE TRANSMISSION: WEST NILE VIRUS IN MARYLAND

Jennifer Orme Zavaleta, Jane Jorgensen, Bruce D'Ambrosio, Hans K. Luh, Fredrick W. Kutz and Philippe A. Rossignol

Abstract

Understanding interactions among pathogens, hosts, and the environment is important in developing a rapid response to a disease outbreak. In order to deploy the most rapid response possible, we must exploit existing data to its maximum extent to determine plausible mechanisms and patterns (temporal and geospatial) of disease spread. These data often are observational in nature, and collected during independent survey efforts. We demonstrate the use of Relational Bayesian Modeling (RBM), a model discovery technique using machine-learning technology and relational data, to construct quantitative and biologically-consistent models of West Nile Virus (WNV) spread. Survey data on WNV cases in mosquitoes, horses, humans, and birds in Maryland collected during 2001, along with information on tire clean-up sites and collection facilities in Maryland were explored using this technique. Our results indicate a strong association between tire license sites and birds infected with WNV, and that WNV positive birds serve as good indicators for infected mosquitoes and humans. Thus, RBM shows promise as a tool to determine complex community interactions relevant to disease transmission that could guide monitoring and control strategies during the early stages of an outbreak or during an ongoing outbreak of a relatively rare disease.

Introduction

The U.S. National Research Council (NRC) Committee on Grand Challenges in Environmental Sciences identified the inter-relationship between infectious disease and the environment as one of four important environmental research challenges for the next generation (NRC, 2000). Research is needed to improve our understanding of the interactions among pathogens, hosts, and the environment to affect change in the infectivity and virulence of organisms posing a threat to populations of plants, wildlife and humans (NRC, 2000).

The emergence of diseases that are transmitted directly from person to person often reflects changes in human population density, where as vector-borne disease emergence is an indication of environmental changes (Epstein, 1994). Vector-borne disease, particularly those that are zoonotic in origin, may be endemic to a particular region or habitat, or result from habitat disruption. Emergence of vector-borne disease has become an issue associated with increased human population and globalization of human society (Patz et al., 1996). Increased human influence on the environment results in habitat alteration leading to disease emergence or reemergence (NRC, 1992; Wilson, 1995). Changes in climate may also result in disease emergence or re-emergence (NRC, 1992; Centers for Disease Control (CDC), 1994; Patz et al., 1996). Community-level models that address the interactions between infectious disease and the environment could be useful tools for understanding and predicting disease outbreak and spread that are tempered by the pressures of an increasing human population.

Traditionally, epidemiologists employ highly structured and comprehensive methods to gather quantitative information, establishing cause and effect relationships between environmental stressor(s) and disease. This approach is time consuming and resource-intensive, particularly during a disease outbreak. As a possible alternative, we demonstrate the use of Relational Bayesian Models (RBM), discovered in relational data using machine-learning technology, as a rapid means of investigating and predicting the mechanisms and temporal and geospatial patterns of disease spread. Relational Bayesian Models can maximally exploit

existing and largely observational data that are collected during independent survey efforts. The models developed from RBMs are represented as Bayesian networks (BNs) that link information from observed and possibly highly correlated data.

We used an RBM to construct qualitative and quantitative, biologically-consistent models of disease spread from sparse, uncertain survey data. Our objective was to determine whether RBM would serve as a rapid, realistic, and practical tool to generate hypotheses related to the transmission of West Nile virus (WNV) in Maryland.

Models and Methods

The necessity of rapid response to a developing disease outbreak often precludes the investigation of plausible mechanisms and temporal and geospatial patterns of disease transmission. In order to deploy the most rapid response possible, we must exploit existing data to its maximum extent. These data are usually collected from independent surveys, containing varying degrees of uncertainty or gaps in quantitative information. For this reason, we chose to discover RBMs in existing observational data using Cleverset 'Modeler' (Jorgensen et al., 2003) in our analysis of disease transmission. Modeler is an RBM tool developed for model discovery and data exploration in relational databases; a beta version of this model can be obtained from Cleverset, Inc., upon request. Relational Bayesian Models are a type of probabilistic relational model that is an extension of a BN (Getoor et al., 2001). The RBMs discovered by Modeler are represented as BNs, which provide complete representations of the joint probability distribution over the entire set of variables in the model. Relational Bayesian models may be used as a tool to frame multiple, simultaneous hypotheses concerning these variables. We constructed a common frame of reference to temporally and spatially relate data collected in independent efforts that reside in independent tables. Using this frame of reference, Modeler heuristically examined all possible models that

could be derived using the available variables to discover those models that parsimoniously describe relationships among the variables in the model. These relationships form the basis for hypotheses about the key factors involved in transmission of the disease and the manner in which disease spreads.

Model Description

Relational Bayesian Models provide qualitative information on the structure of a domain, as well as quantitative information in the form of probability distributions describing correlations among components in the domain. The domain in this case is the community probabilistically associated with a particular disease. The structure of the domain is summarized by a directed acyclic digraph comprised of nodes representing variables and arcs extending from 'parent' nodes to 'child' nodes, representing conditional dependencies (Figure 3.1) (Ramoni and Sebastiani, 2001). The direction of the arc indicates a probabilistic, though not necessarily causal relationship between nodes. The conditional dependencies are quantified by the conditional probability distributions underlying the structure of the graph (Jensen, 2001).

Modeler performs model discovery by examining the set of variables specified for inclusion in the model and those it derives from the data. In this analysis, the data consisted of an enumeration of WNV cases. Modeler aggregated these instances in counts and evaluated the mean number of cases over a range of conditions (for example, the mean number of human cases found in a geographic location where a WNV positive bird had been found the previous month).

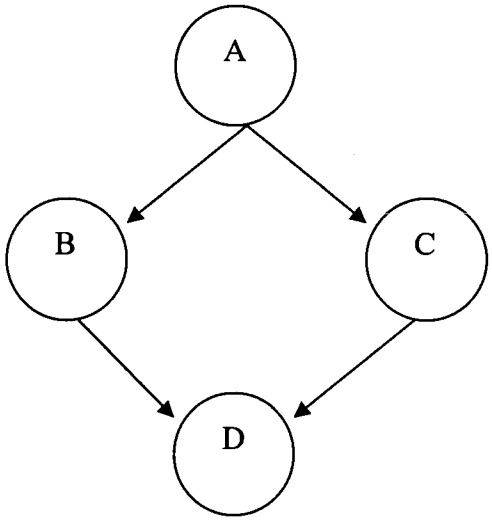


Figure 3.1 Hypothetical directed acyclic graph. "A," "B," "C," and "D" represent variables. The arrows represent arcs indicating conditional linkages between variables. "A" is the parent node of "B" and "C." "B" and "C" are the parent nodes of "D."

To construct the BN, Modeler performs a heuristic search to identify all possible models. Modeler incorporates specific data by selecting a variable pair with the highest mutual information (see Results for further discussion). A BN is then constructed beginning with this variable pair using Bayesian Information Criterion (BIC) as a scoring mechanism (Getoor et al., 2001). The algorithm adds, deletes, or changes the direction of arcs connecting the variables in the BN. The BIC imposes a penalty for those models that have a large number of parameters and is composed of two parts: the prior probability of the structure and the probability of the data given that structure. The BIC balances the complexity of the structure with its fit to the data (Getoor et al., 2001). Thus, those models with the largest BIC have a better fit to the data. When there are no other variables satisfying the BIC for inclusion in the model, the algorithm then tests whether further modifications are needed to identify the best possible BN for that run of the RBM. The models constructed by Modeler can be updated as new information becomes available in the form of additional data or expert knowledge (see Jorgensen et al., 2003).

The final RBM produced by this analysis is a mixed model, part human and part machine involving the interaction between machine learning and expert

knowledge. Through this interaction, different scenarios can be explored to enhance and refine hypothesis generation. Because this technique relies heavily on input of transdisciplinary expert knowledge and interpretation, judgment is used to determine when a meaningful model has been produced. Human experts must be able to transform the data into appropriate formats, construct a relational framework that Modeler will use to analyze the data, and eliminate redundancies in the BN developed by Modeler.

Model Application

We used this modeling approach to explore existing data and to address multiple hypotheses concerning the incidence and spread of WNV in Maryland during 2001. West Nile virus is a disease where the ecological dependency on vector and host populations as well as the ecological conditions necessary for disease outbreak is uncertain (Figure 3.2).

The mosquito-borne WNV causes West Nile encephalitis, considered an emerging infectious disease. It has recently been found in temperate regions such as Europe and North America (Komar, 2000). Known as an arboviral (arthropod-borne) disease of birds, it poses a risk to other wildlife, domestic animals such as horses, and humans. The disease was first described in humans from a case in Uganda in 1937 and was characterized as a mosquito-borne virus in Egypt in the 1950s (Komar, 2000). The virus, endemic in Egypt, has spread over the past 40 years to several countries in Europe, Africa, the Middle East, Asia, and now North America. The first North American human case of WNV occurred in New York City in August, 1999 (CDC, 1999); WNV has rapidly spread across the country reaching the West Coast of the United States by 2002 (CDC, 2002a).

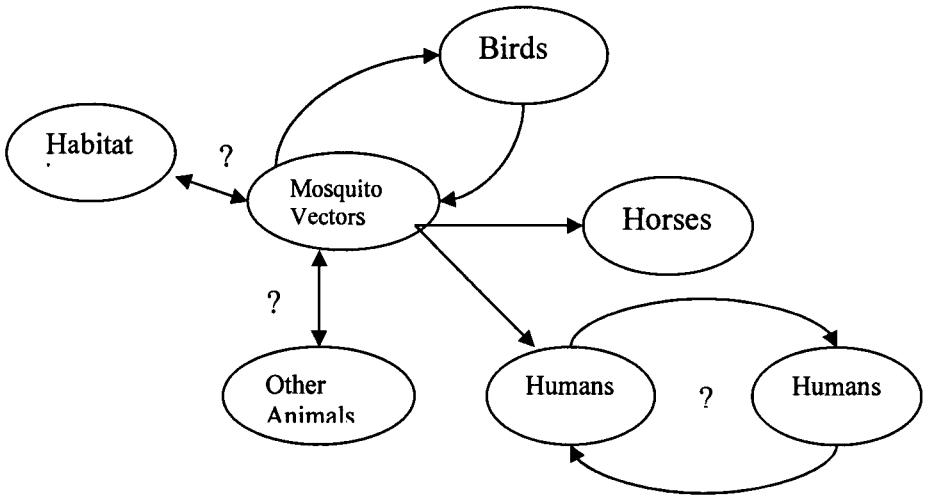


Figure 3.2 West Nile Virus transmission. Arrows represent direction of viral transmission. Those transmissions noted with question marks are postulated but not known. Other animals include native wildlife and domestic or exotic animals.

West Nile virus is transmitted between mosquito vectors and bird hosts (Figure 3.2). Adult mosquitoes acquire the virus in a blood meal from an infected avian host. The virus resides in the mosquito salivary glands where it is amplified through continuous transmissions between mosquitoes and avian hosts. An infected bird can be infectious for 1-4 days after which the bird, if it survives, develops a life-long immunity. Therefore, a sufficient number of vectors must feed on an infective host to cover the extrinsic incubation period of about 14 days (Cornel et al., 1993). Information on the specific species that serve as competent vectors and hosts is still being gathered. In North America, the primary vector species is thought to be *Culex pipiens*. *C. pipiens* was first implicated in the transmission of the virus in the New York City outbreak of 1999, but may be only a moderately effective vector (Komar, 2000); other mosquito species such as *Aedes* were found to be highly susceptible (Enserink, 2000). Transmission to horses, humans and wildlife is thought to occur from non-*Culex* mosquitoes and arthropods. Natural infections in both hard (*Hyalomma marginatum*) and soft (*Ornithodoros maritimus*)

ticks, along with swallow bugs (*Oeciacus hirundinis*) have been reported in Africa, Europe, and Asia (Komar, 2000).

It is not known whether animals other than birds can serve as competent hosts and whether there are specific habitats associated with the transmission of the virus. Early reports on the disease outbreak in the United States suggested that horses and humans were non-competent hosts because they were unable to develop a viremia sufficient to infect mosquitoes (Komar, 2000); however recent findings suggest that infected humans may directly infect other humans through blood transfusions, organ transplants, breast milk, and intrauterine exposure (CDC, 2002b, c, and d).

The information used to conduct this analysis is the same as those summarized by Kutz et al. (2003). Specifically, data on positive and negative mosquito pools (pooled samples from mosquitoes caught in the same trap on a given date) for WNV were provided by the Maryland Department of Agriculture. Mosquitoes were collected in light/CO_2 traps placed in locations typical of mosquito control operations around the country. The main criteria for trap placement included urban areas where residents granted permission, areas that were free of vandalism, and likely mosquito habitat such as a freshwater or saltwater marsh. If a trap did not capture mosquitoes, then it was moved to another location. The traps were emptied daily and mosquitoes frozen to ensure virus isolation. The Maryland Department of Health and Mental Hygiene performed viral analyses.

Information on the number and location of WNV cases reported in horses, birds, and humans from 1999 to 2001 was also available from the Maryland Department of Agriculture. The Maryland Department of Health and Mental Hygiene provided information on human cases. Further, our analysis used information on licensed tire collection facilities and tire clean up sites (potential mosquito breeding sites) provided by the Maryland Department of the Environment.

