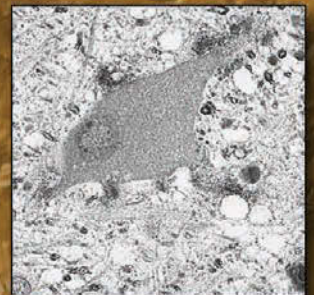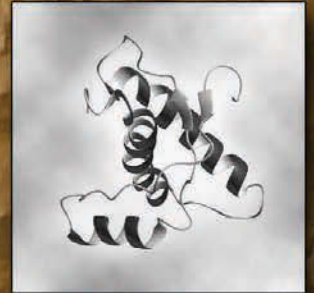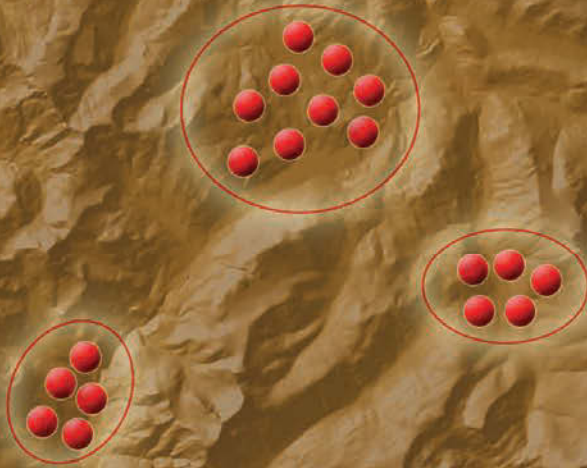# Scale-dependent approaches to modeling spatial epidemiology of chronic wasting disease

$$P_i(x_i, t_0 + s_i) = \int_{-\infty}^{+\infty} G_i(x_i - x_0, s_i)P(x_0, t_0)dx_0$$

$$= \int_{-\infty}^{+\infty} \frac{\exp\left[\frac{-(x_i - x_0 - \beta_i s_i)^2}{4D_i s_i}\right]}{\sqrt{4\pi D_i s_i}} P(x_0, t_0)dx_0$$

Special Report 2007

AUTHORS:
- Mary M. Conner — Utah Division of Wildlife Resources
- John E. Gross — National Park Service
- Paul C. Cross — United States Geological Survey
- Michael R. Ebinger — Utah State University
- Robert R. Gillies — Utah State University
- Michael D. Samuel — United States Geological Survey
- Michael W. Miller — Colorado Division of Wildlife

UTAH DNR WILDLIFE

USGS science for a changing world

ASSOCIATION of FISH & WILDLIFE AGENCIES

Scale-dependent approaches to modeling spatial epidemiology of chronic wasting disease

Special Report 2007

Contributing Authors:

Mary M. Conner	Utah Division of Wildlife Resources
John E. Gross	National Park Service
Paul C. Cross	United States Geological Survey
Michael R. Ebinger	Utah State University
Robert R. Gillies	Utah State University
Michael D. Samuel	United States Geological Survey
Michael W. Miller	Colorado Division of Wildlife

Technical Associates:

Karen G. Wolfe	Utah State University
Samson Y. Gebreab	Utah State University

Executive Summary

This e-book is the product of a second workshop that was funded and promoted by the United States Geological Survey to enhance cooperation between states for the management of chronic wasting disease (CWD). The first workshop addressed issues surrounding the statistical design and collection of surveillance data for CWD. The second workshop, from which this document arose, followed logically from the first workshop and focused on appropriate methods for analysis, interpretation, and use of CWD surveillance and related epidemiology data. Consequently, the emphasis of this e-book is on modeling approaches to describe and gain insight of the spatial epidemiology of CWD.

We designed this e-book for wildlife managers and biologists who are responsible for the surveillance of CWD in their state or agency. We chose spatial methods that are popular or common in the spatial epidemiology literature and evaluated them for their relevance to modeling CWD. Our opinion of the usefulness and relevance of each method was based on the type of field data commonly collected as part of CWD surveillance programs and what we know about CWD biology, ecology, and epidemiology. Specifically, we expected the field data to consist primarily of the infection status of a harvested or culled sample along with its date of collection (not date of infection), location, and demographic status. We evaluated methods in light of the fact that CWD does not appear to spread rapidly through wild populations, relative to more highly contagious viruses, and can be spread directly from animal to animal or indirectly through environmental contamination. We discovered that many of the well-published methods were developed for fast-spreading human diseases, such as influenza and measles. While these methods are applicable to fast spreading wildlife diseases, such as foot-and-mouth disease or West Nile virus, many are not likely to work well for CWD. Only limited data exist to evaluate geographic and spatial spread because many locations where we find CWD tend to be locations where samples have just been taken or sample sizes have just become large enough to have a high probability of detecting a low prevalence. Consequently, methods that work well to describe or predict the spread of foot-and-mouth disease throughout England, which occurred within a year, do not work well for describing or predicting CWD spread. We did not exclude methods that we regarded as inappropriate; rather, we included methods that are commonly used for disease epidemiology and then discussed their applicability for modeling the spatial epidemiology of CWD. We hope including inappropriate methods with an explanation of why they are ill-suited for CWD will make it easier to drop them from consideration and explain to others why they were not recommended for spatial modeling of CWD.

We organized the three chapters by scale and extent for which each method was developed or best suited. The first chapter covers methods appropriate to multi-jurisdictional or multi-state modeling, which we call "regional" scale. The second chapter covers methods appropriate for within state areas such as wildlife management units or metapopulations, which we call "landscape" scale. The third chapter covers methods appropriate for population or individual-based modeling, which we call "fine" scale. We know this rubric is somewhat artificial because many methods work at multiple scales. We hope, however, that this structure addresses some of the challenges faced by managers that work at local, regional, state, and national scales. Further, the resolution of empirical data often changes with spatial scale, which affects the utility of different modeling approaches. For example, individual-based models work best at modeling spread within populations, while risk analysis is most useful for summarizing data over larger scales such as a region. Because some methods are applicable at several scales, however, we included a graphic at the beginning of each method that indicates the range of scales for

which it applies. For example, the graphic to the right indicates that the method is most applicable for regional-scale modeling.

There is also a question of resolution as well as scale and extent for each method. CWD surveillance data have been collected over large areas, such as a wildlife management unit or state, but the resolution of the data may be fine scale with GPS locations for many samples. For each method, we described the required resolution of the data and describe the type of data required, as well as what questions the method could answer and how useful the method is, given typical CWD data.

| Regional Scale | Landscape Scale | Fine Scale |
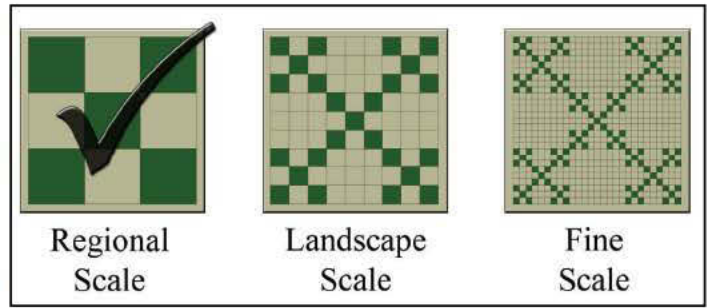
An example of the scale graphic which is found at the beginning of each method in the text. The most applicable scales are indicated by the presence of a check mark, which is the regional scale in this example.

For each scale, we presented a focal approach that would be useful for understanding the spatial pattern and epidemiology of CWD, as well as being a useful tool for CWD management. The focal approaches include risk analysis and micromaps for the regional scale, cluster analysis for the landscape scale, and individual based modeling for the fine scale of within population. For each of these methods, we used simulated data and walked through the method step by step to fully illustrate the "how to", with specifics about what is input and output, as well as what questions the method addresses. We also provided a summary table to, at a glance, describe the scale, questions that can be addressed, and general data required for each method described in this e-book. We hope that this review will be helpful to biologists and managers by increasing the utility of their surveillance data, and ultimately be useful for increasing our understanding of CWD and allowing wildlife biologists and managers to move beyond retroactive fire-fighting to proactive preventative action.

© Wildlife Reflections
Nature Photography

Summary Table of Methods

Model	Scale	Usefulness	CWD Usefulness	Data Required	Potential Output
Risk Analysis	Broad Scale (Regional/ Multi-State)	Non-infectious or slow epidemics	CWD Modeling and management	• Spatial coordinates • Sample date • Factor of interest co-variate data	• Estimation of spatial variation of disease prevalence • Statistics for factors affecting disease prevalence • Model statistics and variable weights
Micromaps	Broad Scale (Regional/ Multi-State)	Facilitates data visualization and analysis of discrete regions	Useful for presenting data from multiple regions	• Positive/negative samples by polygon • Polygon identification	• Linked time/space prevalence maps • Linked time/space statistics
Cluster Analysis	Broad Scale (Regional/ Multi-State)	Identification of "hot spots" with presence/ absence data	Descriptive tool for CWD surveillance	• Positive/negative samples by polygon	• Assignment of positives to clusters
Snapshot Approach	Broad Scale (Regional/ Multi-State) and/or Local/Fine-Scale	Infection spread is independent of temporary barriers	Assumptions about host distribution limits the utility for CWD modeling	• Area disease status area • Centroid coordinates of area • Modeling assumptions	• Identify likely site of future infection. • Estimates of rate of spread
Occupancy Analysis	Broad Scale (Regional/ Multi-State)	Monitor disease status over large spatial scales	First cut at multi-state/ regional scales	• Area disease status • 2-years of data for each area	• Probability of disease presence • Evaluation of covariate data importance to presence/detection
Epidemic Trees	Broad Scale to Fine Scale depending on the unit of analysis (individual or management unit)	Highly infectious, rapid spreading diseases	Inability to link positive cases to original source hinders CWD use	• Location of infections • Some case-tracking data • Dates of when infections began and were reported • Rule set tying subsequent cases to origin	• Rate of spread estimates through time/space • Generation time estimates • Reporting time estimates • Rate and routes of spread
Cluster Analysis	Landscape Level (Statewide)	Depict and describe spatial patterns of disease and identify hotspots of prevalence and estimate risk surface	Descriptive tool for CWD surveillance	• Spatial coordinates • Covariates (if necessary)	• Assignments of positives to cluster • Potential spatial covariates tied to disease pattern • Identification of high-risk areas and cluster maps

Summary Table of Methods

Model	Scale	Usefulness	CWD Usefulness	Data Required	Potential Output
Geostatistical Analysis	Landscape Level (Statewide)	Depict and describe purely spatial relationships (autocorrelations) between positive and negative disease samples	Better evaluation of environmental and ecological covariates after accounting for the spatial structure of the data	• Spatial coordinates of samples or centroid of small areas if data are grouped • Covariates (if necessary)	• Autocorrelation of disease related to distance and/or direction • Spatial prevalence estimates • Potential spatial covariates tied to disease pattern after accounting for spatial autocorrelation
Cellular Automata Models	Landscape Level (Statewide)	Understanding how factors affect spatial spread/distribution of a disease at various scales	General tool useful for various purposes, such as evaluating various locales as CWD origin point. Often assumes homogeneous environment and distribution of hosts.	• Application dependent	• Application dependent
Metapopulation Models	Landscape Level (Statewide)	Evaluation of prediction of management strategies in spatially structured populations	Accounts for the herd structure of deer and elk, but requires one to define the patches or subpopulations	• Dispersal/migration routes and probabilities among patches • Vital rates of population and transmission dynamics	• Spread of disease through a patchy distribution of hosts.
Diffusion Models	Landscape Level (Statewide)	Evaluations in diseases that do not present themselves or slow moving diseases	Hypothesis of CWD spread and potential origin point. Diffusion models often assume a uniform environment (e.g. no habitat heterogeneity).	• Date of first detection for spatial area • Proportion of adjacent spatial areas infected for each time step • Proportion of shared borders • Prevalence at several time steps • Proximity of area to features influencing rate of spread	• Prediction of probability an area will become infected at a given time step • Prevalence prediction in areas through time • Estimates of covariates effects
Trend Surface	Landscape Level (Statewide)	Evaluations in infectious, moderate to rapid spreading diseases presenting themselves	Not useful for CWD because date of detection is unrelated to date of disease initiation	• Infection detection date for spatial area • Centroid coordinates for spatial areas	• Contours of months to first reported case of disease • Rate of disease spread