Modeling Procedure

We constructed RBMs using the WNV data for 2001, the only year out of three where data were available for all the variables of interest. These data included instances of birds that tested positive for WNV (positive birds), positive mosquito pools, negative mosquito traps, licensed tire storage or disposal facilities, human and horse cases recorded in individual data tables. Date of discovery and geographical location of the cases were also included in the data tables. We associated these independent data tables across time and space to construct a composite model of disease spread across positive birds, positive mosquito pools, licensed tire facilities (both storage and disposal), human and horse cases (Figure 3.3).

Knowing the date and location (latitude and longitude) enabled us to establish spatial and temporal links. We did this by imposing scale across both the spatial and temporal dimensions by creating two additional variables, 'geocell' and 'month.' They were defined by dividing the State of Maryland into 5-mile square geocells and time into months. A 5-square mile geocell was based on the average distance flown by birds and mosquitoes in a day (Klowden, 1995; Verbeek and Caffrey, 2002). We associated the data tables containing the instance data with geospatial and temporal adjacency tables creating a relational database.

Modeler examines the location and time of a case and determines the set of geocells included in the model. By setting the initial table to positive birds, we initially incorporated all geocells containing positive birds into the model. If we had started with using a table containing fewer cases, the initial number of geocells incorporated into the analysis would have been fewer. For example, starting with six human cases in three geocells would result in an initial model containing three geocells. The 732 positive bird cases would then be evaluated with respect to these three initial geocells. Thus, by starting with the table with cases in the most geocells, we created the most comprehensive model given the data in the combined data tables.

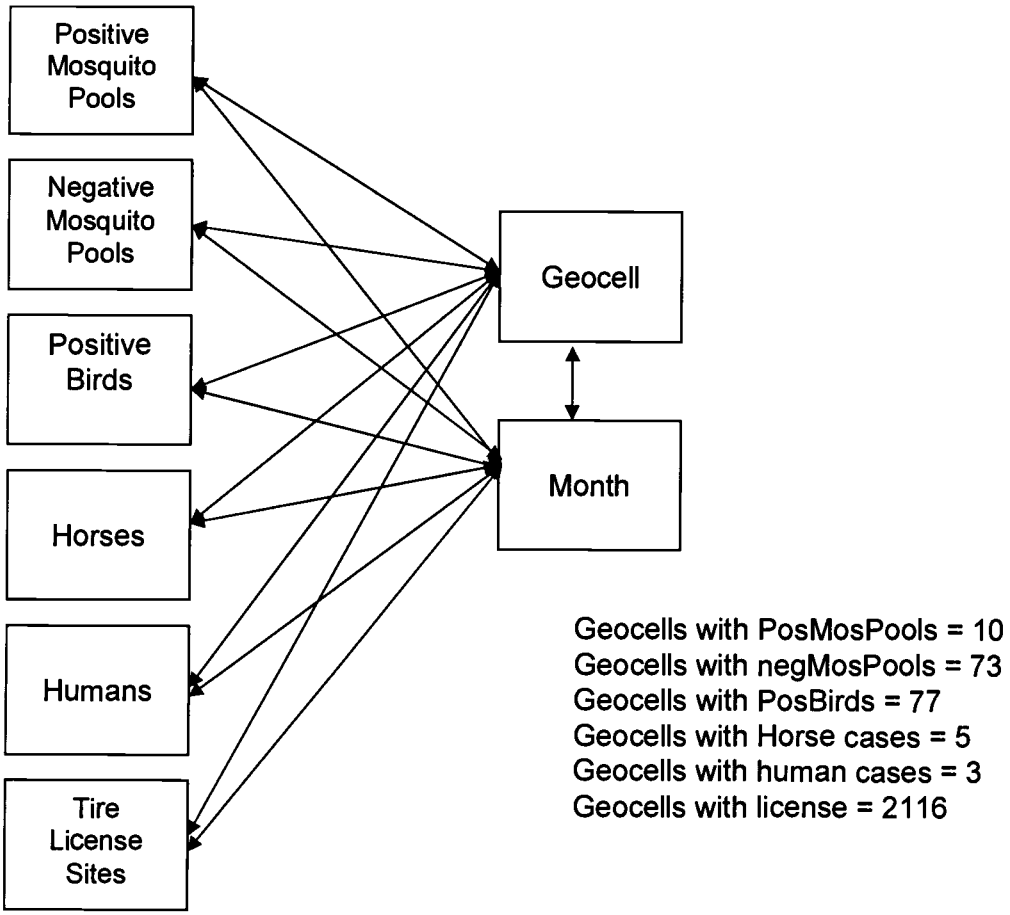


Figure 3.3. Development of relational database for West Nile Virus. Modeler developed a relational database by conducting a series of pairwise correlations between two variables: 'geocells' and 'month' (see Modeling Procedure) and independent data sets. The number of positive and negative mosquito pools, positive birds, horses, humans, and tire license sites per geocell is given.

Those geocells containing positive birds were evaluated with respect to licensed tire storage and disposal sites, negative mosquito pools, positive mosquito pools, human cases and horse cases in the same and adjacent geocells in the same and previous month (Figure 3.4). When the number of licensed tire facilities, positive mosquito pools, human and horse cases contributed information to the model that improved model fit by BIC, these variables were added to the model. Cases that were not located within the same or adjacent geocells to any positive

bird case were not considered in determining the distributions of that particular variable in the BN.

Modeler also considered temporal correlation. Variables in the model were counts for a given geocell for a given time period (current month and previous month; see Figure 3.4). Arcs between variable pairs considering space and time were created when the variables were found to improve the model based on BIC.

The version of Modeler used for this analysis takes discrete data as input. Continuous variables of interest such as the number of positive birds in a geocell in a month were discretized (placed in probability "bins") in order to include them in the model. Modeler automatically discretizes variables on an exponential scale into 'probability buckets.' For example, the discretization for human and horse cases was: Category I: one case; Category II: two cases; and Category III: more than two cases, reflecting the number of positive cases in a geocell in a month. Because we are using Modeler for exploration and hypothesis generation for trends in disease spread, point values are not important, rather, we are interested in the direction and magnitude of the effect that contribute to a hypothesis.

Results

We used Modeler to correlate the independent, observational data for license sites, positive birds, positive mosquitoes, horses, and humans across space and time. We considered including landscape information, (i.e., landscape classified as agricultural, forested, water, urban) into the analysis but the available data were at too coarse a scale for this particular analysis. In general, we found that the structure of the BNs produced using Modeler is consistent with what we know of disease transmission and reflects the biological relationships that are inherent from the data.

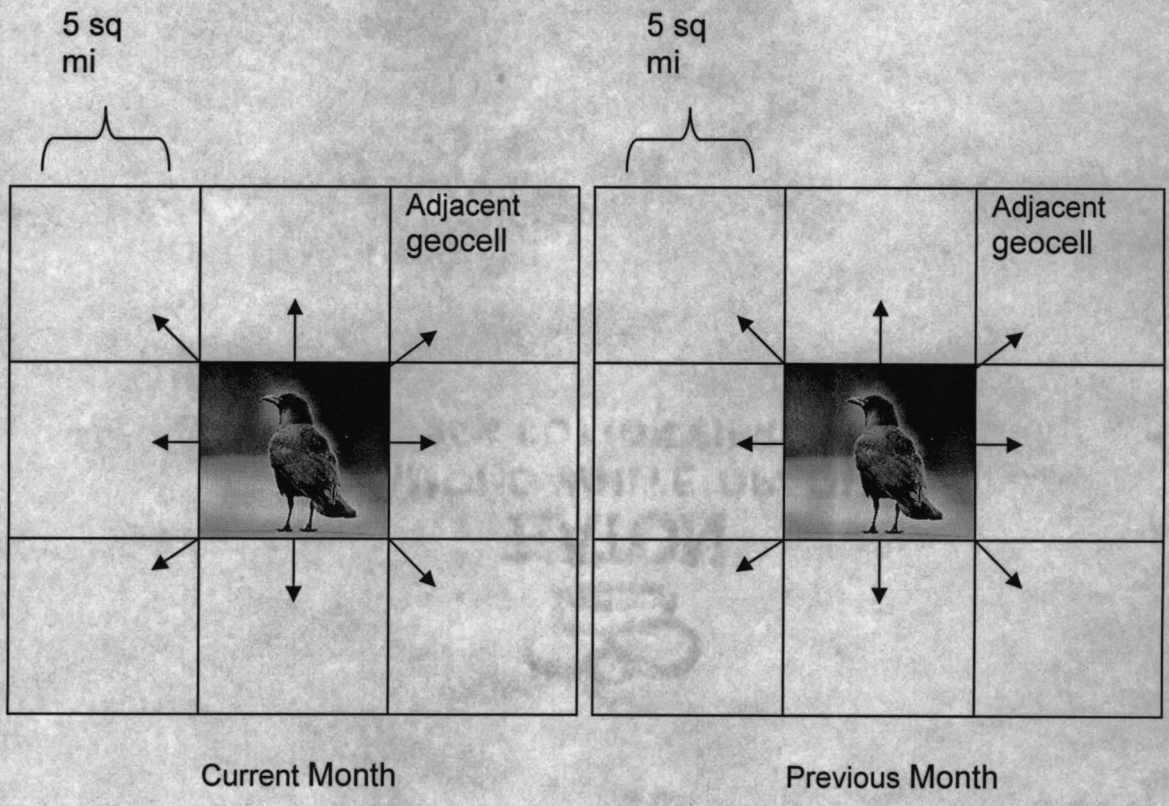


Figure 3.4. Adjacency relationships in space and time. Modeler correlated data by relating data in one geocell with adjacent geocells within the same and previous months.

Starting with positive birds and a full BIC, Modeler constructed a BN (Figure 3.5) showing a correlation between positive birds and license sites as a central organizing feature of the model because it is highly connected to other nodes in the model. We aggregated tire storage and disposal license sites into one variable after finding no difference when they were included in the model as separate variables. The highly connected nodes represent those variables that Modeler found to be linked probabilistically with many other variables.

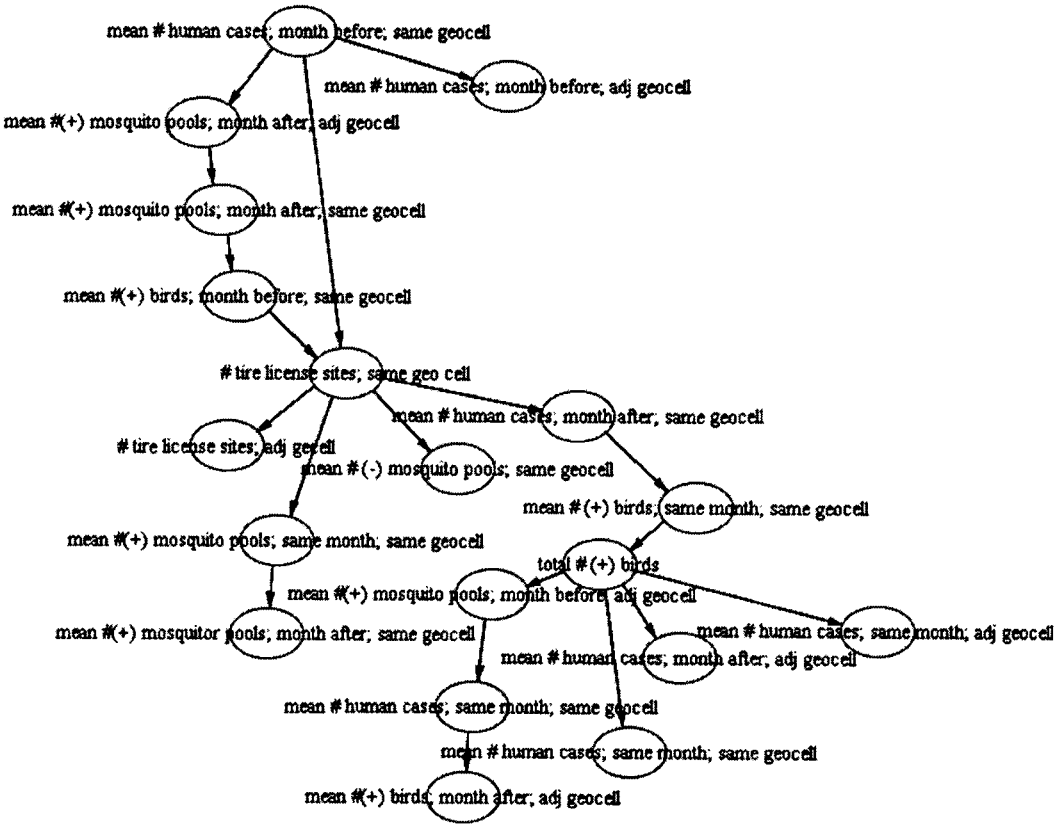


Figure 3.5. Bayesian network of positive birds for West Nile Virus in Maryland during 2001 developed using a full Bayesian Information Criterion. Variables represent mean number of cases found with respect to a positive bird case. Cases may be reported in the same month, previous month or month after a positive bird was found, in the same geocell or adjacent geocell.

Because the data were so sparse, the model constructed with the full penalty in the BIC were relatively uninformative with few links indicating determinants of human cases. In order to introduce more complexity into the model, we relaxed the penalty in the BIC to 0.9 (Figure 3.6) and 0.8 (Figure 3.7) to create a more connected model.

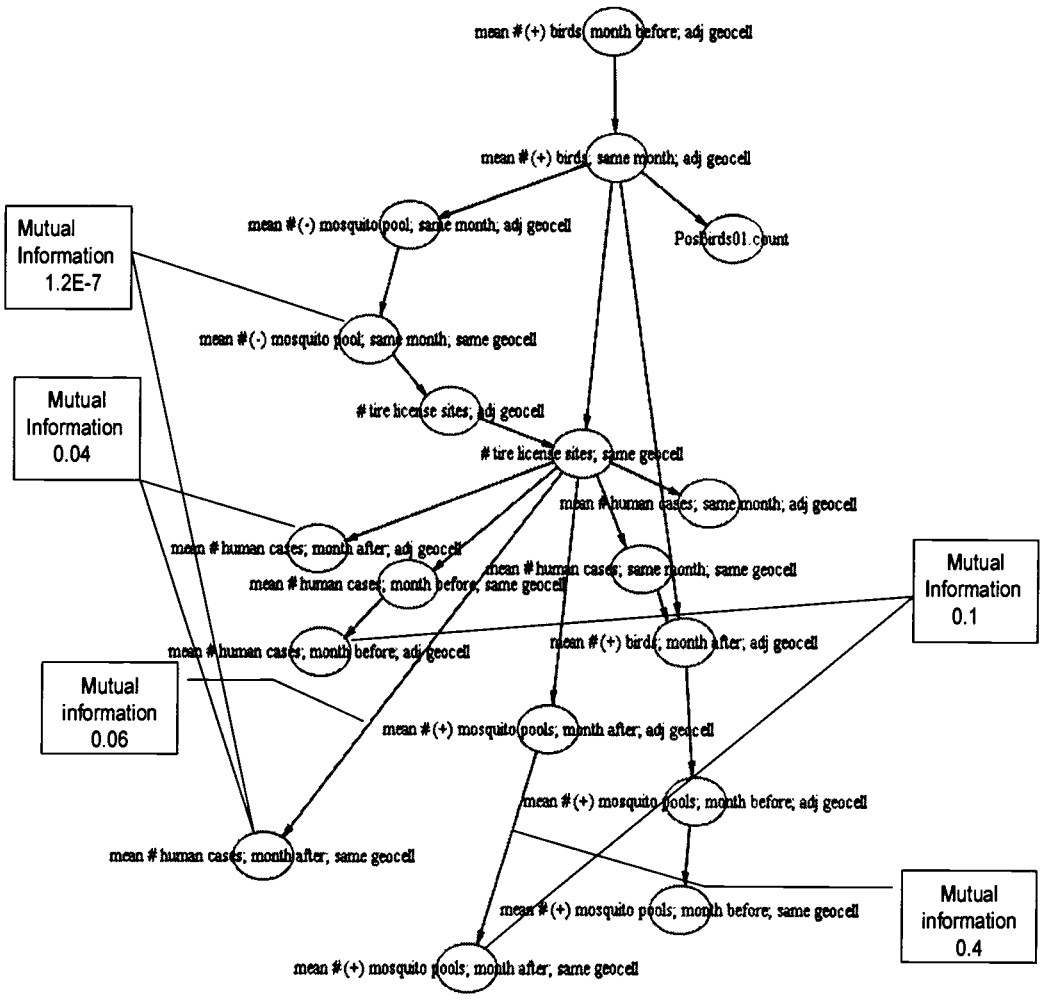


Figure 3.6. Bayesian network of positive birds for West Nile Virus in Maryland during 2001 developed using a Bayesian Information Criterion of 0.9. Variables represent mean number of cases found with respect to a positive bird case. Cases may be reported in the same month, previous month or month after a positive bird was found, in the same geocell or adjacent geocell. Mutual information between variables is a measure of the shared information between variables, and indicates the change in one variable that would result from a change in another.

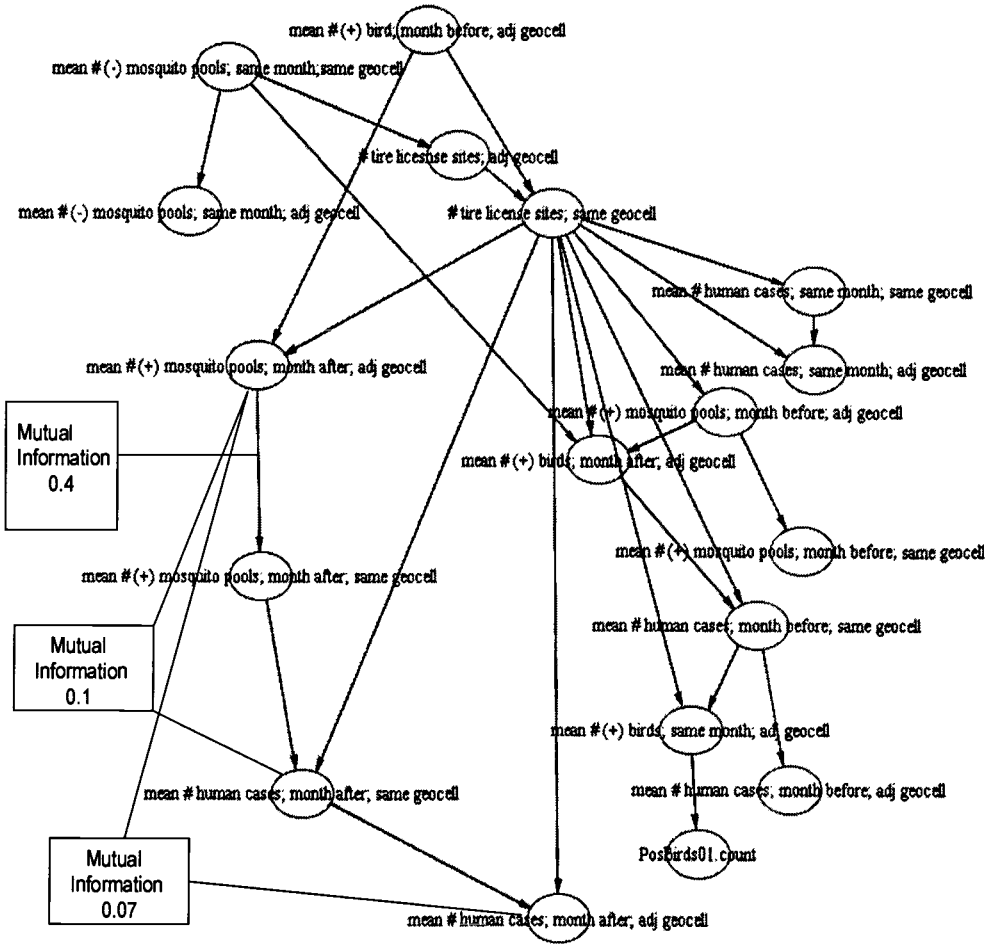


Figure 3.7. Bayesian network of positive birds for West Nile Virus in Maryland during 2001 developed using a Bayesian Information Criterion of 0.8. Variables represent mean number of cases found with respect to a positive bird case. Cases may be reported in the same month, previous month or month after a positive bird was found, in the same geocell or adjacent geocell. Mutual information between variables is a measure of the shared information between variables, and indicates the change in one variable that would result from a change in another.

When we create a BN using a relaxed BIC of 0.8, we add complexity to the model, creating linkages not represented in the models created with a BIC of 0.9, or a full BIC. In all cases, the node representing positive birds associated with tire license sites is an organizing feature. However, in the BN developed with a BIC of

0.8, a spatial and temporal pattern of nodes associating positive birds, positive mosquito pools and humans in the same or adjacent geocell during the same or previous month becomes apparent.

The BN networks we developed using Modeler are not unique and are specific to the relational dataset. For example the addition of hypothetical 40 tire license facilities to geocell 853 (the city of Baltimore) produced a different model (Figure 3.8). However, the commonalities among these models provide information about the problem at hand. For example, the number of horse cases did not contribute additional information to any models regardless of model configuration or settings to construct the model (i.e., full or reduced BIC). This does not mean that horses are not biologically important in the mechanistic model, but that they do not contribute information to the probabilistic model. Positive birds on the other hand are prominent in all models across space and time and thus can be inferred as an important indicator of disease.

The linkages within a BN can be further evaluated by examining the mutual information for a specific node. Mutual information is a measure of shared information between two variables and indicates which variables are more strongly correlated such that a change in one would result in a change in the other. Focusing on positive birds as an indicator of human cases referring to the BN developed using a BIC of 0.9, the variable "mean # human cases; month after; same geocell" (which represents the mean number of human cases found the month after a positive bird was found in the same geocell) has the highest level of mutual information (0.06) shared with the variable "# tire license sites; same geocell" (Figure 3.6). Similarly, the variable "mean # human cases; month after; same geocell" has a high level of mutual information (0.4) with the variable "mean # human cases; month after; adjacent geocell" indicating that positive bird cases serve as a good indicator of human cases in space and time. The least amount of mutual information is shared with this variable "# tire license sites; adjacent geocell."

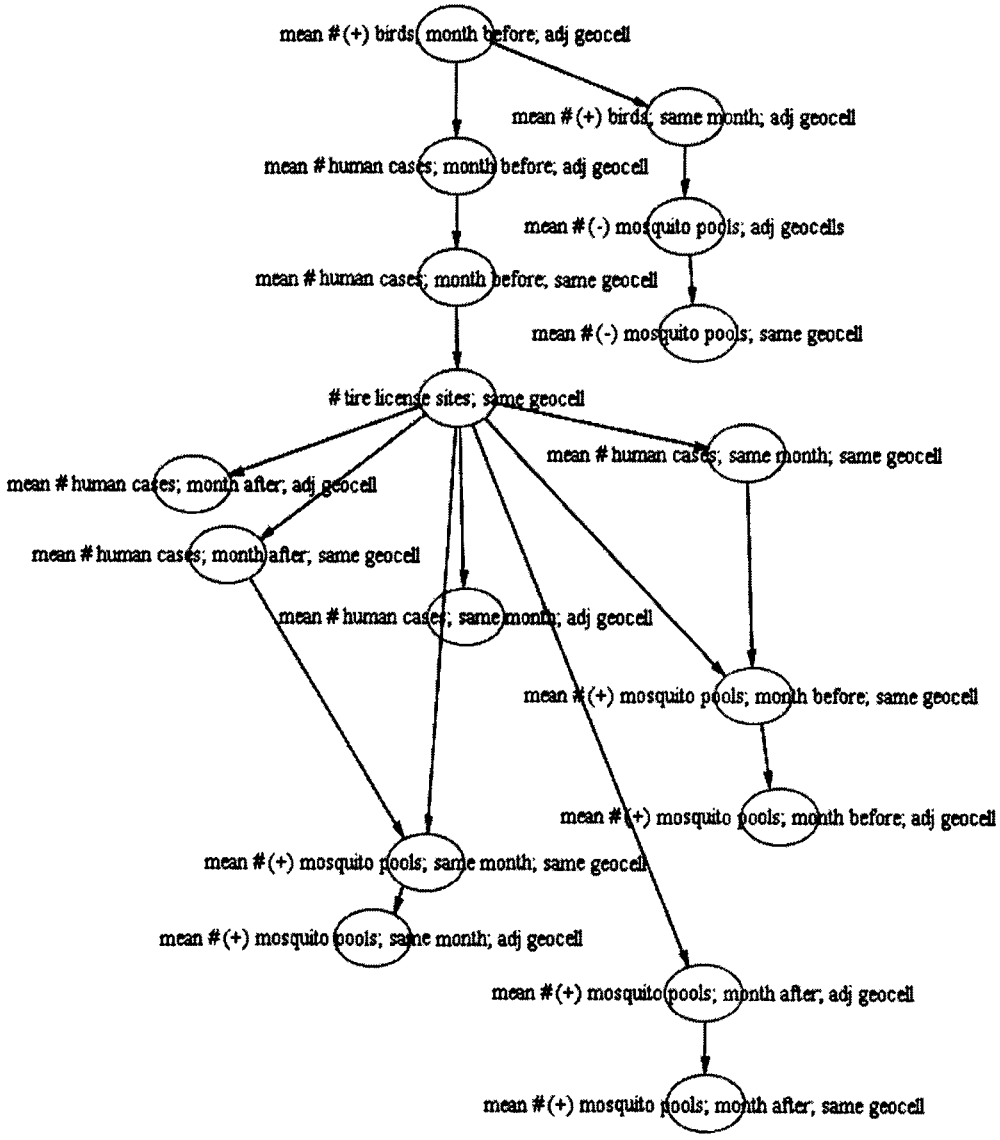


Figure 3.8. Bayesian network of positive birds for West Nile Virus in Maryland with 40 additional hypothetical tire sites and full Bayesian Information Criterion penalty. Variables represent mean number of cases found with respect to a positive bird case. Cases may be reported in the same month, previous month or month after a positive bird was found, in the same geocell or adjacent geocell.

If we examine positive birds as an indicator for positive mosquito pools, the variable "mean #(+) mosquito pools; month after; same geocell" has the highest amount of mutual information (0.4) shared with "mean #(+) mosquito pools; month

after; adjacent geocell" (Figure 3.6). This information also demonstrates a spatial and temporal relationship.

Looking at the series of connected nodes from the BN developed with a BIC of 0.8 (Figure 3.7) we find that the variable relating positive mosquito pools in the adjacent geocell the month after a positive bird has the highest mutual information (0.4) with it's child node representing human cases in the month after a positive bird in the same geocell. The amount of shared information with the positive bird-positive mosquito in the same geocell variable decreases down the series of linkages, with the 'grandchild' node representing human cases occurring a month after a positive bird in the adjacent geocell having a mutual information value of 0.07.