Summary Table of Methods

Model	Scale	Usefulness	CWD Usefulness	Data Required	Potential Output
Compartment Models	Local or Fine-Scale	Wide utility in the disease ecology field.	Can be elaborated to incorporate a number of relevant CWD characteristics, but these models often do not include an environmental transmission component.	• Dependent on model structure and level of detail • Detailed models may require movement rates, sex and age composition, disease state, social contacts, effects of infection on vital rates, and resistance factors	• Minimal: number of individuals in each disease class of susceptible, infected and recovered or dead at each time step
Individual-Based Models	Local or Fine-Scale	Simulate a wide variety of situations in animal ecology and varies by model structure and level of detail	Useful to incorporate known individual heterogeneity in contact patterns.	• Dependent on model structure and level of detail • Theoretical model: existing observations of population structure and disease prevalence • Detailed models may require population age and sex composition, disease prevalence, animal movement rates, contact rates	• Minimal: population structure and disease state through time in the population by age and sex • Detailed models could produce harvest and treatment variables, location of animals, animal densities across landscape, physiological state, genetic composition, number of offspring, indices of genetic diversity, gene flow rates, etc.
Network Models	Local or Fine-Scale	Flexibility in simulating various spatial/social structures	Additional complexity and data requirements are probably unnecessary to adequately model CWD due to the environmental transmission and chronic nature of CWD	• Estimates of interconnectedness or association of individuals/groups	• Predictions of disease spread across the network. • Comparison of management scenarios that focus on particular individuals or groups.
Spatial Stochastic Models	Local or Fine-Scale	Evaluating population dynamics where spatial heterogeneity is important	Application dependent.	• Limited data often restricts the predictive capacity of these models. Potentially useful for "what-if" scenarios where the range of potential outcomes is important.	• Mean and variation in the potential outcomes for a given epidemic, location, or management strategy.

Acknowledgements

Photo courtesy of USDA

HISTORY OF CWD WORKSHOPS FROM WHENCE THIS DOCUMENT ORIGINATES

In 2001, the discovery of chronic wasting disease (CWD) in Wisconsin and in other areas caused many states to initiate CWD surveillance programs. To enhance coordination between states and increase the rate of information transfer and understanding of the disease, the USGS promoted and funded the first multi-state workshop on CWD, held in Madison, Wisconsin in 2002. The workshop focused on designing, developing, and implementing CWD surveillance programs for free-ranging cervids. The main objectives were to define the surveillance goals, establish the key operational and logistical components for conducting a surveillance program, and develop prototype statistical methodologies and procedures for CWD surveillance (detection, distribution, and monitoring). Participants in the workshop included wildlife managers and biologists, epidemiologists, biostatisticians, and ecologists. A white paper reviewing the current state of knowledge on planning, conducting, and evaluating a CWD surveillance program, originated from the first workshop. The white paper from the first workshop has been well utilized by wildlife managers and biologists involved with CWD research as well as research on other wildlife diseases (see www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf).

Left to right: Julie Blanchong, Don McKinnon, Gary White (Kneeling), Matt Farnsworth, Bruce Trindle, Rob Gillies, Steve Morey, Dennis Heisey. (Missing: Dan O'Brian, Mike Samuel, Nohra Mateus-Pinnilla)

With the success of the first workshop, a second multi-state CWD workshop was held in Utah in 2004. The first workshop addressed source data and statistical guidelines to collect surveillance data for a variety of outputs. The second workshop, from which this document arose, followed logically from the first workshop and focused on analysis, interpretation, and use of CWD surveillance and related epidemiology data, with an emphasis on modeling approaches by which to gain a better understanding and describe the spatial epidemiology of CWD. Again, participants included wildlife biologists and managers, along with epidemiologists, biostatisticians, and ecologists from around the country. This document stems directly from the overarching objectives and specified outputs of the second workshop. Specifically, we addressed the goal to produce a white paper for wildlife managers and researchers that discusses the relevance of various spatial epidemiology tools and explanation of how they might be used. We hope this document, "Scale-dependent approaches to modeling spatial epidemiology of chronic wasting disease", fulfills that goal.

Left to right: Paul Cross, John Gross, Mike Miller, Mary Conner, Dennis Heisey, Steve Griffen, David Douglas. (Missing: Tim Van Deelen)

Table of Contents

Introduction

A field collection site where employees take samples from hunter harvested deer to check for the presence of CWD. (photo courtesy of USDA/APHIS)

DISEASE BACKGROUND AND MOTIVATION FOR REPORT

Chronic wasting disease (CWD) (Williams and Young 1980) is a contagious prion disease of deer (*Odocoileus* spp.) and elk (*Cervus elaphus nelsoni*) (Williams and Young 1993, Williams and Miller 2002), which also has been recently found in moose (*Alces alces*) (Kreeger et al. 2006). Although once thought only to be endemic to northcentral Colorado and southeastern Wyoming, CWD has been detected in free-ranging and captive herds throughout the United States and in two Canadian provinces (Figure I.1). CWD is the only prion disease known to affect free-ranging species (Williams and Miller 2002), and currently is found in is found in a growing number of privately owned captive cervid herds. With more sensitive diagnostic capabilities and increased surveillance for the disease, new foci continue to be documented.

Epidemiological investigations of CWD are in a nascent stage and a thorough understanding of the factors that contribute to the occurrence, the onset of clinical symptoms, and the transmission of CWD remains incomplete. In addition to potentially reducing population growth rates of free-ranging deer and elk populations, which may have an economic impact on wildlife agencies and cascading ecological impacts, CWD has already had an economic impact on commercial cervid operations and may have a deep economic impact if it spreads to livestock or proves to be a threat to human health (Williams and Miller 2002).

1
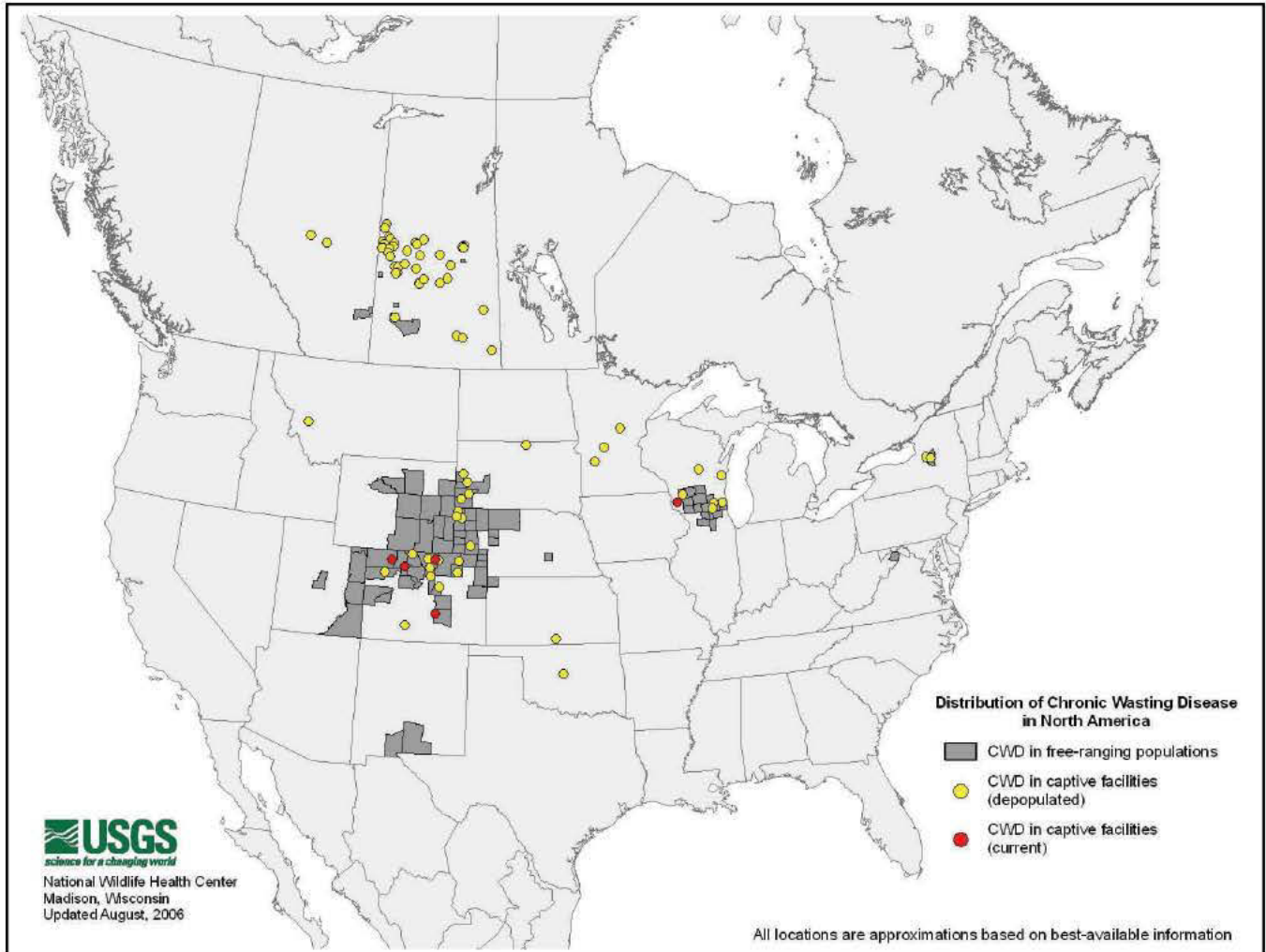
Figure I.1. Distribution of CWD in North America.

Given the potential risk and concern about the spread of CWD, there is need to equip ecologists, biologists, and managers with an array of tools to further understand the disease and enhance management efforts. Modeling is a powerful tool that can facilitate formulation and evaluation of potential management strategies, identify risk factors, predict areas at risk, and guide research to enhance mechanistic and biological understanding of CWD transmission and epidemiology.

This e-book is the product of a second workshop promoted by the United States Geological Survey (USGS) to promote cooperation between states for the management of CWD.

The first workshop addressed issues surrounding the statistical design and collection of surveillance data for CWD (Samuel et al. 2006). The second workshop, from which this document arose, followed logically from the first workshop and focused on appropriate methods for analysis, interpretation, and use of CWD surveillance and related epidemiology data. Consequently, the emphasis of this e-book is on modeling approaches to describe and gain insight of the spatial epidemiology of CWD.

We designed this e-book for wildlife managers and biologist who are responsible for the surveillance of CWD in their state or agency. We chose spatial methods that are popular or

Introduction

common in the spatial epidemiology literature and evaluated them for their relevance to modeling CWD. Our opinion of the usefulness and relevance of each method is based on the type of field data commonly collected as part of CWD surveillance programs and what we know about CWD biology, ecology, and epidemiology. Specifically, we expected the field data to consist primarily of the infection status of a harvested or culled sample along with its date of collection (not date of infection), location, and demographic status. We evaluated methods in light of the fact that CWD does not appear to spread rapidly through wild populations, relative to more highly contagious viruses, and can be spread directly from animal to animal or indirectly through environmental contamination.

We organized this e-book into 3 chapters by scale and extent for which each spatial epidemiology method was developed or best suited. Specifically, we classify methods into regional, landscape, and fine scales. The first chapter covers methods appropriate to multi-jurisdictional (e.g., multi-state or multi-providence) modeling, which we call "regional" scale. The second chapter covers methods appropriate for statewide or large areas within a state, such as wildlife management unit or county, or for metapopulations, which we call "landscape" scale. The third chapter covers methods appropriate for small areas or local populations, which we call "fine" scale. We know this rubric is somewhat artificial because many methods work at multiple scales. However, when integrating empirical data most methods work best at a particular scale. For example, individual-based models work best at modeling spread within populations, while risk analysis is most useful for summarizing data over larger scales such as a region. We chose to organize the e-book by scale to make it easily accessible to wildlife managers and biologists. Because some methods are applicable at several scales, however, we include a graphic at the beginning of each method that indicates the range of scales for which it applies.

Box I.1. PRION DISEASES

Prions are proteinaceous, infectious disease agents that propagate by converting a normal host protein into an abnormal form. The host protein from which prions can arise is called cellular prion protein (PrP) or PrP^C. PrP^C occurs naturally on the cell surfaces of healthy individuals of all mammalian species studied to date. PrP^C becomes a disease-associated prion when the normal protein folds into an abnormal conformation. This novel PrP conformation (PrP^{RES}) is relatively resistant to proteases, the cellular enzymes that normally break down proteins. PrP^{RES} accumulates during the course of disease in brain and, for some strains and host species, other tissues.
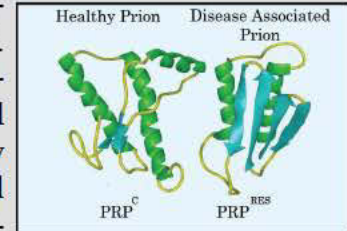
Prion diseases are a family of rare progressive neurodegenerative disorders that affect both humans and animals. Prion diseases are often called transmissible spongiform encephalopathies (TSEs) because, microscopically, the brain shows large vacuoles in neurons associated with prion accumulation. Traditionally, it was thought that all infectious agents responsible for transmissible diseases bore at least some genetic material, either DNA or RNA, in order to propagate in a host. Because prions do not have a nucleic acid genome, the assertion by Stanley Prusiner in 1982 that proteins alone could transmit an infectious disease came as a considerable surprise to the scientific community, and some debate about the true nature of the causative agent continues today. Prion disease transmission is thought to occur because infectious prions are able to induce abnormal folding of normal PrP^C in adjacent cells, allowing the agent to spread within an individual in a crystallization-like process. Of the relatively few known prion diseases, some appear to arise spontaneously, some are heritable, some arise through

Introduction

Box I.1. PRION DISEASES
(continued from last page)

cannibalism, consumption of, or other exposure to infected tissues, and some – including chronic wasting disease (CWD) – are contagious in their natural hosts.

Endogenous PrP^C proteins in their normal largely α-helical form

Abnormal form of the protein, PrP^Sc for scrapie; it is composed primarily of beta sheets

PrP^Sc interacts with the normal proteins, catalyzing the transformation of PrP^C into PRP^Sc

Once initiated, this leads to a chain reaction of PrP^Sc accumulation. This process somehow causes damage to cells of the nervous system, eventually leading to the death of infected animals.

BIOLOGY OF CWD WITHIN AND BETWEEN ANIMALS RELEVANT TO MODELING

There are several aspects of CWD pathogenesis and epidemiology relevant to modeling the disease. First and foremost, CWD is a chronic disease that develops slowly in an individual, taking about 2 years, and sometimes much longer, for infected animals to become clinically symptomatic in captivity (Williams and Miller 2002). Once clinical signs appear, deer in pen-settings typically die in a few weeks to a few months (Williams and Miller 2002). Because the intensity of prion contamination in the environment is likely to be lower for free-ranging populations than captive deer, we speculate that the incubation period may be longer in free-ranging animals. Though, once symptomatic, free-ranging animals may succumb to the disease more quickly due to indirect factors such as increased mortality from predation, hunting by humans, or collision with vehicles. The relationship between disease progression in captive mule deer and free-

ranging deer remains unresolved. The time from infection to onset of clinical symptoms is likely to depend upon host genetics (Jewel et al. 2005, Fox et al. 2006, Johnson et al. 2006), infectious dose, and nutritional status or other factors, all of which tend to vary between captive and free-ranging populations. However, because of the extremely high exposure rates in captive studies, which likely result in high dose rates, the rate of disease progression in captive deer may represent the minimum expected in free-ranging deer.

Second, CWD is laterally transmitted through both direct animal-to-animal contact (Miller and Williams 2003) and through indirect environmental contamination pathways (Miller et al. 2004). Prions are highly resistant to environmental conditions that are lethal to other pathogens, as well as to a variety of treatments that typically kill or inactivate conventional infectious agents, such as bacteria or viruses (Brown and Gajdusek 1991). In addition, it appears that some proportion of CWD prions have the ability to remain infectious in the environment for years (Pálsson 1979, Miller et al. 2004).

Besides disease characteristics, demographic patterns of CWD prevalence in mule deer indicate that deer social structure may influence transmission, although the routes of transmission in free-ranging deer are not fully known (Miller et al. 2006). Prevalence of the disease in breeding-age males is 2-4 times higher than in younger males or females (Figure I.2) (Miller et al. 2000, Miller and Conner 2005). Higher prevalence among males, relative to females, may be due to higher exposure risks, higher susceptibility, or lower disease mortality.

At present, higher exposure risk appears to be the most plausible hypothesis. Observations from captive male and female mule deer suggest they are equally susceptible to infection (Williams and Young 1980, Williams and Miller 2002, Miller et al. 2004). No field data exist to suggest diseased males survive for

Introduction
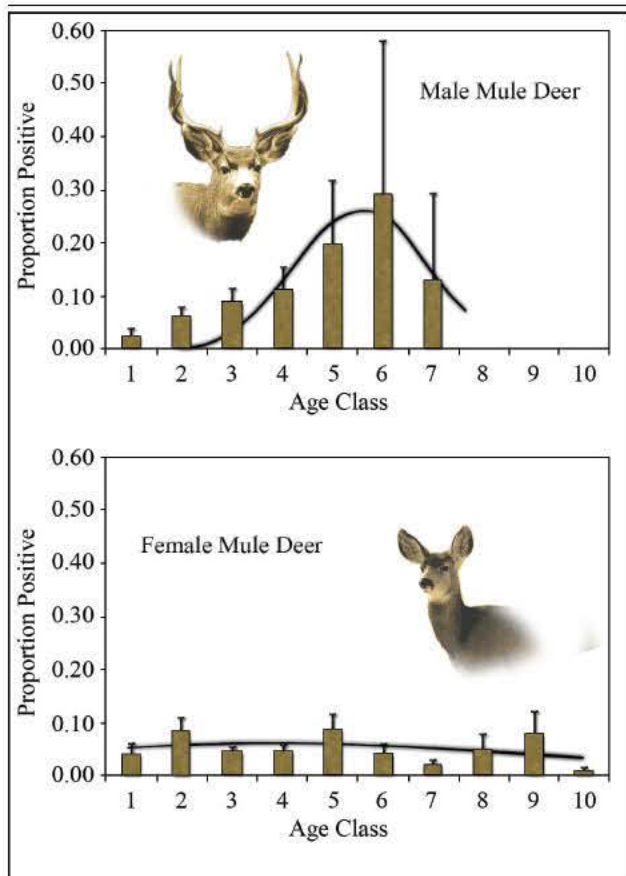
Figure I.2. *Prevalence of CWD by age class for mule deer; based on tooth cementum data collected from surveillance samples in Colorado (1998-2002). Error bars represent 95% confidence intervals. Data from Miller and Conner (2005).*

longer compared to females, and data from pen studies show no difference in survival times of infected males and females (Mike Miller, Colorado Division of Wildlife, unpublished data). Further, male mule deer typically have larger home ranges (Kucera 1978, Kufeld and Bowden 1995) and more social interactions than females (Koutnik 1981), which potentially increases their chance of either direct or indirect contact with infectious CWD prions. Perhaps more importantly, mature male mule deer also practice serial polygyny, where they canvas as many females as possible by sniffing and licking the vulva of females to detect estrus (Estes 1972, Kucera 1978, Geist 1981). Because courtship in mule deer is often protracted (Geist 1981), such contacts likely occur repeatedly during the breeding season. If

Box I.2. PRION PROPAGATION & CHRONIC WASTING DISEASE

CWD-associated prion (PrP^{CWD}) has been identified by use of a special staining process (immunohistochemistry) in neurological tissue of affected cervids, as well as in various tissues of the immune system (collectively called lymphoid tissues), including those associated with the digestive tract (e.g., tonsil, Peyer's patches, and mesenteric lymph nodes). This distribution pattern suggests that PrP^{CWD} may be shed through the alimentary tract.

Brain tissue infected with CWD prions
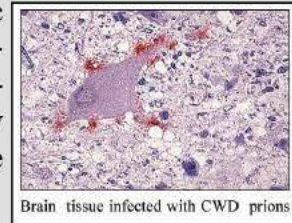
CWD infection is thought to arise after prions are ingested and subsequently taken up by the tonsils and Peyer's patches. Phagocytic cells at these sites likely engulf the agent, which may then propagate and spread to secondary lymphoid organs where propagation continues. Eventually, PrP^{CWD} enters the nerves of the gastrointestinal tract and ascends to the spinal cord and the brain, where further propagation ensues.

PrP^{CWD} propagation typically occurs much earlier in the disease course in the lymphatic system than in the nervous system. Accumulation of PrP^{CWD} in lymphoid tissues occurs without any appreciable physiological consequence, but accumulation in the central nervous system eventually leads to clinical CWD and death of infected animals. The clinical signs of CWD include behavioral changes, loss of body weight, emaciation, excessive salivation, increased drinking and urination, stumbling, trembling, and depression; behavioral changes include decreased interaction with other animals, listlessness, lowering of the head, blank facial expression, and repetitive walking in set patterns. The diagnosis of CWD is based on examination of the brain or lymphoid tissue for spongiform lesions and/or

(continued on next page)

**Box I.2. PRION PROPAGATION &
CHRONIC WASTING DISEASE
(Continued from last page)**

accumulation of PrP^{CWD} by immunohisto-chemistry.

Mule deer doe in the mid to late stages of CWD infection.

CWD is transmitted via excreta (Williams and Miller 2002, Miller and Williams 2003, Mathiason et al. 2006), then these behaviors may increase the exposure risk of males. Such behaviors may increase male contact with infectious agent and, accordingly, increase risk of contracting CWD, explaining the observed demographic differences in prevalence.

OBJECTIVES

"All models are wrong, some are useful."
(Box 1979)

Virtually all decisions on natural resource management occur in the face of incomplete data, within a complex and often politically-driven setting, and in systems subject to "surprises". Conceptual and quantitative models offer a means to formally express our understanding of key processes and system dynamics, articulate hypotheses on how we think the system works, and to facilitate communication among diverse audiences. In general, the accuracy of the quantitative model increases with the availability of reliable data. However, even when the data are inadequate, quantitative models can offer a means to examine key relationships and guide management decisions (Starfield and Bleloch 1986).

Disease models can take on an almost infinite variety of mathematical structures and levels of complexity or reality. They may aggregate spatial processes into one or a few simple equations, or explicitly represent spatial heterogeneities in great detail. A model can include random (stochastic) events, or have fixed parameters (deterministic); it can represent a system at a particular point in time (static model), or permit changes over time (dynamic). Table I.1 provides a relatively simple summary of common model types and the characteristics that distinguish each. Differences in model characteristics can overlap, and in these cases it may be a matter of opinion whether a model falls within a category or not. For example, a statistical model composed only of a linear regression is clearly a statistical model (and not a mechanistic model), but the parameters of most mechanistic models can be estimated using a statistical model.

Models of infectious diseases can integrate epidemiological and biological data to give insights into patterns of disease spread (among animals or geographically) and the effect of management interventions. A wide variation in model structures can be used tackle a range of questions and spatial scales. Examples in wildlife and livestock disease systems include analysis of the factors affecting the spread and control of rabies (Smith et al. 2002), foot-and-mouth disease (Ferguson et al. 2001), and bovine tuberculosis (Cross et al. 2004). Here we focus on modeling approaches that are commonly applied to a single population, or to the interaction among populations. However, as modeling sophistication and computing power improve, our ability to use computationally or data intensive approaches at broader scales is rapidly increasing (Ferguson et al. 2005). Similarly, with the increase in computing power, virtually any temporally dynamic model can be extended to incorporate spatial components with the addition of movement rules.