The BN is essentially qualitative; however, quantitative information can also be obtained from conditional probability tables. This information can be represented in a tree format derived from the raw, not discretized data. The conditional probability table for positive birds in the same geocell as number of tire license sites indicates the association between the two variables (Figure 3.9). In geocells with more than 109 tire license sites, there is a greater probability of finding more than two cases of positive birds. This finding suggest that in the aggregate, geocells with more than 109 tire license sites are different with respect to the number of positive bird cases, than geocells with less than 109 tire license sites. The number of license sites is a result of the modeling algorithm. The number "109" for tire sites could represent the density of potential breeding habitat (standing water in stored tires) for mosquitoes. It could also serve as a measure of urbanization that increases not only breeding habitat but also indicates a certain human and bird population threshold for disease prevalence.

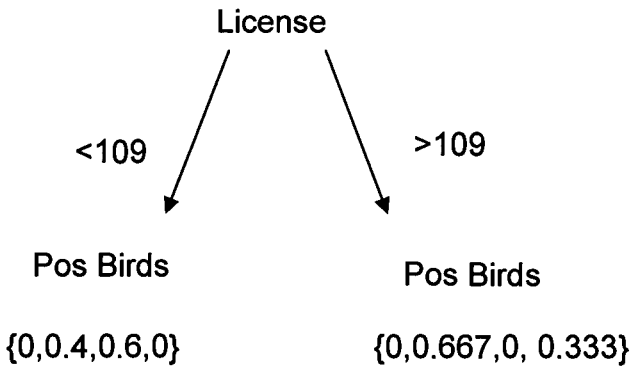


Figure 3.9. Conditional probability tree for positive birds in same geocell with license sites. Probability categories are NA, one case, two cases, and more than two cases. Thus the probability of finding more than two cases of birds positive for West Nile Virus in geocells with greater than 109 tire license sites is 0.333.

Similarly, the distribution of positive birds in adjacent geocells the month after a positive bird is found is also associated with the number of tire license sites. The probability increases for the mean number of positive birds in the adjacent geocell when the adjacent geocell contains more than 109 tire license sites (Figure 3.10).

Conditional probabilities determined for the mean number of human cases in the adjacent geocell the month after a positive bird was found indicates that the probability of finding more than two human cases in the same geocell a month after a positive bird was found is 1 (Figure 3.11). Thus, even with sparse data, a tentative quantitative estimate can be made regarding the use of positive birds as an indicator of disease incidence in humans.

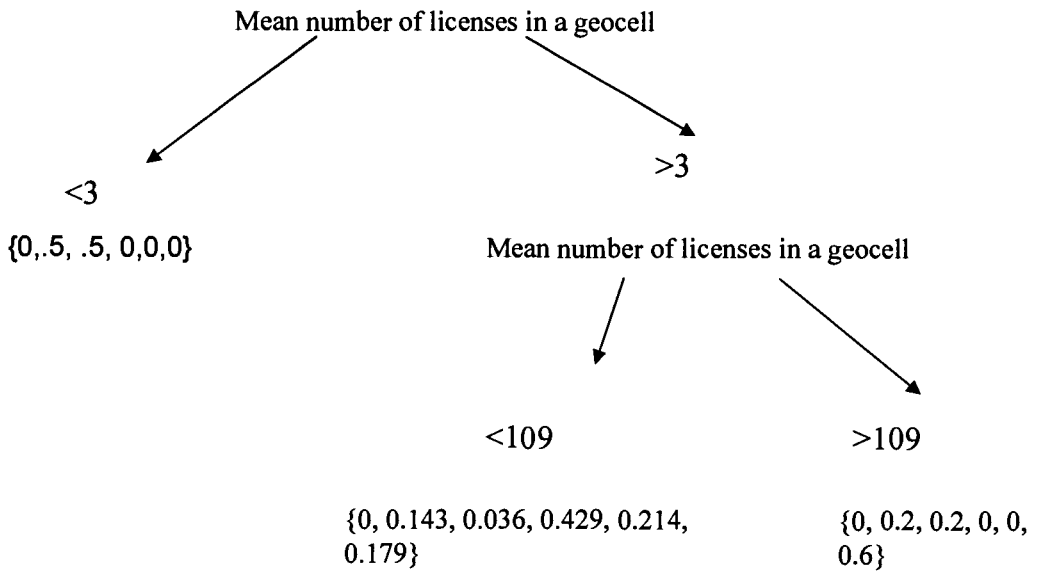


Figure 3.10. Conditional probability tree for the mean number of positive birds in an adjacent geocell the month after a positive bird is found. The probability categories are NA, 1, 2, 6, 20, more than 40. Thus the probability of finding more than 40 positive bird cases in an adjacent geocell one month after a positive bird is found is 0.6.

Discussion

We demonstrate the use of probabilistic RBMs as a new tool to assist public health professionals meet the challenge of responding quickly and effectively during a disease outbreak. The RBM is a tool used to discover models from independently collected, observational data. The RBM can be used to identify probabilistically, those variables that may be important during an outbreak, even from sparse survey data that would not otherwise be useable. These variables may identify indicators or important conditional relationships that can be used to guide disease surveillance or management control strategies. The RBM also has predictive capabilities that could be used to develop hypotheses related to disease spread.

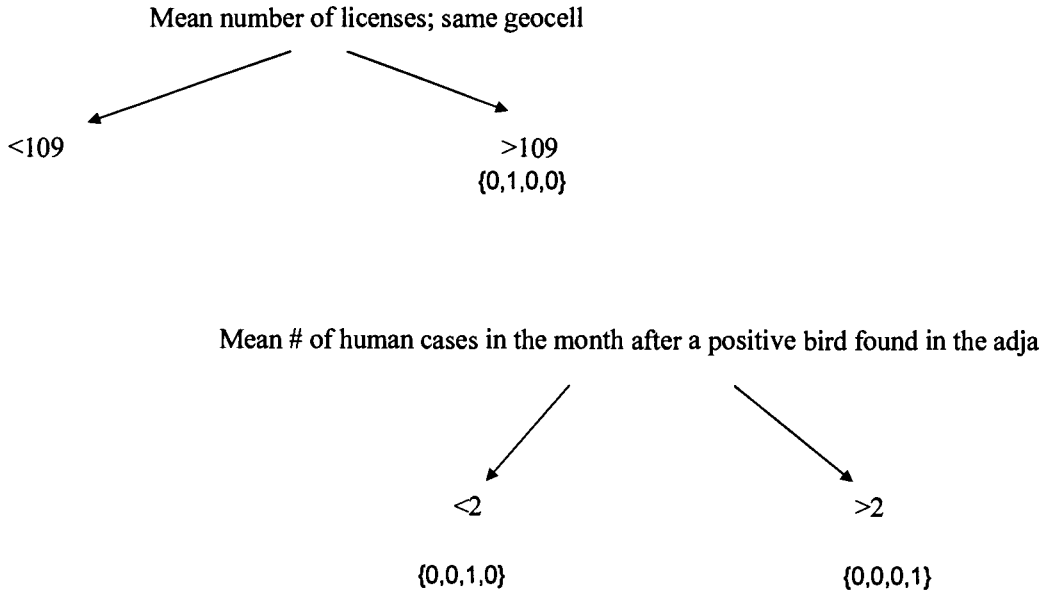


Figure 3.11. Conditional probability tree for the mean number of human cases in the adjacent geocell the month after a positive bird was found. Probability categories are NA, one case, two cases, and more than two cases. Thus the probability of finding more than two human cases after finding a positive bird in geocells with greater than 109 tire license sites is 1.

In the analysis of WNV spread, we found that the associations between positive birds, positive mosquito pools and human cases support the findings of Kutz et al (2003). Kutz and co-workers used geospatial techniques to study the potential impact of two nonnative mosquito species on the epidemiology of WNV in Maryland. Using licensed tire storage sites as a representative source for vectors, they demonstrated a spatial convergence of the WNV, the nonnative mosquito vectors and susceptible hosts, primarily in urban landscapes. They concluded that the two nonnative species had a high potential to serve as 'bridge' vectors transmitting the disease to noncompetent hosts: horses and humans. Our analysis identifies positive mosquito pools, irrespective of species, as an indicator for disease transmission across space and time.

The two primary benefits to this modeling approach are visualization and model discovery (Jorgensen et al., 2003). The RBM produces a visually intuitive interpretation of disease transmission with a quantitative foundation. The BN

developed is a graphical representation of all the joint probability distributions of all the variables contained within specified databases. Therefore, analysts can explore the strengths of relationships contained within the data. Further, we are able to generate and update hypotheses as more data become available. Thus, this modeling technique allows for a continuum of analysis in which variables can be added, deleted, or modified to generate a new model, without starting anew. The model discovery approach is scalable to large datasets with greater complexity.

Although not conducted in the analysis of WNV, RBMs can be used to model bias (such as selection or reporting) and confounders that may exist within the data, to determine the significance of its influence on hypothesized associations between variables. Modeling bias and confounding occurs through manual insertion of a variable valued, for example, as high, medium or low, in the database. Associations between the inserted variable and disease could quantitatively indicate the influence of bias or confounding in that particular analysis.

A limitation to this technique is that the models developed are not necessarily unique. The models developed reflect the data at hand and the particular settings of the Modeler to explore changes in model structure. While the exact configuration of the network is limited to the specific data set, general trends in disease can be observed across and within other datasets. Further, this technique does not allow testing of point values, a feature of more traditional, frequentist statistics.

The RBM technique has been used to explore biological data. Getoor et al. (2001) describes the application of this technique to generate hypotheses relating patients from a tuberculosis clinic, various risk factors, and specific strains of tuberculosis. Jorgensen et al. (2003) demonstrated the ability of the RBM to identify factors related to water clarity in exploring the behavior of the Crater Lake ecosystem. Here, we demonstrate the use of the RBM technique to determine those complex community interactions that are relevant to the transmission of WNV. We found a strong probabilistic relationship between the number of licensed tire storage and disposal sites and the mean number of birds that are positive for the

virus. These positive birds serve as spatial and temporal indicators for the mean number of positive mosquito pools and mean number of human cases. These results support and extend the findings of Kutz et al. (2003) using a different, geospatial technique.

Seldom do public health professionals have the luxury of formulating disease control campaigns with complete knowledge. In nearly all cases they must make decisions given the data at hand balanced with risks should decisions be deferred. The pressure is to act early and effectively. As such, this novel technique may prove to be a valuable tool for evaluations of disease outbreak, particularly in instances where little is known about transmission and data are sparse.

Chapter 4
INTEGRATIVE APPROACHES TO RISK ASSESSMENT

Jennifer Orme Zavaleta, Anne Fairbrother, Philippe A. Rossignol,
and W. Daniel Edge

Abstract

Environmental risk assessment and its use are changing from stressor-endpoint specific assessments for use in command and control types of decisions to an integrated approach for application in community-based decisions. This change reflects the challenge for environmental public policy to address more complex scientific problems. As a result, the process of risk assessment and supporting risk analyses are evolving to characterize the human-environment relationship. Integrating risk paradigms combine the process of risk estimation for humans, biota, and natural resources into one assessment to improve the information used in environmental decisions (Suter et al., 2003). A benefit to this approach includes a broader, system-wide evaluation that considers the interacting effects of stressors on humans and the environment, as well the interactions between these entities. To improve our understanding of the linkages within complex systems, risk assessors will need to rely on a suite of techniques for conducting rigorous analyses characterizing the exposure and effects relationships between stressors and biological receptors. Many of these analytical techniques are narrowly focused and unable to address the complexities of an integrated assessment. In this paper, we discuss qualitative community modeling and Relational Bayesian modeling techniques that address these limitations and evaluate their potential for use in an integrated risk assessment.

Introduction

The environmental risk assessment paradigm is shifting from independent analyses of human health or ecological effects to a more integrative, or unifying, approach. The idea of integrating risk assessments has been the topic of extensive discussion over the past decade. Integration ideally would combine the process of risk estimation for humans, biota, and natural resources into one assessment to improve the information used in environmental decisions, resulting in more effective protection of both humans and the environment (Suter et al., 2003). A benefit to this approach is a broader, system-level evaluation that considers the interactions of the effects of stressors on humans and the environment, as well the interactions between these entities. In addition, stressors other than chemicals need to be considered. The basis for such an integrated approach would be the perspective that ecosystems serve as part of the foundation defining human well-being.

Risk assessments are important tools for informing public health and environmental protection decisions. They constitute the scientific reasoning for estimating the likelihood of an adverse human or ecological effect resulting from exposure to a stressor. Although the human health and ecological risk assessment paradigms were developed independently, they are related. In both paradigms, risk characterization is a key step providing a description of the weight of the evidence concerning the hazard, potential exposures, and the uncertainties, variability, and assumptions used in the assessment. Thus, the integration of risk assessment approaches could be encapsulated in risk characterization and the analytical processes it entails.

The shift in risk assessment to an integrated approach is consistent with changes in the scientific approach to complex problems. In many instances, a multidisciplinary approach is a necessity to fully evaluate cause and effects relationships. Wilson (1998) noted that science is no longer a specialized activity, but involves the synthesis of causal explanations. Thus, scientific research is

shifting towards understanding linkages within highly complex systems (Vitousek et al., 1997; Wilson, 1998; NRC, 2000; Forget and Lebel, 2001).

To improve our understanding of the linkages of complex systems as part of an integrated risk assessment, risk assessors must rely on a suite of techniques for conducting rigorous analyses characterizing exposure and effects relationships among stressors and biological receptors. Current analytical techniques have been criticized as inadequate and irrelevant; they can be misinterpreted due to a lack of understanding of the problem and the inability to deal with uncertainty (NRC, 1996; Peterman and Anderson, 1999). Further, many of the commonly used techniques are narrow in focus and unable to adequately evaluate complex systems. In this paper, we review community-level modeling techniques that account for these limitations and evaluate their potential for integrated risk assessment.