In the sections that follow, we attempt to summarize the utility of different modeling

Model	Explanation
Mathematical	Explicitly represents mechanisms, such as physiological processes, breeding or disease transmission between individuals, or other system-relevant processes. Parameters for process equations generally have an identifiable link to reality. Also called process-based or control model.
Statistical	An equation that describes an observed relationship, but based only on a statistical relationship. Equation parameters usually have no meaningful interpretation, and the domain of inference is usually restricted to very specific areas and conditions. Also called descriptive or phenomenological model.
Dynamic	Explicitly represents change over time.
Static	Describes a system state at a particular point in time.
Non-spatial	Spatial relationships are not explicitly represented , although an area may be implied by representing population density or other area-specific measures. Most models of single populations are non-spatial. Also called spatially homogenous.
Spatially distributed	A non-spatial model applied across an area, but without communications or transfers of energy or materials between spatial units of the model. An example is a point based ecosystem that is applied to a grid of cells, where the model in each cell has unique weather, soils, etc., but seeds or other propaguls are not exchanged between grid cells.
Spatially explicit	A model with communication of attributes between distinct areas that are explicitly represented in the model, and material and energy are exchanged between areas. For example, a disease model where each county is characterized by population density, topography, and disease transmission occurs between counties based on county attributes and disease incidence. Also called spatially heterogeneous.
Stochastic (random)	Random events are included in the model so each instance of the model generates a different result. Model results are typically presented as a distribution or probability of an outcome. Typical random events include weather (temperature, rainfall), variation in contact between infected and susceptible individuals, or random variation in mating or recruitment (especially in small populations). Many consider stochastic models to be more realistic than deterministic models, but they can be much more difficult to evaluate.
Deterministic	Parameters are fixed and each model outcome is exactly the same for a given set of inputs. Common for models composed of differential equations. Deterministic models may have a closed-form (analytical) solution, which can generally simplify analysis.
Analytic	A mathematical model that is solved solely through the use of mathematical arguments and not by numerical approximations or other simulations.
Individual-based	Each individual is explicitly represented in the model. For example, a disease model where the sex, age, disease status, and contacts of each animal in a population are followed for the entire life of each animal.
Spatially explicit	Any model whose behaviors and solutions are obtained by numerical approximations and not by mathematical arguments. Virtually always involves the use of computers.

Table I.1. *Characteristics used to describe or classify models routinely used to model disease dynamics. These terms are not all exclusive, and it may be accurate (and appropriate) to describe a model as having two or more model types. Modified from Haefner (1996).*

approaches to extend our understanding and management of CWD. In particular, we describe and evaluate spatial modeling approaches previously used for wildlife and/or human diseases in the context of CWD. We focus on three spatial scales and present approaches relevant to each. We use spatial scale because for wildlife systems the data available often dictate which scale and resolution are possible. At finer spatial scales, managers and researchers may track known individuals. At broader spatial scales however, this becomes impossible due to logistical and financial constraints. At a state or national scale often the only data available are cross-sectional disease surveys. In the case of CWD, this is primarily done through surveillance of hunter-harvested deer and elk. It is important to note that although data are often limited, the modeling approaches we discuss may be implemented at different spatial scales. For each approach we present the questions addressed, input data required and existing data, model outputs and interpretation, and the use of these results. In addition, we discuss the overall effectiveness of the specific approach to CWD modeling.

We also select a single modeling approach for each scale, which we define as the focal approach. The focal approach is a method that we believe to be applicable to modeling CWD, as well as being a useful tool for identifying management strategies. Note that the epidemiological focus and goals of models vary by scale and these aspects will be discussed later. Using a close facsimile (but not exact to protect unpublished data) of previously obtained CWD surveillance data, we work through the method as an illustration of the strengths and shortcomings of that specific approach. The approaches for each scale are be presented in order, based on our conclusions, from most to least relevant for modeling CWD. At each scale, besides presenting a focal method, we summarize pros and cons of all methods, discuss current data limitations, identify any data gaps for relevant approaches, and describe where future work may be most fruitful.

LITERATURE CITED:

Box, G. E. P. 1979. Some problems of statistics and everyday life. Journal of the American Statistical Association 74:1-4.

Brown, P., and D. J. Gajdusek. 1991. Survival of scrapie virus after 3 years interment. Lancet 337:269-270.

Cross, P. C., J. O. Lloyd-Smith, J. Bowers, C. Hay, M. Hofmeyr, and W. M. Getz. 2004. Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. Annales Zoologici Fennici 41:879-892.

Estes, R. D. 1972. The role of the vomeronasal organ in mammalian reproduction. Mammalia 36:315-341.

Ferguson, N. M., C. A. Donnelly, and R. M. Anderson. 2001. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. Science 292:1155-1160.

Ferguson, N. M., D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437:209-214.

Fox, K.A., J. E. Jewell, E. S. Williams, and M. W. Miller. 2006. Patterns of PrP(CWD) accumulation during the course of chronic wasting disease infection in orally inoculated mule deer (*Odocoileus hemionus*). Journal of General Virology 87:3451-3461.

Geist, V. 1981. Behavior: adaptive strategies in mule deer. Pages 157-223 *In* Mule and black-tailed deer of North America. O. C. Wallmo, Editor. University of Nebraska Press, Lincoln, Nebraska, USA.

Haefner, J. W. 1996. Modeling biological systems: principles and applications. Chapman and Hall, New York, New York, USA.

Jewell, J. E., M. M. Conner, L. L. Wolfe, M. W. Miller, and E. S. Williams. 2005. Low frequency of PrP genotype 225SF among free-ranging mule deer (*Odocoileus hemionus*) with chronic wasting disease. Journal of General Virology 86:2127-2134.

Johnson C, J. Johnson, J. P. Vanderloo, D. Keane, J. M. Aiken, and D. McKenzie. 2006. Prion protein polymorphisms in white-tailed deer influence susceptibility to chronic wasting disease. Journal of General Virology 87:2109-2114.

Koutnik, D. L. 1981. Sex-related differences in the seasonality of agonistic behavior in mule deer. Journal of Mammalogy 62:1-11.

Kreeger, T. J., D. L. Montgomery, J. E. Jewell, W. Schultz, and E. S. Williams. 2006. Oral transmission of chronic wasting disease in captive Shira's moose. Journal of Wildlife Diseases 42:640-645.

Kucera, T. E. 1978. Social behavior and breeding system of the desert mule deer. Journal of Mammalogy 59:463-476.

Kufeld, R. C., and D. C. Bowden. 1995. Mule deer and white-tailed deer inhabiting eastern Colorado plains river bottoms. Colorado Division of Wildlife Technical Publication 41, Fort Collins, Colorado, USA.

Mathiason C K., J. G. Powers, S. J. Dahmes, D. A. Osborn, K. V. Miller, R. J. Warren, G. L. Mason, S. A. Hays, J. Hayes-Klug , D. M. Seelig, M. A. Wild, L. L. Wolfe, T. R. Spraker, M. W. Miller C. J. Sigurdson, G. D. Telling, and E. A. Hoover. 2006. Infectious prions in the saliva and blood of deer with chronic wasting disease. Science 314:133-135.

Miller, M. W., E. S. Williams, C. W. McCarty, T. R. Spraker, T. J. Kreeger, C. T. Larsen, and E. T. Thorne. 2000. Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. Journal of Wildlife Diseases 36:676-690.

Miller, M. W., and E. S. Williams. 2003. Horizontal prion transmission in mule deer. Nature 425:35-36.

Miller, M.W., E. S. Williams, N. T. Hobbs, and L. L. Wolfe. 2004. Environmental sources of prion transmission in mule deer. Emerging Infectious Diseases 10:1003-1006.

Miller, M. W. and M. M. Conner. 2005. Epidemiology of chronic wasting disease in free-ranging mule deer: spatial, temporal, and demographic influences on observed prevalence patterns. Journal of Wildlife Diseases 41:275-290.

Miller, M. W., N. T. Hobbs, and S. J. Tavener. 2006. Dynamics of prion disease transmission in mule deer. Ecological Applications 16:2208-2214.

Pálsson, P. A. 1979. Rida (scrapie) in Iceland and its epidemiology. Pages 357-366 In Slow transmissible diseases of the nervous system: Volume I. S. B. Prusiner and W. J. Hadlow, Editors. Academic Press, New York, New York, USA.

Samuel, M. D., D. O. Joly, M. A. Wild, S. D. Wright, D. L. Otis, R. W. Werge, and M. W. Miller. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk. USGS-National Wildlife Heath Center Report. Madison, Wisconsin, USA. Available at: http://www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf .

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

Starfield, A. M., and A. L. Bleloch. 1986. Building models for conservation and wildlife management. Burgess International Group Inc., Edina, Minnesota, USA.

Walters, C. 1986. Adaptive management of renewable resources. MacMillan, New York, New York, USA.

Williams, E. S., and S. Young. 1980. Chronic wasting disease of captive mule deer: a spongiform encephalopathy. Journal of Wildlife Diseases 16:89-98.

Williams, E. S., and S. Young. 1992. Spongiform encephalopathies in Cervidae. In Transmissible spongiform encephalopathies of animals. R. Bradley and D. Mathews, Editors. Revue Scientifique et Technique Office International des Epizooties 11:551-567.

Williams, E. S., and S. Young. 1993. Neuropathology of chronic wasting disease of mule deer (*Odocoileus hemionus*) and elk (*Cervus elaphus nelsoni*). Veterinary Pathology 30:35-45.

Williams, E. S., and M. W. Miller. 2002. Chronic wasting disease in deer and elk in North America. Revue Scientifique et Technique Office International des Epizooties 21:305-316.

Introduction

Approaches to Regional-Scale Modeling
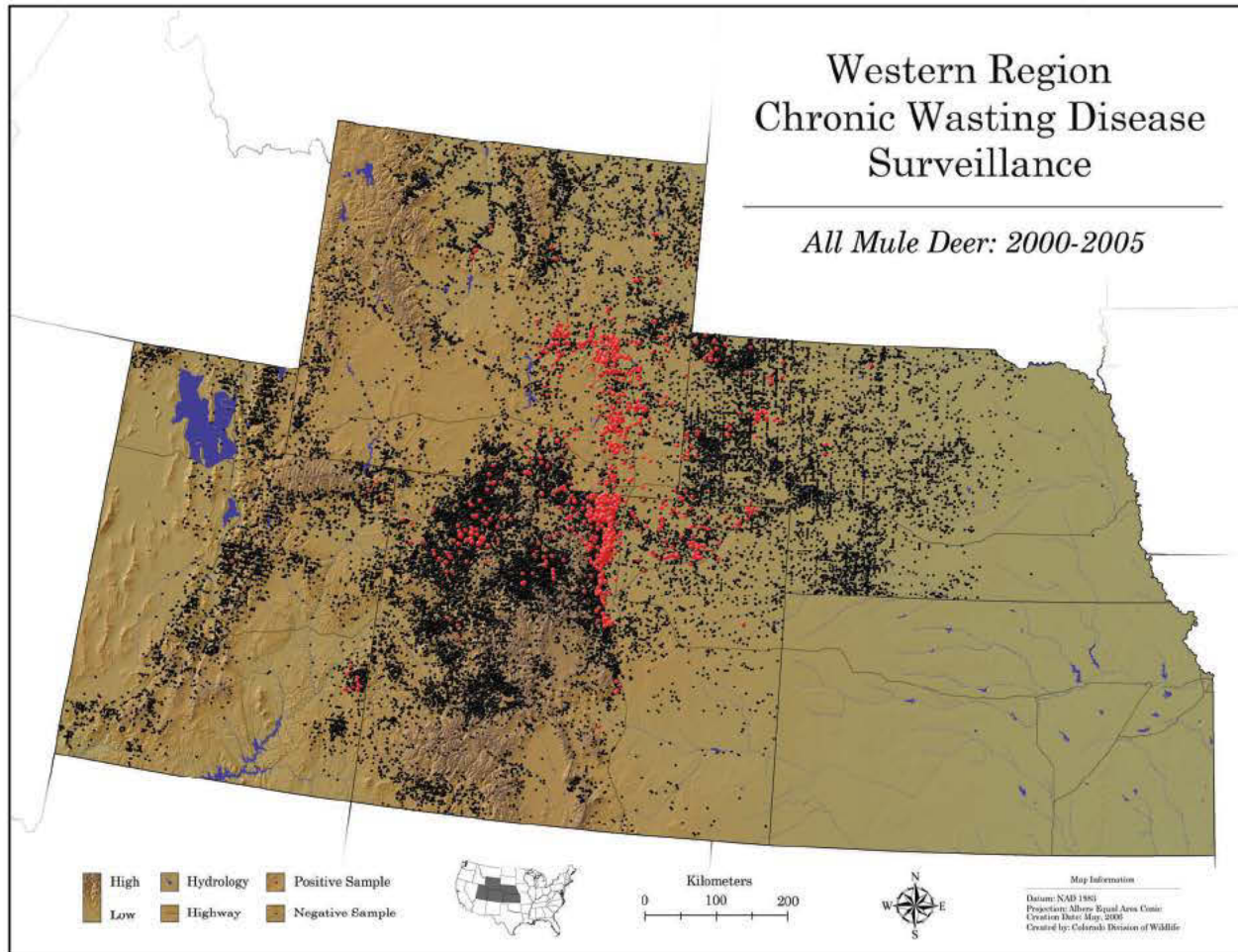
Chronic wasting disease samples from the Western United States region (2000-2005).