Risk Assessment Paradigms

The human health risk assessment paradigm (Figure 4.1) was first popularized by the National Academy of Sciences (NAS) (NRC, 1983). Their intent was to bring about consistency in health assessments within the U.S. federal government, but their influence extended throughout the national and international scientific communities. The NAS paradigm focused initially on humans exposed to chemical stressors posing a cancer risk and was quickly applied to other, non-cancer health effects (e.g., developmental, reproductive, or neurotoxicity). It begins with a qualitative description of hazard to determine whether exposure to a substance results in an undesired effect. Once a hazard has been identified, a dose-response assessment determines the potential magnitude of the hazard. Relying largely on experimental animal studies, or human studies to the extent available, the dose-response assessment develops a quantitative estimate relating an exposure dose to the human biological response. An exposure assessment describes the fate and transport of the substance from source to the receptor, including the likely delivered dose to the site of toxicity. All of this information is then coalesced into a

risk characterization. Risk characterization describes the overall likelihood and magnitude of an adverse effect resulting from exposure to a substance. The adequacy of the database, models used, assumptions, uncertainties, and overall confidence in the risk estimate are communicated through the risk characterization (USEPA, 1995). This step is particularly useful in informing risk management decisions.

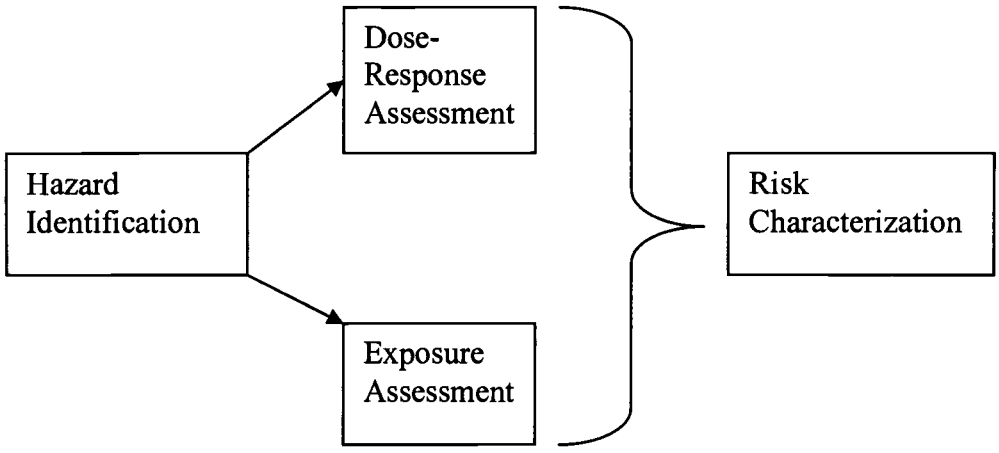


Figure 4.1. Human health risk assessment paradigm. Adapted from NRC, 1983

Building on the NAS paradigm, the *Guidelines for Ecological Risk Assessment* (USEPA, 1998) was developed to evaluate risks of chemicals and other stressors in complex ecosystems, including problems that may extend across temporal and spatial scales at different levels of biological organization. Ecological risk assessment (Figure 4.2) begins with problem formulation, a conceptualization of the problem, including an assessment plan. Problem formulation explores working hypotheses and defines the analytical steps to be included in the assessment. It also includes the identification of assessment endpoints and measures of effect. Many of these elements are policy decisions that are informed by science (Lackey, 1997). After problem formulation, an analysis step involves

characterization of both exposure and effects. This phase is similar to the human health paradigm of exposure assessment and dose-response assessment. The analysis step produces exposure and stressor-response profiles summarizing the relationship between exposure and receptors. Stressor-response profiles may be developed for chemical as well as non-chemical stressors. Risk characterization makes up the final phase of the assessment. Under the ecological guidelines (USEPA, 1998), risk characterization includes a discussion of the supporting evidence and overall degree of confidence in the risk estimate, along with an interpretation of the adversity of ecological risks.

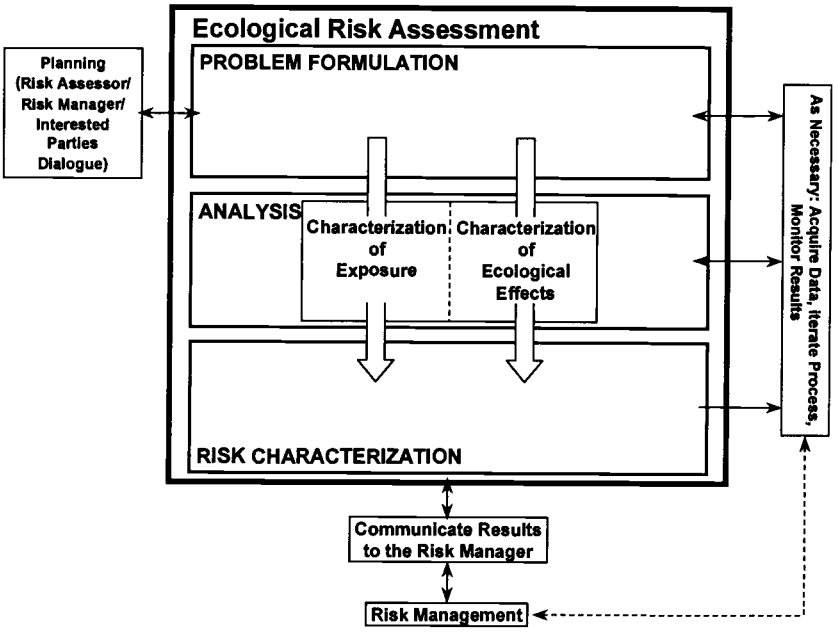


Figure 4.2. Ecological risk assessment paradigm (USEPA, 1992, 1998).

Integrated Risk Assessment Paradigms

Over the past decade several frameworks for integrating risk have been proposed that are based on the human health and ecological paradigms described above. For example, Harvey and coworkers (1995) developed a 'holistic' approach that consisted of parallel *and* integrated health and ecological assessments. Their

process followed the steps originally outlined by the NRC (1983) conducting human health and ecological assessments in parallel. A series of risk choices is produced for the risk manager by integrating the results of two parallel assessments during the risk characterization step. Using mercury as a case study, they developed a risk characterization consisting of a series of risk estimates developed for humans exposed through inhalation or ingestion that address neurological or reproductive effects, and for wildlife exposed through the aquatic food chain addressing reproductive success and decreased species distributions. The authors suggested that the series of risk estimates would provide options for risk managers to choose from in making a decision (Harvey et al., 1995).

Although cast as a holistic process, the Harvey et al. (1995) approach is not really integrative, but rather a comparison of different risk values generated for different exposure scenarios and toxicity endpoints; protective of different species. Thus, this approach may be too generic and unresponsive to a particular problem or management decision.

A special forum of the World Health Organization's International Programme on Chemical Safety (IPCS) developed another approach. They outlined an integrated process combining elements of both human health and ecological processes (Suter et al., 2003). This paradigm (Figure 4.3) is more closely aligned with the concepts of the *Guidelines for Ecological Risk Assessment* (USEPA, 1998). Here, hazard identification becomes an element of problem formulation, and dose response assessment occurs as part of the effects characterization. Most importantly, this approach considers the interactions among stressors and receptors such as wildlife or humans, and the abiotic environment.

One distinct difference of the IPCS integrated approach from the Harvey et al. (1995), NRC (1983) and ecological risk paradigms (USEPA, 1998) is the involvement of stakeholders and risk managers in the process. The human health and ecological risk paradigms were designed to be independent from risk management so that the outcome reflects scientific analyses that are not influenced by socio-political bias. In the IPCS approach, stakeholder and risk management involvement throughout the process is viewed as essential to ensure buy-in and

responsiveness of the assessment to the specific problem, considering both human and ecological risks where applicable (Suter et al., 2003). While this, in and of itself does not ensure integration, it increases the potential depending on how the problem is defined at the onset of the risk assessment.

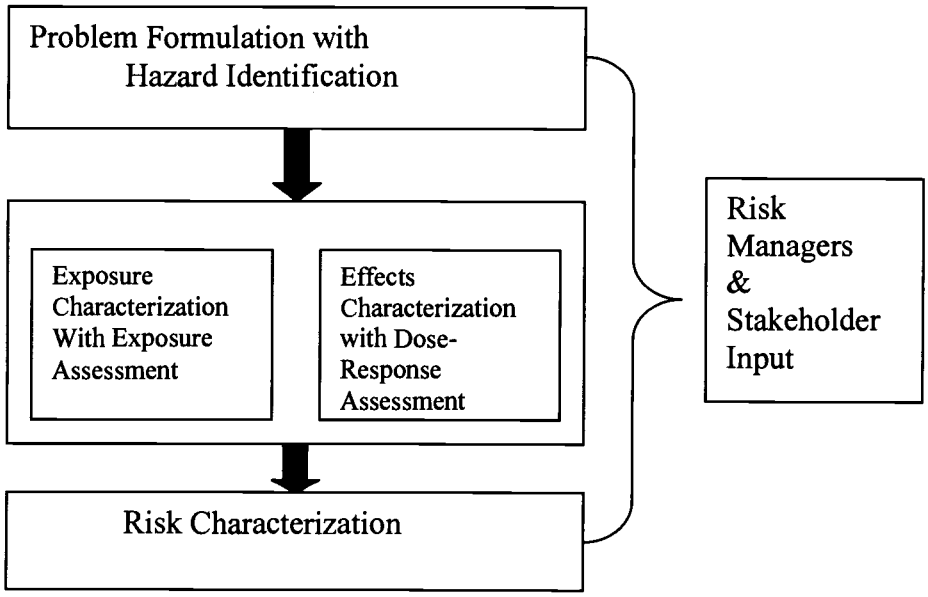


Figure 4.3. Integrated risk assessment paradigm. Adapted from World Health Organization (WHO), 2001.

The IPCS approach combines the process of risk estimation for humans, biota, and natural resources into one assessment for the purpose of improving the information used in environmental decisions, resulting in more effective protection of resources valued by society (Miranda et al., 2002; Suter et al., 2003). Integration is achieved through all phases of the risk assessment process (Suter et al., 2003). Under problem formulation, integration entails the development of stressor-driven assessment questions common to both health and environmental questions that focus on potential susceptible human and ecological endpoints. Exposure and effects characterizations are integrated through an evaluation of all the possible sources of exposure and an understanding of common modes of toxic action in humans and other organisms. Similar to the holistic approach (Harvey et al., 1995), the IPCS risk characterization includes multiple estimates of risk from which a best

estimate of human and ecological risk is selected using a common and consistent approach (Suter et al., 2003). The authors go on to indicate that evidence for health and ecological risks would be integrated when appropriate but do not describe how this would be achieved.

The IPCS integrated approach was applied to several complex environmental problems (Table 4.1). The case studies developed using the integrated approach identify aspects of where integration can or should occur with respect to exposure and effects characterization, but they do not actually conduct an integrated assessment. Rather, they describe how to integrate risks. The risk characterization section in each of the case studies largely reflects parallel risk comparisons. Two studies (Ross and Birnbaum, 2003, Vermeire et al., 2003) propose a common quantitative approach, a Toxic Equivalency Factor (TEF) approach as a means of integrating risks. It is not clear, however, that having a common quantitative approach to estimate risks for different species is actually integrative, but rather reflects the commonalities in the toxic endpoints and mechanisms of toxicity for the exposures and species of interest. Thus, the IPCS approach goes beyond the holistic approach in describing levels of integration throughout the risk assessment process. However, the information included in the risk characterization step presents parallel risk estimates for human and ecological endpoints under different exposure and effect scenarios. The responsiveness of the assessment to a particular problem is likely to be greater under the IPCS approach given the interaction with risk managers and stakeholders throughout the process.

Other approaches to integrative assessments have been proposed that focus on human and environmental linkages including socioeconomic and political factors. Epstein (1994) developed an integrated assessment framework of climate change and ecosystem vulnerability. His generalized framework depicted overlapping and interacting climate and social systems with ecosystems whose intersection directly or indirectly produces various outcomes ranging from changes in health, crop yields, and demography to economic productivity (Figure 4.4).

Table 4.1. Summary of International Programme for Chemical Safety (IPCS) integrated risk assessment case studies.

Environmental Problem	Assessment Endpoints	Areas of Integration	Proposed Risk Characterization	Reference
"Dioxin-like" Persistent Organic Pollutants	Humans and upper trophic level wildlife	• Route of exposure • Mode of action • Toxicity	Apply Toxic Equivalency Approach (TEF) to both humans and wildlife	Ross and Birnbaum, 2003.
Tributyl- and triphenyltins	Humans and piscivorous wildlife	• Route of exposure • Mode of action • Toxicity	Species and exposure-specific human and ecological risk estimates	Sekizawa et al., 2003.
UV-Radiation	Amphibians, coral, humans, and oceanic primary productivity	• Exposure pathways • Mechanistic pathways	Parallel characterization of risk across assessment endpoints.	Hansen et al., 2003.
Organophosphorous pesticides	Humans and wildlife	• Exposure pathways • Toxicity	Species-specific TEFs	Vermeire et al., 2003.