We consider the term regional-scale to describe large areas; large enough to be relevant to multi-jurisdictional (e.g., multi-state or multi-providence) management efforts and regional description of CWD epidemiology. Regional-scale epidemiology focuses on the biotic and abiotic factors that influence the observed spatial pattern of a disease, and its spread between populations (see Hess et al. 2002) or across large areas. At the regional scale, data are generally summarized by county, wildlife management unit, or other relatively large areas, such as a state or providence. Data collected typically include time to next case (rate) or counts of positive and nega-

tive samples within a given area (area is also called "tract" in epidemiology literature). For a region, major goals include evaluating biological and ecological risk factors and predicting high-risk areas. At this scale, researchers and managers also attempt to identify likely corridors of disease spread and potential barriers that could be used to arrest proliferation.

From a regional perspective, an introduced wildlife disease may appear as a point source with diffusion (Figure 1.1), as observed for bovine tuberculosis in white-tailed deer (Schmitt et al. 1997, Hickling 2002) in Michigan, USA, and raccoon rabies in the northeast-

Approaches to Regional-Scale Modeling
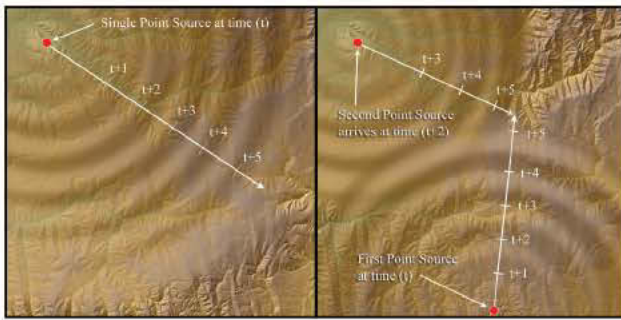
Figure 1.1. Visual depiction of wave front phenomena emanating from single (left) and multiple sources (right).

ern USA (Jenkins and Winkler 1987, Moore 1999). Established epidemics may show a diffusion wave front, as seen in fox rabies in Europe (Kallèn et al. 1985, Smith and Harris 1991), or multiple point sources, as seen in anthrax epidemics in African ecosystems (Prins and Weyerhaeuser 1987). Although a regional view of an epidemic may suggest diffusion across a landscape, finer resolution may reveal a more patchy distribution and heterogeneity in the rate of spread. For example, the pattern of raccoon rabies in Pennsylvania, USA, appeared consistent with simple diffusion when viewed from a large, geographic perspective. Subsequent analyses, however, revealed areas of slow spread (barriers), high prevalence areas, and rapid local spread (corridors) that did not conform to simple diffusion model predictions (Moore 1999). Predictions at the regional-scale can be useful to describe disease occurrence, prevalence, or spread in general terms or for political reasons. Yet, patterns of disease spread at a biologically relevant scale may be poorly represented by values averaged over a large spatial area.

Spatial epidemiological models on a large scale are geared toward one of two distinct forms of disease: highly contagious infectious diseases that generate rapidly moving epidemic fronts, such as foot-and-mouth disease, or non-infectious diseases, such as non-viral cancers or toxin related illness. Because of the chronic nature of CWD infection in animals, the putative slow rate of transmission within populations, and the relatively slow rate of geo-

graphic spread, CWD epidemiology falls somewhere between these two rubrics and thus we describe approaches for both types of disease.

RISK ANALYSIS / ASSESSMENT:

Risk analysis is a term often used when evaluating effects of a known point-source (such as a power plant) or line-source (such as contaminated streams or rivers) that emits toxins or pollutants relative to proximity to the source (Morris and Wakefield 2000). Risk analysis, or risk assessment, also describes geographical studies in terms of spatial distribution of putative risk factors, and their relationship or correlation to disease risk (Briggs 2000). Overall or total risk is the response variable, which can be a variety of functions, usually summarized in a map. For example, risk could be a krigged surface of prevalence or probability of infection, or a relative risk surface or odds-ratio surface. Risk can also be expressed qualitatively, where areas are described as having high, medium, or low risk. The risk response surface could be course, wherein all risks are averaged over a cell, or continuous, such as for krigged data. The choice of a risk response variable and its spatial resolution depends on study goals and available data.

Diseases have multiple causes, and disentangling the risk factors can be complex. However, this problem is analogous to other spatial predictive models and analyses are typically conducted in a GIS framework, in which environmental or ecological landscape characteristics are used to predict presence (MacKenzie et al. 2006), distribution (Murwira and Skidmore 2005), abundance (Dunford and Freemark 2005), or other spatial traits, such as home range size (Anderson et al. 2006) of wildlife species. Spatial data have frequently been used to link predictions about species distribution and abundance to habitat characteristics, but it is less common to use this approach in predicting disease distribution or prevalence.

Perhaps this is because regionally-oriented risk models are not applicable to highly infectious, fast-spreading diseases, which have been the historical focus of spatial epidemiological modeling for regional scales. Risk analyses are typically more appropriate for endemic diseases that are spreading slowly, or not at all. In the case of a slowly spreading disease, one wants to be reasonably sure that absence of the disease is related to factors associated with the site (or individual) rather than it simply being farther away from the site where the disease was introduced. Note, however, that in more sophisticated analyses, distance from the epicenter and time since introduction could be included as covariates prior to investigating the effects of other variables. Several groups have employed spatial risk models to predict the probability of disease, or disease-vector presence, based on environmental factors: tsetse flies (Rogers et al. 1996), Lyme disease (Allen et al. 2003, Schauber et al. 2005), human induced disease of great apes (Sleeman 2005), and disease-carrying *Ixodex ricinus* ticks (Merler et al. 1996).

Questions addressed / model predictions:
1. Estimates the relationship between environmental/ecological (abiotic and biotic) factors and disease risk.
2. Potentially estimates mechanistic relationships between disease and environmental/ecological (abiotic and biotic) factors.
3.

Data required:
1. If disease cases present themselves, such as cancer cases, then only spatial locations of the positive samples are required (note that "targeted" surveillance samples for CWD may fall in this category).
2. Spatial coordinates of positive and negative samples are required if disease cases do not 'present' themselves – such as for CWD cases.
3. If any temporal aspect to the models, then the date samples were collected (as a proxy for date of infection of individuals.
4. Spatial environmental/ecological (abiotic and biotic) data for factors of interest.

Output:
1. Estimates spatial variation in disease prevalence and/or risk.
2. Estimates risk parameters, effect sizes, covariate effect sizes, and other relevant statistics for factors/variables affecting the probability of disease.
3. Provides model selection statistics and the relative weight of different models/ variables.

General usefulness:
Spatial risk models are potentially useful for non-infectious disease or diseases with low infection rates or slow epidemic fronts (i.e., relatively slow spread). Because wildlife hosts are mobile, this approach is most applicable for diseases with short dormancy or latency periods. Longer latencies would dilute the effect of some factors unless the risk factors were relatively constant and migration rates were relatively low (e.g., animals do not contact disease in one location and then move to another so that risk is unhinged from location of infected animals), whereas short latency diseases can be more directly tied to risk factors. Thus, ecological risk factors associated with short latency period are more readily identified. Spatial risk models could be used to test various hypotheses about environmental and biological covariates and risk factors, as well as to identify potential areas of disease risk.

Usefulness to CWD modeling and/or management:
If relevant spatial data are available, risk analysis has high potential applicability to CWD modeling and management. CWD could make use of a risk analysis approach, given the fact that geographical spread appears to be slow in nature (we find it where we look for it, and, at present, have seen no evidence of rapid or even moderate rate of spatial spread). Surveillance data would satisfy the three data requirements, but availability of appropriate environmental and ecological spatial data would need to be assessed, collected, and evaluated. This approach could identify environmental/ ecological risk factors to potentially target for

management, as well as high prevalence and high risk areas to target management intervention or other actions. For example, a recent study found that prions bind with varying affinity to different soil minerals (Johnson et al. 2006) and Farnsworth et al. (2005) found greater CWD risks in urban areas. Risk analyses could be used to predict how soil types and housing density may affect current and future hotspots of CWD, but the power of the analyses would depend upon host movement, spatial scale of variation in soil type and the amount and resolution of available data.

MICROMAPS

Micromaps represent a different paradigm for graphical visualization of diseases (Box 1.1) compared to other common techniques such as chloropleth maps (see Figure 2.6) or smoothed preva-

Box 1.1. A CLOSER LOOK AT MICRO-MAPS

lence or risk surfaces (see Figure 1.7). The chloropleth map, which uses either shading or color to differentiate the values from one region to another, is somewhat problematic as noted by Dent (1993) and Harris (1999). The main issue is the color representation of continuous data, which usually necessitates the conversion of continuous data into discrete intervals and going from infinite color gradation to a limited number of colors (Symanzik and Carr 2007). Symanzik and Carr (2007) provide details on issues pertaining to region area size, the loss of information when making continuous information discrete and the inability to represent confounding variables.

Given that most CWD data is georeferenced, and the response variables of primary interest are numerical, linked micromap plots can be used to order and present this multivariate data in a contextual structure. In addition, confidence intervals for variables, such as prevalence, can be shown on micromaps. This is an important piece of information missing from chloropleth maps. Thus, linked micro-

A typical, but hypothetical, linked micromap plot showing the four key features of the technique (Carr and Pierson 1996). Note: maps display simulated (not real) data. The first (leftmost) panel, Maps, contains a map of the region. The second panel, ID, provides the names of the geographical regions (here, Region 1 through Region 10). The third through the fifth panels display statistical summaries. These panels may represent many forms of statistical summaries including box-plots, dot-plots (as shown), time series plots, confidence intervals, etc. Sorting the geographic regions based on the statistical variable(s) of interest is the second feature. Sorting improves perception between consecutive panels from top to bottom of the display. The

Hypothetical Micromap

Box 1.1. A CLOSER LOOK AT MICROMAPS *(continued from last page)*

third feature is the partitioning of the regions into perceptual groups of size five or less to allow the viewer's attention to focus on explicit areas at a time. The fourth feature is color and location that links corresponding elements within the parallel sequence panels, i.e., the color red in the topmost panels relates to the geographic region in the Northeast of the map, the area name Region 5, and a red dot in each of the three statistical panels. The color red is reused in the next consecutive set of panels for Region 2, but there is no relationship between Region 5 and Region 2 as one might at first assume. Simply, there are not enough distinguishable colors to populate an entire display (with, say, 50 different regions); consequently, colors have to be reused in different panels.

The data displayed in statistical panels 1 and 2 show a strong positive association (the correlation r calculated as 0.99), expressed in the almost parallel behavior of the dots and lines representing the values for these two variables. In contrast, the statistical data in panel 3 and 1 (as well as 3 and 2) show a strong negative association (the correlation r calculated as –0.94 for 3 and 1 and as –0.92 for 3 and 2). This negative association is seen in the movement of the dots and lines in opposite directions for these variables. Moreover, the data in panel 3 show an unusual outlier, the value for Region 1. It is this outlier that considerably reduces the almost perfect negative association otherwise present in this data. Just a simple numerical calculation of r might not be able to reveal the influence of a single region on the overall relationship.

The map panels of the linked micromap plots from the previous page exhibit a strong geographic pattern: Highest occurrences with respect to the statistical panel 1 can be found in the North and in the East; lowest occurrences can be found in the West and in the South. Additional features of linked micromap plots exist and are described in more detail in Symanzik and Carr (2007).

maps plots are visual, but also are a representation of statistical data without certain shortcomings of the chloropleth depiction (Carr 2001). This provides benefits through (a) more meaningful representation of the data and, (b) communicating complex data sets in a manner that facilitates interpretation.

An ample set of templates are available that offer users considerable flexibility in data visualization (Carr et al. 1998). For example, the statistical panel of linked micromap plots can use any number of forms including box-plots, bar-plots, histogram-plots, or time series plots (Box 1.1). These alternate statistical plots offer additional avenues to represent the underling structure of the data and examine patterns and relationships in the data.

Questions addressed / model predictions:
1. Depicts prevalence or other relevant disease statistics or information across space and time simultaneously.
2. Allows queries of the underling structure of the data.
3. Facilitates examination of patterns and relationships in the data.

Data required:
1. Summary of positive and negative samples by polygon, such as wildlife management unit, county, etc., for each time step of interest.
2. The polygon from which each positive and negative sample was taken.

Output:
1. Prevalence maps that are linked across space and time simultaneously.
2. Any statistics of interest linked across space and time simultaneously.

General usefulness:
Linked micromap plots facilitate data sort-

ing and division into smaller groups, which can be used to highlight specific spatial and temporal patterns. Linked micromap plots provide considerably more information than would otherwise be provided by a series of tables or an overall map representation (e.g., a chloropleth map) alone. Viewers can navigate through the linked micromap plot to a place of interest in order to review prevalence and related statistics, such as confidence intervals. Thus, estimates and the equally important variance of those estimates can be portrayed. This is an advantage over typical spatial summaries wherein variance is not portrayed. A second key advantage is that linked micromaps present statistical summaries and estimates in a spatial context. Unlike traditional graphical methods, linked micromap plots combine both exploratory analysis and traditional statistical graphics while maintaining the spatial context; this is important in CWD epidemiology because of its intrinsic spatial nature.

Usefulness to CWD modeling and/or management:

Linked micromaps provide highly useful visualization techniques for presenting temporal data from a large number of areas, such as wildlife management units or counties, along with associated statistics. These plots can allay information overload and facilitate interpretation of large and complex data sets; this property is extremely useful to managers who need to make timely decisions about CWD management. Linked micromap plots are a constructive GIS representation coupled to a statistical visualization tool, which provide exploratory capabilities. Linked micromap plots can be used to augment the presentation of CWD data. One example may be to micromap the results of a risk assessment model that would stratify areas based on risk to facilitate planning and mitigation practices.

CLUSTER ANALYSIS

Cluster analysis refers to a widely used set of grouping al-

gorithms that identify meaningful structures (often spatial for spatial disease epidemiology) in observed data. The conceptual approach involves grouping data so that patterns within a valid cluster are more similar to each other than to patterns belonging to different clusters (Jain et al, 1999). The scope of problems addressed by cluster analysis includes many disciplines and has led to the development of a large assortment of clustering methods. No single clustering technique is universally appropriate for uncovering the structures that may be present in high dimensional data sets. For example, while many algorithms might get close, relatively few would be able to group the data as shown below (Figure 1.2).

Each clustering technique has its weaknesses and strengths, and these must be considered in conjunction with the goals of the analysis and the nature of the data. For example, many parametric clustering algorithms tend to find
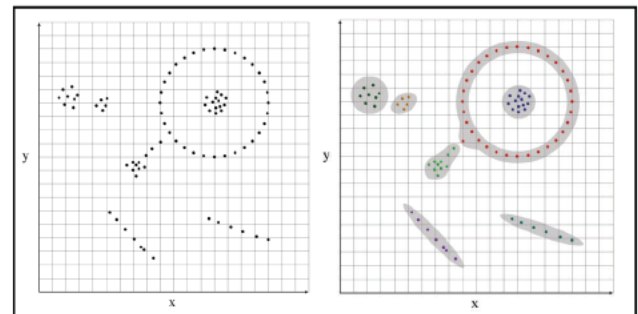
Figure 1.2. Hypothetical cluster assignment of points (adapted from Jain et al. 1999).

clusters of a particular shape (e.g., spherical), or of equal variance. However, CWD surveillance data from harvested deer are often distributed in elongated patterns (e.g., following a drainage or roadway, or within a valley.) and are not likely to yield clusters of equal variance. The spatial irregularity of CWD surveillance data reduces the choice of algorithms to those able to detect irregularly shaped clusters. Determining the distributional nature of the data set and selecting an appropriate clustering algorithm can be a time consuming and confusing process. It is important to remember that there is no "best" method covering every

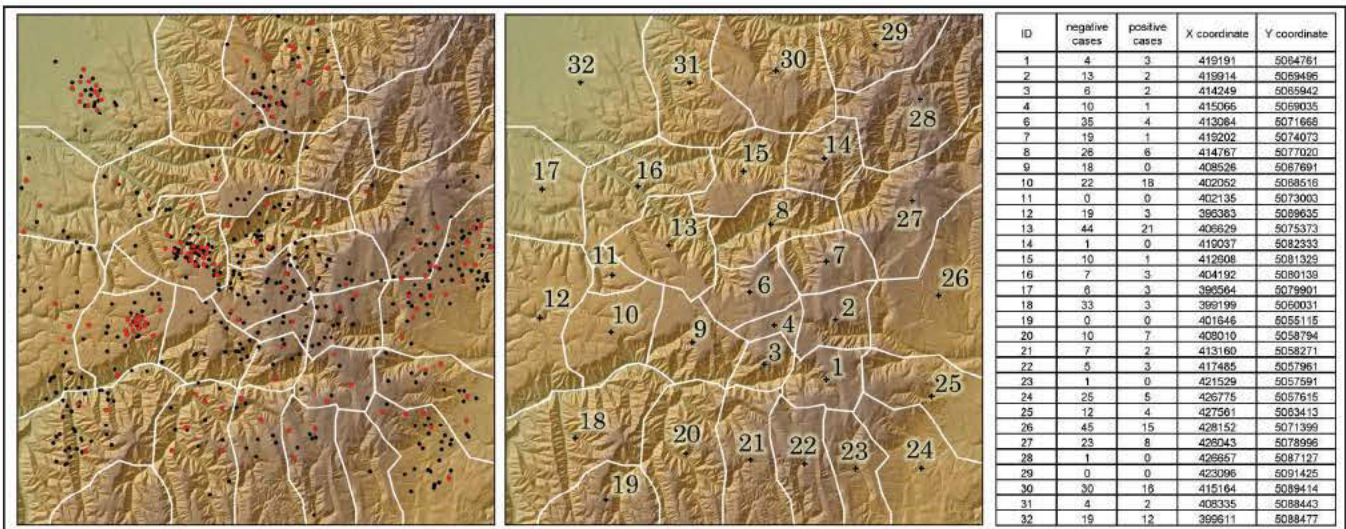situation. Accordingly, understanding strengths and weaknesses of each method is an important part of the clustering process. For example, clustering methods such as *k*-means and Ward's minimum variance method (i.e., least squares criterion algorithms) tend to select clusters with roughly the same number of observations in each cluster. Algorithms based on nonparametric density estimation such as single linkage and density linkage are generally considered to be the least biased methods for selecting clusters (SAS Institute Inc., SAS OnlineDoc® 9.1.3, Cary, NC: SAS Institute Inc., 2002-2005). However, this comes at the cost of reduced power to detect clusters. The specific goals of the analysis along with the knowledge of researchers and managers involved will be integral to the selection of appropriate clustering methods.

Clustering analysis at the regional scale can be used to determine the location of clusters as well as to answer questions about their significance (Kulldorff and Nagarwalla 1995). For the clustering of disease data it is necessary to compensate for the uneven distribution of the sampled data (Kulldorff 1997). This is particularly true with CWD surveillance efforts which vary widely across larger spatial scales. We advocate the use of spatial scan statistics for the clustering of disease data and cover this method in greater detail as the focal approach in the next section on landscape level methods.

While the use of scan statistics to detect clusters at the regional scale follows the methods outlined in the next section (i.e. at landscape level), there is one important difference: the analytical unit used for spatial clustering. For CWD data across large regions, it is often the case that information about the exact location of all samples is missing. For example, hunter harvested deer, which contribute overwhelmingly to CWD data sets, often lack geographic coordinates, but contain courser spatial information (e.g., wildlife management unit, county, etc). To make the data spatially explicit the analytical unit is shifted from the individual points to predetermined polygons. In other words, geographic coordinates (e.g., UTMs) are not required for individual CWD surveillance samples and data are effectively transformed into polygon count data (the number of positive and negative cases) with measures being concentrated at the central coordinates of the polygons (Figure 1.3). Polygons can take many forms including agency units (e.g., wildlife or game management units), administrative units (e.g., counties or postal codes), environmental units (e.g., water-sheds),

ID	negative cases	positive cases	X coordinate	Y coordinate
1	4	3	419191	5064761
2	13	2	419914	5069495
3	6	2	414249	5065942
4	10	1	415066	5069035
6	35	4	413084	5071668
7	19	1	419202	5074073
8	26	6	414767	5077020
9	18	0	408526	5087691
10	22	18	402052	5068516
11	0	0	402135	5073003
12	19	3	396383	5089635
13	44	21	406629	5075373
14	1	0	419037	5082333
15	10	1	412908	5081329
16	7	3	404192	5080138
17	6	3	390564	5079901
18	33	3	399199	5060031
19	0	0	401646	5055115
20	10	7	408010	5058794
21	7	2	413160	5058271
22	5	3	417485	5057961
23	1	0	421529	5057591
24	25	5	426775	5057615
25	12	4	427561	5063413
26	45	15	428152	5071399
27	23	8	426043	5078996
28	1	0	426657	5087127
29	0	0	423096	5091425
30	30	16	415164	5089414
31	4	2	408335	5088443
32	19	12	399611	5088477

Figure 1.3. (left) A set of hypothetical polygon "units" overlaid on a set of positive (red) and negative (black) cases. (middle) The centroid locations (black crosses) and id numbers for each of the 32 polygon units. (right) Associated table with counts of positive and negative cases along with centroid UTM coordinates.

or even arbitrary grid cells or quadrats. Abundant digital GIS data permits wide discretion in the choice of criteria for selection spatial units.

We note that aggregating into polygons, such as wildlife management units, may be the only reasonable choice if a most of the surveillance samples lack geographic coordinates. The primary disadvantage of "collapsing" data from points to polygons is loss of spatial heterogeneity within the polygons. For example, the most basic polygon attributes will contain only spatial information on centroids (or boundaries) and counts of the number of cases and non-cases within each polygon (Figure 1.3). If such "global,' or 'first-order' clustering methods mask significant variation within polygons, the analysis should be supplemented with additional 'second-order,' or local, results based smaller spatial units, or even the points themselves (if geographic data are available).

Questions addressed / model predictions:

1. *Identifies high prevalence disease areas.*
2. *Identifies regional-scale spatial disease pattern.*
3. *Identifies potential spatial covariates to disease pattern.*

Data required:
1. The polygon, which could be a large area for a regional analysis (e.g., wildlife management unit or county), from which each positive and negative sample was taken.

Output:
1. Assigns positive cases or areas to a particular cluster.
2. Maps of disease clusters.

General usefulness:
At the regional scale, the main strength of cluster analysis is that relatively course resolution data can be used to identify areas of high disease prevalence (at which management intervention could be targeted). Cluster analyses are exploratory, but can be useful for hypothesis generation. Cluster analysis is a valuable initial step in examining the spatial epidemiology of a disease.