Epstein noted that integration was dependent on the use of specific biological, social or geochemical indicators depicting the functions of complex systems. Referring to the complex relationship between disease emergence and changes in climate and ecosystems, Epstein (1994) proposed a number of principles for modeling and monitoring complex ecosystems. He emphasized the need to account not only for direct impacts to the different systems but also those indirect effects resulting from the interactions among factors within the three overlapping systems. He noted that those diseases transmitted directly from person to person reflect changes in population density with little interaction among the three systems, while vector-borne disease reflects environmental changes involving all

three systems in his integrated model. Integration in Epstein's approach also occurs through scientific and political collaborations. While he did not present an overall assessment of risk, he did suggest guidelines for identifying system vulnerabilities affecting overall stability and resilience; key elements in his view for mitigating disease emergence.

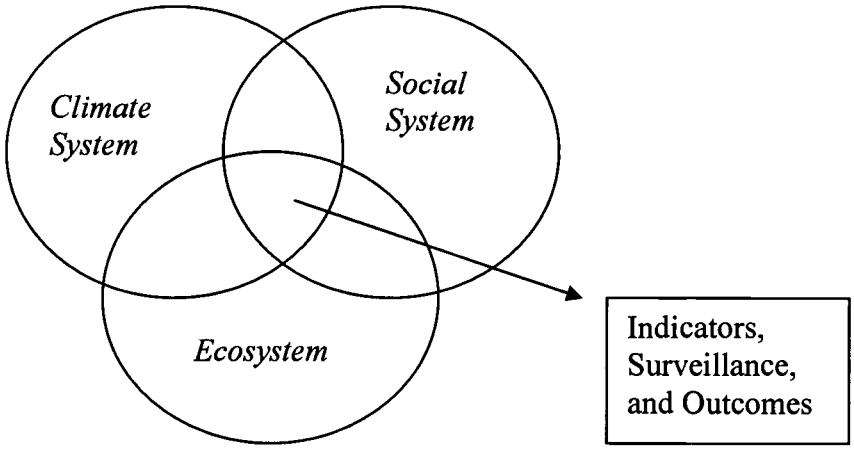


Figure 4.4. Framework for integrated assessment of climate, social systems and ecosystems. Integrated assessment occurs in the area of overlap and involves indicators for use in surveillance systems and predicting outcomes. Adapted from Epstein, 1994.

VanLeewen et al. (1999) also presented a conceptual 'butterfly' model that focused on human health in an ecosystem context. Human health is determined from the intersection of biophysical and socioeconomic environments. Biochemical and behavioral filters separate humans from each of these environments to protect against disease. The boundaries of the butterfly could be at the community, watershed, or population level and include the interactions between humans and the nonhuman environment. Their model is not an approach for assessing risk *per se* but can be viewed as a mechanism for determining risk factors influencing human health. Health is influenced by the structural or functional elements of an ecosystem (VanLeeuwen et al., 1999). As the authors noted, this model focuses

only on human health and does not determine health for other species in the ecosystem.

Integrative Analytical Approaches to Risk Assessment

The integrated paradigms described above provide a framework for considering human and environmental interactions but fall short of demonstrating specific analytical techniques for conducting an integrated risk analysis. The examples include a mix of conceptual, integrated approaches that are either descriptive or consist of parallel risk assessments. Considering the models presented by Epstein (1994) and VanLeeuwen et al. (1999) it is clear that an evaluation of interactions among human populations, their environment, and other important ecological factors are needed in conducting an integrated analysis. This type of evaluation is similar to that encompassed by an ecoepidemiological approach. Similar to human epidemiology, ecoepidemiology has been used to study the ecological effects that are prevalent in certain areas among population groups, communities and ecosystems and their potential causes (Bro-Rasmussen and Løkke, 1984; Martens, 1998). This approach focuses on a description of the effects, identification of causes, and understanding their linkages. Humans are considered as part of the environment.

Bro-Rasmussen and Løkke (1984) used the ecoepidemiological approach to describe possible associations between lesions observed in fish, discharge of high carbohydrate wastewaters, and discharges of chlorophenoxy and phenoxy acid herbicides in Køge Bay, Denmark. They determined that herbicide exposure alone was not sufficient to explain lesions observed in fish. Lesions were highest in fish found in coastal waters contaminated with high levels of organic materials and chemical pollutants from pulp and paper industries. This resulted in low oxygenated waters promoting the growth of bacterial flora. Fish in these waters had compromised immune systems that increased their vulnerability to facultative pathogens. Only when considering the complexities of the system and exposure to

multiple stressors were the investigators able to understand the dynamics and possible etiology for the observed effects.

An ecoepidemiological approach is similar to community and systems-level ecological risk assessment with respect to understanding relationships between biotic and abiotic factors. Levins (1973) noted that addressing more complex systems required breaking down disciplinary boundaries to create an integrated process that addresses management goals in which community structure, and other mechanistic factors could be examined as a whole. A system in this context is defined as a habitat, geographic area, human community or network of communities (Levins, 1998). As complexity increases, the ability to gather quantitative information is complicated by the impracticality of the number of parameters to measure and the loss of realism (Levins, 1966; Puccia and Levins, 1991).

Qualitative models can simplify complex systems without sacrificing realism (MacArthur and Levins, 1965; Levins, 1966) and enable an integrated analysis of a system. Qualitative modeling in the form of signed digraphs, 'loop analysis,' and matrix analysis facilitates the understanding of a system where there is incomplete information (Figure 4.5). Because qualitative models involve only the signs of the interactions among variables, (positive, negative, or no change), variables representing poorly quantified aspects of the system can be included in the analysis (Puccia and Levins, 1991). Such variables may represent different species, resources, climate, or socioeconomic factors that influence community structure and function. When constructing models, qualitative modeling methods can help determine which variables should be included, what should be measured, and how system dynamics might be affected under different perturbation scenarios (Levins, 1998).

Loop analysis and the corresponding community matrix is a useful analytical tool for exploring and understanding the effects of natural and anthropogenic stress on a system. Dambacher et al. (1999) used this modeling procedure to characterize a predator-prey system involving the snowshoe hare and

arctic fox. This technique also proved useful in predicting the impact of species introductions into a community (Li et al., 1999; Castillo et al., 2000) and explaining complex transitions in community composition over time (Bodini, 1998; Ortiz and Wolff, 2002). Loiselle et al. (2000; 2002) used loop analysis to examine different economically-based management scenarios in a wetland ecosystem to identify management options and guide monitoring programs.

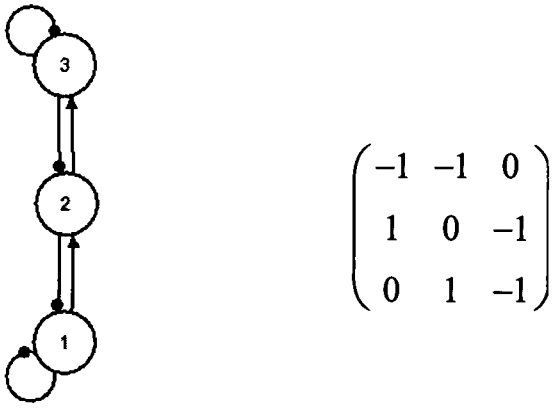


Figure 4.5. Signed digraph and corresponding community matrix. Circles 1, 2, and 3 represent variables. Lines represent linkages in the communities with arrows indicating positive effects, dark filled circles as negative effects. Half circles on the variable indicate self-regulating effects.

In the context of integrated risk, Levins (1998) extended qualitative modeling to the problem of vector-borne disease. In his system, he identified the invasiveness of vectors and disease reservoirs as core variables that would be important in an epidemic, adding vector habitat requirements, vector and host behaviour, host health status, and economic variables as other factors to be considered. With an increasing 'web of causation,' Levins (1998) argued that internal processes critical to community function could be examined. In Chapter 2, we developed a procedure to predict disease risk that combines recent developments in qualitative community modeling with biomathematical theory of vector-borne disease transmission. This procedure predicts the change in risk of

vector-borne disease following perturbations such as increases in vector abundance, animal control measures, habitat alteration, or global warming. Like Levin's postulated epidemic-disease community, this procedure allows the consideration of a complex community structure linking ecological factors to human disease. This procedure results in a rigorous prediction of an ecological community response to a perturbation with minimal to no quantitative parameterization. It generates focused hypotheses to guide data collection and control management strategies as interventions.

Bayesian analyses in the form of Bayesian networks are another tool that can be useful in an integrated risk analysis. A Bayesian approach is based on probability theory and is a useful decision-making or inferential technique when there is incomplete information or it is not possible to gather enough information to reduce uncertainties (Reckhow, 2003). A Bayesian network is used to model a system containing uncertainty, offering both qualitative and quantitative information in the form of conditional probabilities. It can be applied to multivariate problems involving complex relationships among variables (Reckhow, 2003). A Bayesian network consists of a directed acyclic graph and a probability distribution. The network characterizes variable relationships through interrelated nodes and arcs (Figure 4.5). The nodes represent variables and the arcs represent conditional dependencies between the nodes. Bayesian networks are used to identify those key variables influencing relationships within a system, and thus are an integrative analytical tool.

The use of Bayesian networks is increasing in scientific analyses of complex problems. Crome et al. (1996) applied a Bayesian approach to evaluate the impact of logging on bird and mammal species in rain forests. The investigators had too few data to detect potential impacts using traditional statistical analysis. However, results of a Bayesian analysis suggested a correlation between canopy cover and impacted bird species that was not previously apparent. Further, of the 76 species of birds in question, only 4 species were identified as having a high probability of being adversely impacted by logging.

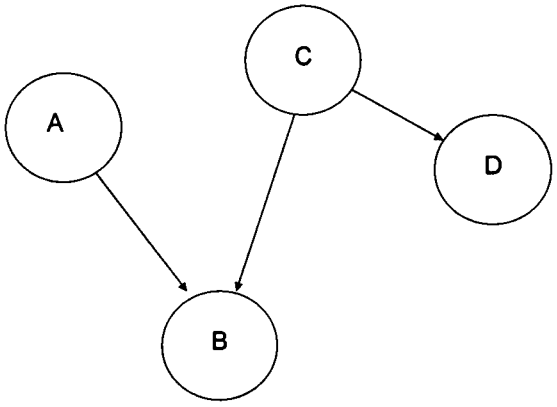


Figure 4.6. Bayesian network where the probability of "B" is dependent on "A" and "C." The probability of variable "D" is dependent on "C." The arcs indicate correlations between variables, which are not necessarily causal.

Bayesian networks have also been used to guide such diverse analyses as land management decisions (Marcot et al. 2001), fish stock assessment (Varis et al., 1993; Hammond and Ellis, 2002), and potential risk factors associated with heart disease (Buntine, 1991). Each of these cases started with a hypothesized model that could be updated as additional information became available, and involved large uncertainties, the pooling of information from different datasets, and expert judgment in the analysis.

When a specific model is not known, a data discovery technique, Relational Bayesian Modeling (RBM), can perform a heuristic search to discover models from data (Jorgensen et al., 2003). This technique involves machine learning, guided by expert judgment to develop a probabilistic model. The RBM extends Bayesian networks to the relational level, modeling uncertainty related to variables, their properties, and relationships among variables (Getoor et al., 2001). These relational data may be obtained from multiple sources, such as observational data stored in independent data tables that are related in space and time. The probabilistic relationship between variables is such that the distribution of any one variable in any one table, or any variable derived from any such variable, is

affected by, or probabilistically conditioned on, all other linked variables in the model. More importantly, variables that are not linked are conditionally independent of one another. Thus, RBMs are well suited for application to complex systems.

There are a few examples of where RBM has been used to evaluate complex problems. Getoor et al. (2001) described an RBM analysis to determine possible probabilistic relationships between patients from a tuberculosis clinic, certain risk factors, and specific strains of tuberculosis. In a second example, Jorgensen et al. (2003) used an RBM approach to explore the long-term changes in the clarity of Crater Lake using information summarized in multiple databases. The RBM analysis enabled the investigators to construct multiple, complex hypotheses concerning the entire lake ecosystem given data obtained from the long-term studies of the lake.

In Chapter 3, we used Relational Bayesian modeling to identify probabilistic relationships associated with the transmission of West Nile Virus in Maryland. Similar to the Crater Lake study (Jorgensen et al., 2003), the RBM approach was used to explore relationships among multiple, independent databases. Multiple hypotheses were generated suggesting spatial and temporal relationships between key vector, host and habitat variables related to disease transmission.

Thus, the RBM technique appears to be an effective means of conducting an integrated risk analysis through the qualitative and quantitative evaluation of complex community interactions. The hypotheses generated by the RBM analysis can be used to guide further quantitative testing of specific relationships between probabilistically linked variables.

Discussion

In our view, an integrated risk assessment should go beyond parallel comparisons of risk for species with common exposures and toxicities to chemical or other types of stressors. It is important to consider the interactions between

biotic and abiotic components of a system that influence human and environmental health. An integrated risk assessment should involve appropriate techniques that facilitate integration of risk analysis, identifying those components of the system that contribute most to risk.

To conduct an integrated risk assessment, a suite of tools is needed that integrates human and environmental health in the problem formulation (for hypothesis generation) and analysis phases of the assessment, not simply during the risk characterization phase. Such tools should consider the interacting system as a whole. Although this adds complexity in the analysis, models and other decision support methods are available that can simplify and reduce complexity.

Ideally, models should strive to characterize natural systems, optimizing generality, realism, and precision (Levins, 1966). However, such models would consist of too many parameters to measure, be difficult to solve, or if solved, the results would have little meaning. Considering that models can reflect only two of the three areas at best, Levins (1966) favored qualitative models that are flexible and emphasize realism and generality over precision. He argued that while quantitative models are useful in testing hypotheses, understanding qualitative relationships is most important in the long-term in understanding the system (Levins, 1966).