Usefulness to CWD modeling and/or management:
Cluster analysis is a valuable descriptive tool for CWD surveillance data. In general, we recommend that cluster analysis be conducted as a first step in the examination and evaluation of large-scale CWD surveillance data. The usefulness of cluster methods for CWD is the same as described above in "General usefulness." All CWD surveillance data, georeferenced to point or area, can be used in cluster analysis. This characteristic makes cluster analysis especially viable for multi-state data where some states collect sample coordinates and others do not.

If data are not collapsed into polygons for regional-scale applications, then all the spatial aspects of location-based cluster analysis need to be addressed. In the next section on landscape-level modeling, we discuss issues relevant to location-based cluster analysis in detail. In the section on Risk Analysis/Assessment, we consider the use of kernel density estimators, a special case of cluster analysis, to generate a risk surface.

SNAPSHOT APPROACH:
(FOLLOWING KEELING ET AL. 2004)

The goal of the snapshot approach is to identify areas to which a disease is likely to spread and the rate of spread given a single snapshot of the locations and disease status of an area. This method is appealing because it only requires one data collection effort, which is less expensive and time-consuming than other approaches. While Keeling et al. (2004) refer to this as a lattice-based, grid-based, or a cellular automata approach, the applicable area could be a county or wildlife management area rather than a square grid cell. Pairwise status (infected:infected, infected:non-infected, or non-infected:non-infected) between each area
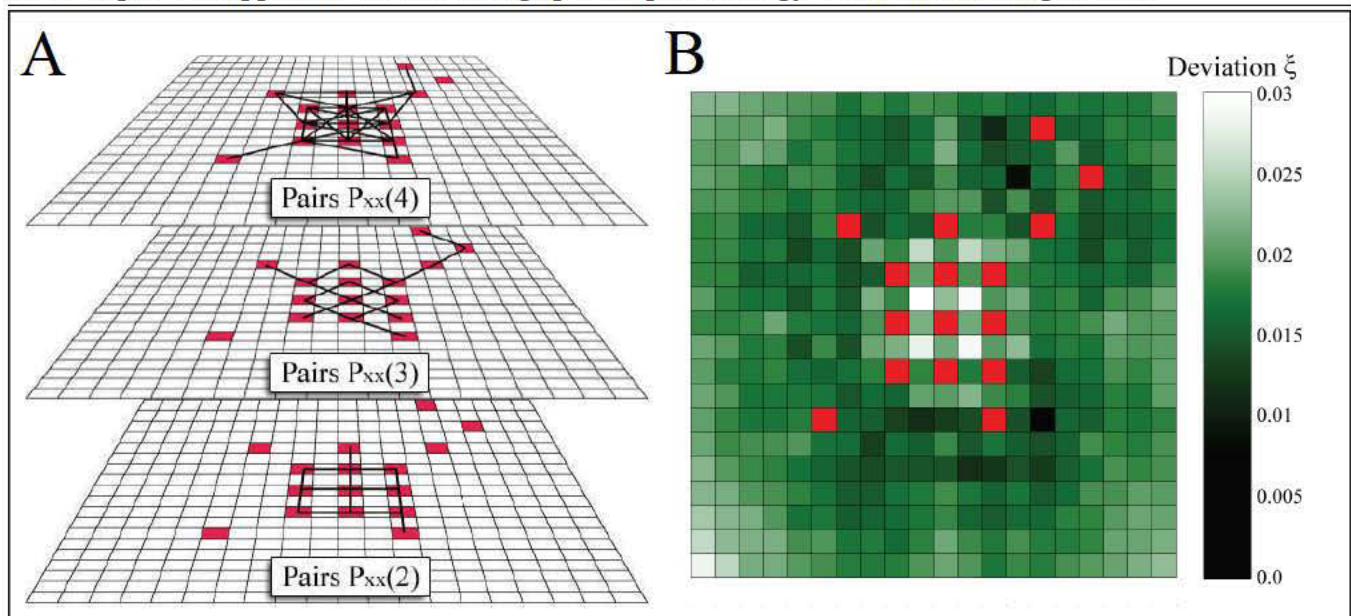
Figure 1.4. *(A) Pairwise status of infected:infected (P_{xx}) at distances of 2, 3, and 4 for a simple spatial distribution of positive CWD cases (red squares). Two cells joined by a black line represent a single pair at distance d. This pattern shows aggregation at intermediate distances and strong density dependence at d = 1 (because no red squares are adjacent to each other). (B) Likely sites of future CWD infection based on the spatial structure from (A) across all possible distances. Non-occupied sites are shaded (green) to indicate the deviation from the spatial structure in the observed snapshot. Dark squares represent the least change, and therefore represent sites that are most consistent with the observed spatial pattern. If the system is in statistical equilibrium (stationary prevalence), then these dark cells represent the most likely areas for positive cases of CWD. (Adapted from Keeling et al. 2004)*

or cell are calculated, along with the distances between areas. From the data, a function is fit to estimate how the risk of infection changes with distance from any positive cell. There are four main assumptions required for this method: 1) the spatial sample is a representative sample of the disease on the landscape, 2) the spatial area is homogeneous, 3) the spread is not limited by boundaries (e.g., roads, rivers, or other boundaries to animal movement), and 4) estimated parameters and mechanisms were constant during the formation of the spatial pattern. If these assumptions are met, the rate of change in the pairs and the probability of areas to become positive are identifiable (Figure 1.4).

Questions addressed / model predictions:

1. Estimates rate of spread as a function of distance from an infected area.
2. *Estimates the probability that an area will*

become infected.

Data required:

1. Disease status of spatial areas, such as wildlife management units (presence-absence).
2. Centroid coordinates of the spatial areas (distances are calculated between area centroids).
3. Meet the 4 assumptions listed above.

Output:

1. Estimates the probability that an uninfected area will become infected.
2. Estimates the rate of spread as a function of distance from an infected area.

General usefulness:

The snapshot model is probably most useful at very broad or very fine scales, where the spread of an infection may be independent of temporary barriers. That is, at very broad

scales, many 'local' barriers are not of great interest or importance to modeling probability of infection. At very fine scales, areas within or between barriers can be modeled. The main strength of the snapshot method is that data are required at only one point in time, and the data are relatively straightforward to collect. As a result, the snapshot method permits a rapid assessment of where the disease may spread to next.

Usefulness to CWD modeling and/or management:

The data required, disease status of an area, and its centroid, are easily determined from CWD surveillance data. However, some of the model assumptions may be difficult to meet. Surveillance sampling could be designed to meet the first assumption over large areas such as wildlife management units, but none of the remaining three model assumptions are realistic for CWD epidemiology. In particular, the patchy distribution of hosts across the landscape is likely to be a strong confounding factor in a snapshot analysis. Given the restrictive assumptions, the snapshot approach is unlikely to be useful for modeling and management of CWD. However, the idea of using area-based, presence-absence disease data to model the probability of disease in an area (even when disease is not detected) is explored in more detail below.

OCCUPANCY ANALYSIS:

 If presence-absence data are collected and modeled within a mark-resight framework (MacKenzie et al. 2002, MacKenzie et al. 2003), they can be used to estimate and monitor occupancy, colonization, or extinction probabilities of a wildlife species in a given area. In this case, occupancy is a proxy for abundance in large-scale monitoring studies. Data collection for occupancy estimation has the added advantage of requiring a less intensive field protocol and being potentially less costly as compared to methods for density or abundance estimation.

Recent theoretical advances in the development of occupancy models and their implementation have dramatically increased the viability of using this technique for landscape-scale modeling (McKenzie et al. 2002, McKenzie et al. 2005, Royle and Dorazio 2006. Freeman et al. 2007).

CWD data could be modeled using occupancy modeling. Probability of disease occupancy in an area, such as a wildlife management unit, could be estimated and modeled similarly to occupancy of a species in an area. Because CWD does not appear to spread rapidly, one could assume that over a short time frame, colonization and extinction probabilities are sufficiently low to be irrelevant (although these may be the parameters of interest for diseases with fast-moving epidemic fronts). If one controls for sampling intensity (sub-sampling could be used to do this) and prevalence is relatively constant over time, then occupancy probabilities could be used as a proxy for prevalence probabilities over the same area. The disadvantage of this method, however, is a great loss of resolution. That is, prevalence could drop dramatically, but this drop would not be detected by an occupancy model. Its main use would be for very large areas, such as multiple states, where it could be adapted for areas with sparse sampling and different sampling protocols.

Similar to occupancy for monitoring wildlife species, data collection would be far simpler and less costly to collect compared to data collected to generate precise prevalence estimates or a risk surface. This approach readily lends itself to a model selection framework, similar to the risk assessment approach. Hypothesized environmental and ecological risk factors could be evaluated with respect to the probability of disease occurrence. Occupancy models would be especially tractable for very broad scales, although it may be an important challenge to correctly identify the size of the areas sampled for disease "occupancy". The area should be large enough to be efficiently sampled, but small enough to capture the spa-

tial heterogeneity of CWD prevalence (e.g., if you sample a large enough area, all areas will be positive) and to reflect the epidemiology of the disease and the use of space by hosts.

Questions addressed / model predictions:
1. Over a very broad scale, identifies areas for which the disease is present or absent.
2. Identifies areas having a low probability of disease detection due to sampling schemes.
3. Identifies environmental/ecological (abiotic and biotic) factors associated with the probability that the disease is present.
4. Identifies environmental/ecological (abiotic and biotic) factors/covariates associated with the probability that the disease is detected

Data required:
1. Disease status (presence or absence) of an area such as wildlife management unit.
2. At least 2 years of data for each area.

Output:
1. Estimates probability of disease presence or detection in an area, such as a wildlife management unit or county.
2. Estimates probability surface of disease occupancy using centroid of a given area.
3. Estimates effect sizes (e.g., difference in disease presence between treatment and control areas, between species, etc.), and other relevant statistics for factors in model.
4. If spatial environmental/ecological (abiotic and biotic) factor/covariate data were collected, then estimates their importance to presence and detection.
5. Provides model selection statistics.

General usefulness:
If designed correctly, an occupancy approach is, potentially, a cost-efficient method to monitor status of disease presence over very broad spatial scales. Methods to determine required sample sizes for predicted disease detection probabilities and disease prevalence have been described (Samuel et al. 2006). The method appears viable for highly infectious, fast-spreading diseases, as well as non-infectious or slow-spreading diseases. Also, the binomial and multinomial mark-resight methods underlying the estimation of occupancy probabilities explicitly estimate and accounts for spatial covariance (McKenzie et al. 2005). The occupancy approach is attractive because it is simple compared to other methods that model spatial correlation.

Similar to risk assessment, occupancy models would be more applicable for diseases with short dormancy/latency periods than for those with long latency periods. Environmental or ecological factors associated with short latency period, would be more readily identified because longer latencies may dilute the effect as evaluated in future time periods unless these factors were relatively constant and migration rates relatively low (e.g., animal stays in the place where the risk occurs).

Usefulness to CWD modeling and/or management:
Although occupancy models have not been used for disease modeling, they appear potentially useful for initial modeling of CWD at regional or other broad scales. Because relatively little data is required to determine CWD 'occupancy' status (i.e., is there one or more infected animals in an area?), most states have comparable data. Differences in data collection protocols can make data incomparable and hamper prevalence estimation and use of other approaches (e.g., cluster analysis and risk analysis, based on risk surface). The main drawback of occupancy modeling is a reduction in the resolution of biological inferences. That is, the output is probability that the disease is present, but it could be present at a very low prevalence or very high prevalence. This reduced resolution in the output would dilute the ability to realistically evaluate spatial factors and covariates. Again, choosing the correct spatial area over which to estimate disease presence or absence would be critical to reduce the loss of resolution.