The 'integrative' models we reviewed are not robust enough to integrate multiple stressors or multiple endpoints, but use either parallel assessments or deductive reasoning to remove stressors from consideration. The analytical techniques employed in these models to characterize risk are applied to either human health or ecological assessments. Qualitative modeling and Bayesian methods provide an integrated risk analysis framework that identifies relationships important in the system and thus, guide the application of quantitative models or provide sufficient information for management decisions. Experimental comparison of various community theories suggests that loop analysis was the theoretical approach best suited for predicting the behaviour of complex community structures following a perturbation (Hulot et al., 2000). Both techniques rely on community structure to aid in formulating the problem, identifying limits,

and for generating hypotheses and testing predictions. Used in conjunction with mechanistic models, the integrated analytical techniques provide a balanced, iterative approach for not only assessing risk, but also evaluating possible consequences of different management decision scenarios.

CHAPTER 5
CONCLUSIONS

The environmental statutes from the 1970s and 1980s, called for risk assessments that were human centric and cancer phobic in nature. These assessments focused on environmental research and policy to reduce emissions or prevent contamination from potentially cancer-causing chemicals in the environment. During the 1990s, revisions to these environmental statutes placed greater scrutiny on the scientific analyses supporting environmental regulation through an increased emphasis on costs and benefits associated with risk management decisions. As our society strives for innovation, new technology and a sound economy, there is a greater recognition of the interdependence between environmental and human health and thus, the need to evaluate risk in an integrated fashion. However, several integrated risk models do not employ analytical techniques that integrate multiple stressors or multiple endpoints. Instead, integrated risks are represented as parallel assessments or involve deductive reasoning to identify key stressors and endpoints.

As noted in Chapter 4, the conceptualization of a holistic system should include abiotic factors as well as traditional endpoints of human health and ecological integrity. To consider these factors in risk analysis adds complexity that is beyond the capabilities of some commonly used analytical approaches aimed at estimating the likelihood and severity of risk. Thus, to conduct an integrated risk assessment, a suite of tools is needed that is responsive to the conceptualized problem and considers the interacting system as a whole, addressing all appropriate biotic and abiotic components of a system.

In this dissertation I demonstrate the use of both qualitative and quantitative community-level modeling techniques for integrating risk analysis associated with vector-borne disease. Vector-borne disease provides an example of a risk common to both humans and wildlife. The emergence of vector-borne disease has increased over the past several decades due, in part, to increased human activity that disrupts

the natural environment (NAS, 1992; Daszak et al, 2000). As a result of this increase, the NRC emphasized the need to improve our understanding of the interactions among pathogens, hosts/receptors, and the environment (NRC, 2000). Thus, to evaluate disease, in particular zoonoses, an integrated approach could account for the ecology of the disease agent, vector, host, and other abiotic factors such as climate, economic conditions, or control strategies that can affect the distribution, frequency or severity of disease (Wilson, 1994).

In Chapter 2, I develop a new qualitative modeling procedure based on community-level interactions that predicts a change in risk from vector-borne disease following a disturbance to the community structure. This procedure combines the parameters from a traditional, quantitative biomathematical model, vectorial capacity, with qualitative community interactions as determined from a community matrix to qualitatively predict changes in risk without quantitative parameterization. This procedure predicts the change in risk of vector-borne disease from press perturbations, such as control measures, anthropogenic habitat alteration or global warming. I demonstrated this procedure using a documented example of an ecological community associated with Lyme disease, a tick-borne disease affecting humans. This new modeling procedure predicted that a positive press to deer, for example, increased tick abundance, increasing disease risk. This prediction is consistent with observations in the literature (Wilson et al., 1983). Further, the model results also indicated that a positive press to acorns increases the abundance of mice, other small mammalian hosts, and ticks. However, it does not necessarily increase the risk of Lyme disease, as suggested in the literature (Ostfeld et al., 1997). A benefit of this procedure is the ability to generate focused hypotheses to guide quantitative models and evaluate potential intervention strategies.

As described in Chapter 3, I explored the application of another community-level technique Relational Bayesian Modeling (RBM), to develop hypotheses associated with the transmission of another vector-borne disease, West Nile Virus (WNV) that affects both humans and wildlife. The uncertainties concerning the ecology of WNV suggest the application of a Bayesian analysis.

Using an RBM I was able to model patterns of WNV transmission across space and time in Maryland from sparse, independently collected observational data. The RBM is a model discovery technique that uses observational data to construct quantitative and biologically-consistent models in the form of Bayesian networks. This probabilistic modeling technique provides both qualitative and quantitative information enabling investigators to conduct a continuum of analyses in which information can be added, deleted, or modified to generate a new model, without starting anew. Issues of uncertainty can be accounted for using an analysis of probability distributions. In the analysis of WNV in Maryland during 2001, I found that there was a spatial and temporal pattern of a probabilistic association between the number of tire license sites, infected birds, infected mosquitoes and humans within the same or neighboring geographic locations and in the same, previous, or following month in which a positive bird was found. Thus, even when there are uncertainties and limited data regarding disease transmission, this novel modeling technique may prove to be a valuable tool for evaluations of disease outbreak.

In Chapter 4, I explain how both the qualitative and RBM models show promise in integrating risk analysis. Relying on community structure, these analysis techniques provide a general, but realistic and practical approach for developing hypotheses concerning the interacting relationships of community members.

In summary, the iterative and heuristic nature of these models, improves our ability to predict the impacts of human or natural activity on complex ecosystems. Their use involves a multidisciplinary approach, improving their utility as an integrative tool to provide a realistic analysis of community interactions. In the context of an integrated environmental risk assessment, these models can frame the problem through the development of multiple and simultaneous hypotheses and generation of testable predictions. These hypotheses and predictions then guide the rest of the risk assessment in the analysis, and risk characterization phases through improved conceptualization of the risk problem and relationships among stressors and receptors, and evaluation of different management actions.

As a benefit to public health, these modeling tools provide a new and different analytical approach to public health to evaluate zoonotic disease at a

community level. Reasonable and rigorous predictions of disease risk can be developed using the qualitative community model. The RBM model can use observational data of varying quality to identify, probabilistically, those variables that are important in a disease outbreak. These variables could then serve as indicators to guide disease surveillance or control strategies. In addition, the RBM can be used to determine the influence of bias or confounding in the data. Using integrative tools such as these, will lead to more informed environmental and public health decisions.

BIBLIOGRAPHY

Anderson, R.M. and R.M. May. 1979. Population biology of infections: Part 1. *Nature* 280:361-367.

Bailey, N.T. 1982. *The Biomathematics of Malaria.* Charles Griffin & Co. LTD. London 210pp.

Bender, E.A., T.J. Case, and M.E. Gilpin.1984. Perturbation experiments in community ecology: Theory and practice. *Ecology* 65:1-13.

Bodini, A. 1998. Representing ecosystem structure through signed digraphs; model reconstruction, qualitative predictions and management: The case of freshwater ecosystem. *Oikos* 83:93-106.

Bro-Rasmussen, F. and H. Løkke, 1984. Ecoepidemiology – A casuistic discipline describing ecological disturbances and damages in relation to their specific causes: Exemplified by chlorinated phenols and chlorophenoxy acids. *Regulatory Tox. and Pharm.* 4:391-399.

Buntine. W. 1991. Theory refinement on Bayesian networks. *In: Proceedings of the Seventh Conference of Uncertainty in Artificial Intelligence.* M. Kaufmann (Ed.) San Francisco. pp. 52-60.

Castillo, G.C., H.W. Li, and P.A. Rossignol. 2000. Absence of overall feedback in a benthic estuarine community: A system potentially buffered from impacts of biological invasions. *Estuaries* 23:275-291.

Centers for Disease Control and Prevention (CDC). 1994. Addressing emerging infectious disease threats: A prevention strategy for the United States. US Department of Health and Human Services, Public Health Service, Atlanta, GA. *Morb. and Mort. Weekly Report* 43:1-18.

Centers for Disease Control and Prevention (CDC). 1999. Update-West Nile-like viral encephalitis-New York. *Morb. and Mort. Weekly Report* 48:890-892.

Centers for Disease Control and Prevention (CDC). 2002a. West Nile Virus activity– United States, November 7-13, 2002. *Morb. and Mort. Weekly Report* 50:1026-1027.

Centers for Disease Control and Prevention (CDC). 2002b. Update: Investigations of West Nile Virus infections in recipients of organ transplantation and blood transfusion—Michigan, 2002. *Morb. and Mort. Weekly Report* 51:879.

Centers for Disease Control and Prevention (CDC). 2002c. Possible West Nile Virus transmission to an infant through breast-feeding--Michigan, 2002. *Morb. and Mort. Weekly Report* 51:877-878.

Centers for Disease Control and Prevention (CDC). 2002d. Intrauterine West Nile Virus infection – New York, 2002. *Morb. and Mort. Weekly Report* 51:1135-1136.

Cornel, A.J., P.G. Jupp, and N.K .Blackburn. 1993. Environmental temperature on the vector competence of *Culex univittatus* (Diptera: Culicidiae) for West Nile Virus. *J. Med. Entomol.* 20:449-456.

Crome, F.H.J., M.R. Thomas, and L.A. Moore. 1996. A novel Bayesian approach to assess the impacts of rain forest logging. *Ecol. Appl.* 6:1104-1123.

Dambacher, J.M., H.W. Li, J.O. Wolff, and P.A. Rossignol. 1999. Parsimonious interpretation of the impact of vegetation, food, and predation on the snowshoe hare. *Oikos* 84: 530-532.

Dambacher, J.M., H.W. Li, and P.A. Rossignol. 2002. Relevance of community structure in assessing indeterminacy of ecological predictions. *Ecology* 83:1372-1385.

Dambacher, J.M., H.W. Li, and P.A. Rossignol. 2003a. Qualitative predictions in model ecosystems. *Ecol. Model.* in press.

Dambacher, J.M., H-K. Luh, H.W. Li, and P.A. Rossignol. 2003b. Qualitative stability and ambiguity in model systems. *Am. Nat.* in press.

Dambacher, J.M., R. Levins, and P.A. Rossignol. Submitted. Age structure in perturbed communities: Derivation and appraisal. *Math. Biosci.*

Daszak, P., A.A. Cunningham, and A.D. Hyatt. 2000. Emerging infectious diseases of wildlife—Threats to biodiversity and human health. *Science* 287:443-449.

Dobson, A. and R.M. May. 1986. Chapter 16. Disease and Conservation. *In: Conservation Biology.* M Soule (Ed.) Sinauer Press. Sunderland Mass. pp 345-365.

Dobson, A. and P. Hudson. 1994. The interaction between the parasites and predators of Red Grouse *Lagopus logopus scoticus*. *Ibis* 137:S87-S96.

Dobson, A. and J. Foufopoulos. 2001. Emerging infectious pathogens of wildlife. *Phil. Trans. R. Soc.* London 356:1001-1012.

DiGiulio, R. and W. Benson. 2002. Chapter 1: Rationale and Overview. *In: Interconnections Between Human Health and Ecological Integrity.* R.T. DiGiulio and W.H. Benson (Eds). SETAC Press. Pensacola, Fla. pp.1-13.

Enserink, M. 2000. The enigma of the West Nile. *Science* 290:1482-1484.

Epstein, P.R. 1994. Framework for an integrated assessment of climate change and ecosystem vulnerability. *In: Disease in Evolution: Global Changes and Emergence of Infectious Diseases.* ME Wilson, R Levins, A Spielman (Eds). New York Academy of Science, New York. pp: 423-435.

Forget, G. and J. Lebel. 2001. An ecosystem approach to human health. *Internat'l J. Occup. and Env. Health.* Supplement to 7:3-36.

Friend, M., R.C. McLean, and F.J. Dein. 2001. Disease emergence in birds: Challenges for the twenty-first century. *The Auk* 118:290-303.

Garrett-Jones, C. 1964. Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. *Nature* 204:1173-1175.

Getoor, L., N. Friedman, D. Koller, and A. Pfeffer. 2001. Chapter 1. Learning probabilistic relational models. *In: Relational Data Mining.* S Dzeroski and N Laurec (Eds.) Springer Verlag Publ. New York. pp 7-35.

Goodwin, Jr. M.H. 1958. Needs for ecologic work in control of communicable diseases. *Ecology* 39: 162-165.

Hammond, T.R. and J.R. Ellis. 2002. A meta-assessment for elasmobranches based on dietary data and Bayesian networks. *Ecol. Indicators* 1:197-211.

Hansen, L., S.F. Hedtke, and W.R. Munns, Jr. 2003. Integrated human and ecological risk assessment: A case study of ultraviolet radiation effects on amphibians, coral, humans, and oceanic primary productivity. *Human and Ecol. Risk Assess.* 9: 359-378.

Harvey, T., K.R. Mahaffey, S. Velazquez, and M. Dourson. 1995. Holistic risk assessment: An emerging process for environmental decisions. *Reg. Tox. and Pharm.* 22:110-117.

Hess, G. 1996. Disease in metapopulation models: Implications for conservation. *Ecology* 77:1617-1632.

Hethcote, H.W. 1976. Qualitative analysis of communicable disease models. *Math. Biosci.* 28:335-356.

Hudson, P.J., A.P. Dobson, and D. Newborn. 1998. Prevention of population cycles by parasite removal. *Science* 282: 2256-2258.

Hulot, F.D., G. Lacroix, F. Lescher-Moutoue, and M. Loreau. 2000. Functional diversity governs ecosystem response to nutrient enrichment. *Nature* 405:340-344.

Jensen, F.V. 2001. *Bayesian Networks and Decision Graphs*. Springer-Verlag Publ. New York. 268 pp.

Jorgensen, J., B. D'Ambrosio, and P.A. Rossignol. 2003. Data-driven construction of community models of Crater Lake. *NSF Biocomplexity Workshop -The vertical organization of energy, carbon, and nutrient cycles in an ultraoligotrophic ecosystem: A workshop on Crater Lake, Oregon.* February 16 -18.

Kermack, W.O. and A.G. McKendrick. 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London. Ser A 115:700-721. As cited in Onstad D.W. and R.I. Carruthers. 1990. Epizooltiological models of insect diseases. *Ann Rev. Entomol.* 35:399-419.

Klowden, M. 1995. Blood, sex and the mosquito. *BioScience* 5:326-331

Komar, N. 2000. West Nile Viral encephalitis. *Rev. Sci. Tech. Off Int. Epiz.* 19:166-176.

Kutz, F.R., T.G. Wade, and B.P. Pagac. 2003. A geospatial study of the potential of two exotic species of mosquitoes to impact the epidemiology of West Nile Virus in Maryland. *J. Am. Mosquito Control Assoc.* In press.

Lackey, R. 1997. If ecological risk assessment is the answer, what is the question? *Human and Ecol. Risk Assess.* 3:921-928.

Lane, R.S., J. Piesman, and W. Burgdorfer. 1991. Lyme Borreliosis: Relation of its causative agent to its vectors and hosts in North America and Europe. *Ann. Rev. Entomol.* 36:587-609.

Levins, R. 1966. The strategy of model building in population biology. *Am. Nat.* 54:421-431.

Levins, R. 1968. *Evolution in Changing Environments: Some Theoretical Explorations. In: Monographs in Population Biology.* R. MacArthur (Ed). Princeton University Press, Princeton New Jersey. pp. 3-120.

Levins, R. 1973. Fundamental and applied research in agriculture. *Science* 181:523-524.

Levins, R. 1975. Evolution in communities near equilibrium. *In: Ecology and Evolution of Communities.* M.L. Cody and J. Diamond (Eds.). Harvard University Press. Cambridge Mass. pp. 16-50.

Levins, R. 1998. Chapter 11. Qualitative mathematics for understanding, prediction, and intervention in complex ecosystems. *In: Approaches to Assessing Ecosystem Health.* D. Rapport, R. Costanza, P. Epstein, C. Gaudet, and R. Levins. (Eds). Blackwell Mass. pp 178-204.

Levins, R., T. Awerbuch, U. Brinkmann, I. Eckardt, P. Epstein, N. Makhoul, C.A. DePossas, C. Puccia, A. Spielman, and M.E. Wilson. 1994. The emergence of new diseases – Lessons learned from the emergence of new diseases and the resurgence of old ones may help us prepare for future epidemics. *Am Scientist* 82:52-60.

Li, H.W., P.A. Rossignol, and G. Castillo. 1999. Chapter 30: Risk analysis of species introductions: Insights from qualitative modeling. *In: Nonindigenous Freshwater Organisms. Vectors, Biology, and Impacts.* R. Claudi and J.H. Leach (Eds). Lewis Publ. Boca Raton, Fla. pp 431-447.

LoGuidice, K., R.S. Ostfeld, K.A. Schmidt, and F. Keesing. 2003. The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. *Proc. Nat. Acad. Sci.* 100:567-571.

Loiselle, S., G.M. Carpenato, V. Hull, T. Waller, and C. Rossi. 2000. Feedback analysis in reserve management: studying local myths using qualitative models. *Ecol. Model.* 129:25-37.

Loiselle, S., V. Hull, E. Permengeat, M. Falcucci, and C. Rossi. 2002. Qualitative models to predict impacts of human interventions in a wetland ecosystem. *Web Ecology* 3:56-69.

Lotka, A.J. 1923. Contribution to the analysis of malaria epidemiology. *Am. J. Hygiene* 3:1-121.

Lotz, J.M., A.O. Bush, and W.F. Font. 1995. Recruitment-driven, spatially discontinuous communities: A null model for transferred patterns in target communities of intestinal helminthes. *J. Parasitol.* 81: 12-24.

Lundstrom, J. 1999. Mosquito-borne viruses in Western Europe: A review. *J. Vector Ecol.* 24:1-39.

MacArthur, R.H. and R. Levins. 1965. Competition, habitat selection, and character displacement in a patchy environment. *Proc. Nat. Acad. Sci.* 51:1207-1210.

Macdonald, G. 1952. The analysis of equilibrium in malaria. *Trop. Disease Bull.* 49: 813-829.

Macdonald, D. 1980. *Rabies and Wildlife*. Oxford University Press. Oxford. pp 151.

Marcot, B.G., R.S. Holthausen, M.G. Raphael, M.M. Rowland, and M.J. Wisdom. 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *For. Ecol. and Manage.* 153:29-42.

Martens, P. 1998. *Health and Climate Change*. Earthscan Publs. Ltd. London pp. 15-26.

Mather T.N., M.L. Wilson, S.I. Moore, J.M. Ribeiro, and A. Spielman. 1989. Comparing the relative potential of rodents as reservoirs of the Lyme disease spirochete (*Borrelia burgdorferi*). *Am. J. Epidem.* 130:143-150.

May, R.M. 1993. Chapter 6: Ecology and evolution of host-virus associations. *In: Emerging Viruses.* S Morse (Ed.) Oxford University Press. Oxford. pp 58-68.

Miller, T.E., J.M. Kneitel, and J.H. Burns. 2002. Effect of community structure on invasion success and rate. *Ecology* 83:898-905.

Miranda, M.L., P. Mohai, J. Bus, G. Charnley, E.J. Dorward-King, P. Foster, W.R. Munns. 2002. Chapter 2: Policy Concepts and Applications. *In: Interconnections Between Human Health and Ecological Integrity*. R.T. DiGiulio and W.H. Benson (Eds.) SETAC Press. Pensacola, Fla. pp.15-40.

National Research Council (NRC), National Academy of Sciences. 1983. *Risk Assessment in the Federal Government.* National Academy Press. Washington D.C. 189 pp.

National Research Council (NRC), National Academy of Sciences Institute of Medicine. 1992. J. Lederberg, R.E. Shope, and S.C. Oaks Jr. (Eds). Committee on emerging microbial threats to health. National Academy Press. Washington D.C. pp. 34-112.

National Research Council (NRC), National Academy of Sciences. 1996. *Understanding Risk: Informing Decisions in a Democratic Society*. National Academy Press, Washington D.C. 249 pp.

National Research Council (NRC). 2000. *Grand Challenges in Environmental*

Sciences. National Academy Press. Washington DC. pp. 36-48.

Nicholson, M.C. and T.N. Mather. 1996. Methods for evaluating Lyme disease using geographic information systems and geospatial analysis. *J. Med. Entomol.* 33: 711-720.

Ortiz, M. and M. Wolff. 2002. Application of loop analysis to benthic systems in northern Chile for the elaboration of sustainable management strategies. *Marine Ecol. Prog. Ser.* 242:15-27.

Ostfeld, R.S., C.G. Jones, and J.O. Wolff. 1996. Of mice and men. *BioScience* 46:323-330.

Ostfeld, RS. 1997. The ecology of Lyme-Disease risk. *Am. Sci.* 85:338-346.

Ostfeld, R.S. and F. Keesing. 2000. Biodiversity and disease risk: the case of Lyme disease. *Cons. Biol.*14: 722-728.

Patz, J.A., P.R. Epstein, T.A. Burke, and J.M. Balbus. 1996. Global climate change and emerging infectious disease. *J. Am. Med. Assoc.* 275:217-223.

Pavlovsky, E.N. 1966. *Natural Nidality of Transmissible Disease*. English translation by F.K. Plous, Jr. N.D. Levine (Ed). University of Illinois Press. Urbana. pp. 1-36.

Peterman, R.M. and J. Anderson. 1999. Decision analysis: A method for taking uncertainties into account in risk-based decision making. *Human and Ecol. Risk Assess.* 5:231-244.

Plowright, W. 1982. The effects of rinderpest and rinderpest control on wildlife in Africa. *Symp. Zool. Soc.* London. No. 50. pp. 1-28.

Post, W.M., D.L. DeAngelis, and C.C. Travis. 1983. Endemic disease in environments with spatially distributed heterogeneous host populations. *Math. Biosci.* 63:289-302.

Puccia, C.J. and R. Levins. 1985. *Qualitative Modeling of Complex Systems*. Harvard University Press. Cambridge, Mass. 259 pp.

Puccia, C.J. and R. Levins. 1991. Chapter 6. Qualitative modeling in ecology: Loop analysis, signed digraphs, and time averaging. *In: Qualitative Simulation Modeling and Analysis*. P.A. Fishwick and P.A. Luker (Eds.) Springer-Verlag Publ. New York. Pp. 119-143.

Ramoni, M. and P. Sebastiani. 2001. Analysis of survey data with Bayesian networks. *In: Bayesian Methods: Selected Papers from the Sixth Annual*

Meeting of the International Society for Bayesian Analysis. European Community. Luxembourg. pp 429-438.

Rappole, J.H., S.R. Derrickson, and Z. Hubalek. 2000. Migratory birds and spread of West Nile Virus in the western hemisphere. *Emerging Infectious Diseases* 6:1-15.

Real, L.A. 1996. Sustainability and the ecology of infectious disease. *BioScience* 46: 88-97.

Reckhow, K. H. 2003. Bayesian approaches in ecological analysis and modeling. IN:*The Role of Models in Ecosystem Science.* C.D. Canham, J.J. Cole, and W.K. Lauenroth (Eds). Princeton Univ Press. Princeton, New Jersey. In press

Ross, R. 1908. *Report on the Prevention of Malaria in Mauritius.* Waterlow and Sons Ltd. London. pp. 30-104

Ross, R. 1910. *The Prevention of Malaria.* EP Dutton & Co. NY. pp. 151-253.

Ross, P. and L. Birnbaum. 2003. Integrated human and ecological risk assessment: A case study of persistent organic pollutant (POPs) risk to humans and wildlife. *Human and Ecol. Risk Assess.* 9: 303-324.

Roundy, RW. 1978. A model for combining human behavior and disease ecology to assess disease hazard in a community; Rural Ethiopia as a model. *Soc. Sci. and Med.* 12: 121-130.

Sattenspiel, L. 1990. Modeling the spread of infectious disease in human populations. *Yrbk. Phys. Anthropol.* 33:245-276.

Sekizawa, J., G.W. Suter, and L. Birnbaum. 2003. Integrated human and ecological risk assessment: A case study of tributyltin and triphenyltin compounds. *Human and Ecol. Risk Assess.* 9: 325-342.

Suter, G., T. Vermeire, W. Munns, and J. Sekizawa. 2003. A framework for the integration of health and ecological risk assessment. International Programme for Chemical Safety (IPCS) Workshop summary. *Human and Ecol. Risk Assess.* 9: 281-302.

Tompkins, D.M. and M. Begon. 1999. Parasites can regulate wildlife populations. *Parasitol. Today.* 15:311-313.

U.S. Environmental Protection Agency (USEPA). 1992. Framework for Ecological Risk Assessment. Office of Research and Development. EPA/620/R-92/001. February.

U.S. Environmental Protection Agency (USEPA). 1995. Guidance for Risk Characterization. Office of Science Policy. February.

U.S. Environmental Protection Agency (USEPA). 1998. Guidelines for Ecological Risk Assessment. Office of Research and Development. *Federal Register* 63:26846-26924.

U.S. Geological Survey (USGS). 2000. West Nile Virus. National Wildlife Health Center. found at www.umesc.usgs.gov.

VanLeeuwen, J.A., D. Waltner-Toews, T. Abernathy, and B. Smitt. 1999. Evolving models of human health toward an ecosystem context. *Ecosystem Health* 5:204-219.

Varis, O., S. Kuikka, and J. Kettunen. 1993. Belief networks in fish stock assessment – The Baltic salmon case. *International Council for the Exploration of the Sea (ICES) Statutory Meeting. Ref M. Statistics Committee. Ref. Anacat Committee.* pp 1-18.

Vermeek N.A.M. and C. Caffrey. 2002. American Crow. *The Birds of North America* No. 647. 36 pp.

Vermeire, T., R. MacPhail, and M. Waters. 2003. Integrated human and ecological risk assessment: A case study of organophosphorous pesticides in the environment. *Human and Ecol. Risk Assess.* 9: 343-358.

Vitousek, P.M., H.A. Mooney, J. Lubchenco, and L.M. Mellilo. 1997. Human domination of Earth's ecosystems. *Science* 277:494-499.

Wilson, E.O. 1998. Integrated science and the coming century of the environment. *Science* 279:2048-2049.

Wilson, M.L., J.F. Levine, and A. Spielman. 1984. Effect of deer reduction on abundance of deer tick. *The Journal.* The Yale Journal of biology and medicine. 57:697-705.

Wilson, M. 1994. *Disease in Evolution.* New York Academy of Sciences. pp.1-12.

Wilson, M. 1995. Travel and the emergence of infectious diseases. *Emerging Infectious Disease.* 1:1-12.

World Health Organization (WHO). 2001. International Programme on Chemical Safety. Report in Integrated Risk Assessment. WHO/IPCS/IRA/01/12. Geneva, Switzerland.