EPIDEMIC TREES – A NETWORK ANALYSIS (FOLLOWING HAYDON ET AL. 2002)

 Epidemic trees allow estimation of two primary values, R_0 and R_t. R_0 defines the average number of secondary cases that arise from a single case at the start of a disease outbreak or epidemic. R_t is the average number of secondary cases arising from a single infection during time = t. Traditional Susceptible-Infected-Recovered (SIR) models also estimate R_0 based on a theoretical model. The epidemic-tree approach is novel in that it is an empirical method of direct estimation of R_0 from the history of the observed cases (Haydon et al 2003). The epidemic tree approach is contingent upon data that accurately tracks the historical progression of the disease, specifically the temporal path from initial case to subsequent cases.

Questions addressed / model predictions:
1. Estimates disease transmission rate (R_0 and R_t) through time and space.
2. Evaluates, retrospectively, the effectiveness of different control strategies; i.e., estimates reduction of R_t for a control strategy instituted at a given location and time.
3. Allows retrospective comparisons, sensitivity, and cost-benefit analyses of different control measures, different timing of control actions (relative to onset of disease or relative to season), and different control locations.
4. Evaluates the influence of long-range transmission events compared to short-range transmission.
5. Identifies the best strategies available for a future outbreak or outbreak of similar disease.

Data required (* indicates data not currently collected as part of any CWD surveillance program):
1. Location of infected/infection.
2. Putative date of start of infection.*
3. Date a suspected infection reported.*
4. Date when infection confirmed.*

5. Viable rule set that ties subsequent cases to the case from which they originated.*

Output:
1. Estimates of R_t and its variance for specific time intervals and locations (estimates of rate of spread in time and space).
2. Estimates of generation time = interval between infection and subsequent case arising from it.
3. Estimates of reporting time = time between infection and subsequent reporting of case arising from it.
4. Estimates of routes of spread.
5. Estimates of rate of spread.

General usefulness:
Epidemic tree modeling is good for highly infectious, fast spreading epidemics. This method will underestimate R_t if all cases are not identified. However, as long as R_t is not biased by area (e.g., bias could arise if fewer infections identified away from urban areas due to different detection probability, whereas all areas would be equally under-represented a relatively non-biased situation), this method is usable for comparing the effect of different management strategies on R_t. Finally, for fast-spreading diseases that may re-invade, the strategies gleaned from retrospective analysis could be applied to future invasions.

Usefulness to CWD modeling and/or management:
The epidemic tree approach is not useful for modeling spatial epidemiology of CWD because positive animals cannot be linked to an originating case. Because CWD may be transmitted indirectly through a prion-contaminated environment, there may be no specific originating case or location. Also, it would be difficult to estimate a time of infection, since little is known about the course of disease in free-ranging deer, or about potential individual variation in disease progression. In addition, the timescale of CWD may be too long to take timely advantage of retrospective strategies.

BROAD SCALE FOCAL APPROACHES:

For the regional scale, we decided to present 2 focal approaches, micromaps and risk analysis. We present micromaps as a useful visualization technique to display data over large regions because they integrate spatial and temporal aspects of data. This technique would be the first step to summarize and examine regional data, while risk analysis is a logical second step. Risk analysis, at the regional scale, focuses on the two main questions: 1.) What are the significant CWD risk factors for a free-ranging mule deer population? and, 2.) Can we predict where CWD is likely found or will spread next? We focus on mule deer for all focal approaches, but these approaches could be applied to white-tailed deer and elk. Note: we do not use real data or perform a detailed analysis; our goal is simply to provide a general overview and illustrate the potential of these methods.

FOCAL APPROACH #1: LINKED MICROMAPS

Step #1- Generating Linked Micromap Plots:

After delineating game management unit (GMU) boundaries, linked micromap plots for the GMUs and the state of Colorado were created using the S-plus statistical software package. The sample S-plus code for creating linked micromap plots is available at Dan Carr's ftp site (ftp://galaxy.gmu.edu/pub/dcarr/newsletter/micromap/).

For CWD in northcentral Colorado, we simulated a sample that was limited to only a few GMUs. Therefore, micromap visualization can consist only of those GMUs with sampled deer. Prevalence and associated confidence intervals were subsequently plotted in a linked micromap representation.

Step #2- Interpreting the Final Linked Micromap Plots:

Figure 1.5 shows a series of linked micromap plots for the GMUs sampled for CWD. In the first linked micromap (1.5.A), four vertical panels (columns) are linked by geographic location, which is the GMU in this case. The map (panel 1) shows the boundaries for GMUs inside a CWD endemic area. The legend to the right of the map (panel 2) shows the GMUs designation with a dot in the linking color. The set of four graphs (panels 3 and 4) illustrate two statistical variables. In this particular example, dot-plots represent prevalence for the 2
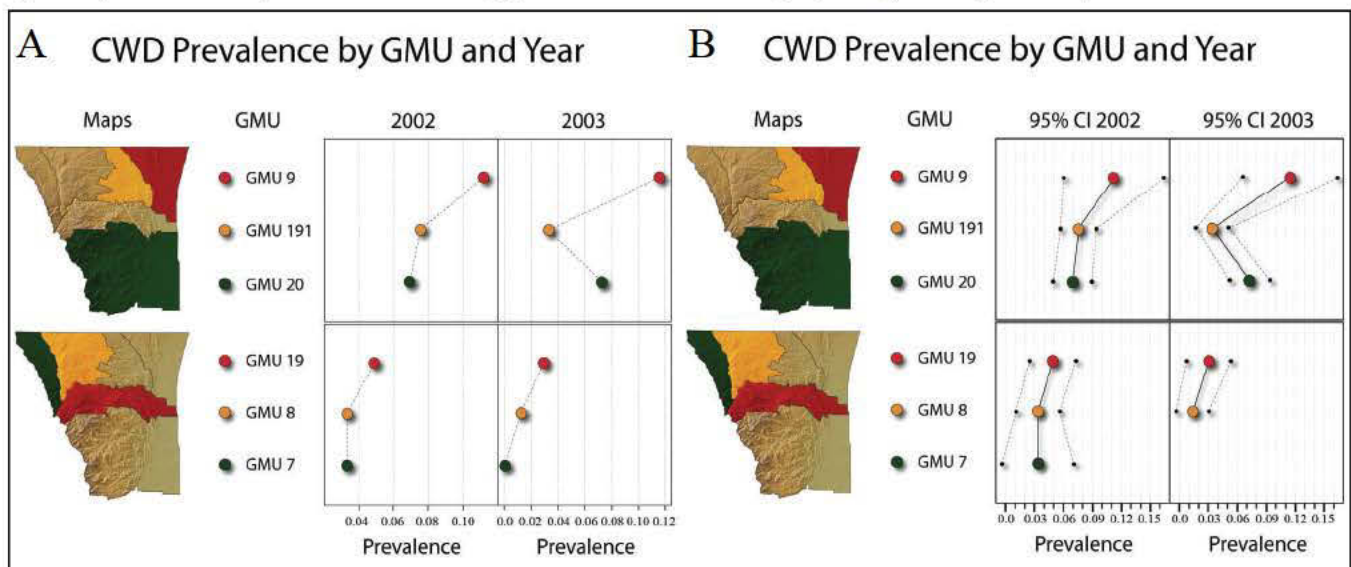
Figure 1.5. *Various micromaps for CWD from simulated mule deer data for northcentral Colorado. (A) CWD prevalence in 6 GMUs for 2002 and 2003. (B) An example of portraying prevalence across space (6 GMUs) and time (2 years) with 95% confidence intervals.* ***Note: this is not real data.*** *See text for more detailed explanation*

years in question, i.e., 2002 (panel 3) and 2003 (panel 4). All corresponding micromaps, labels, and statistical panels are linked by their colors. Note that three distinct colors are used to distinguish the GMUs within a particular micromap frame and are unique to that micromap. The GMUs are ranked according to simulated prevalence of 2002 from highest to lowest and are partitioned into two micromaps. The ranking is user-defined and could be 2003 if desired, while the partitioning and number of micromaps are based on the number of geographical units to be represented (see Symanzik and Carr (2007) for details). While offering no interpretation of the CWD data, it is immediately obvious which GMU had the highest and lowest prevalence in 2002. The GMUs in the lower micromap panel show a decline in prevalence from 2002 to 2003, but still maintain the same ranking. GMU9 was markedly static between years with the highest prevalence.

Step #3- Displaying and Interpreting Supplemental Statistical Information:

A further capability (i.e., supplementary statistical representation) of micromaps, displayed in Figure 1.5.B, is the addition of confidence intervals as a component of the prevalence panel. The confidence intervals (panels 3 and 4) represent the 95% lower and upper confidence limits. The larger colored dots refer to, as before, the prevalence in each GMU. One can now appreciate the fact that the prevalence of each GMU are not quite the "true" (actual) prevalence and that the confidence intervals describe uncertainties of the estimates. Moreover, readers can also observe that GMUs where simulated prevalence was estimated from limited data versus ample data as signified by the width of the confidence interval. As an example, consider how GMU9, which had the highest prevalence in both years, compares to GMU191. Upon initial examination of the prevalence information, it appears that GMU9 has a higher prevalence than GMU191. However, GMU9 has a wider confidence interval indicating that the prevalence for GMU9 is less reliable.

FOCAL APPROACH #2: RISK ANALYSIS USING PREVALENCE (OR ODDS RATIO) SURFACE

Step #1- Development of GIS Database:

Geographical risk modeling begins with the development of a GIS database that will be used to store, access, update, and model risk associated with a disease. Special attention should be placed on checking the quality, type, and spatial resolution of each data layer. Data quality is crucial to the modeling process and can have a significant impact on overall uncertainty. For example, the layer with the coarsest spatial resolution typically constrains the spatial resolution of the analysis to no less than that particular scale. For multi-state modeling, data for each risk factor must be standardized across the entire region. Because data often come from different sources, standardization may require reprocessing data to a common coordinate system or reclassifying categorical variables. This is an important process and maps of an individual risk factor can generate useful insights. These maps can help identify influential variables and/or areas that are problematic to overall risk analysis.

Examples of Potential Risk Factors:
- Proximity to known CWD positive deer and/or elk based on surveillance locations.
- Proximity to captive deer and/or elk facilities (e.g., farms or hunting ranches).
- Proximity to deer and/or elk feeding or wintering areas.
- Proximity to deer and/or elk processing facilities.
- Proximity to deer and/or elk research facilities.
- Proximity to taxidermy operations.
- Historical intensity of sheep grazing and current sheep density.
- Black-faced sheep density.
- Proximity to known scrapie infected or exposed sheep flocks.
- Deer and/or elk movement corridors.

Step #2 – Transforming the Data:

Variables usually require transformation to

a common format. GIS data can be stored as points, lines, polygons, or in raster format (2 dimensional arrays where each cell, or pixel, contains a single values). Raster data is well suited to modeling exercises such as risk analysis (Longley et al 2002). However, "rasterizing" data results in homogeneity within pixels (i.e., each cell is assigned a single value); therefore it is important to consider potential interactions between resolution and the nature of the data being rasterized.

In the case of risk analysis, the ultimate goal is to generate a value that represents the risk in any given cell as a function of the risk variables. This often requires restructuring the data into more biologically meaningful terms. For example, captive cervid facilities are a known potential source of CWD exposure risk. Locations of these facilities are usually provided in the form of polygons or points, and each location will be represented as either present or absent when transformed into a raster format. For each cell containing a facility, the entire cell is assigned a 'present' value, regardless of the size of the facility relative to the cell (e.g., the facility might be 5 acres, but the cell size could be 5x5 miles, or 16,000 acres – the entire 16,000 acres would be considered occupied by the facility). The real value of interest is how far any particular cell is from the nearest game farm, and the estimated distance may depend substantially on the process used to standardize the underlying data. In cases where distance is the parameter of interest, the surface data for that variable is expressed in terms of distance (Figure 1.6).

Step #3 – Assigning Risk:

Risk is assigned to individual cells for each risk layer. The output is the result of applying a "risk" equation to each pixel, where each layer is a variable and the relative risk factors are coefficients between 0 and 1. The final map illustrates results from the equation. Figure 1.7 shows a visualization for two variables which are considered to be important factors in the spread of CWD: proximity to existing cases of CWD and proximity to captive cervid facili-
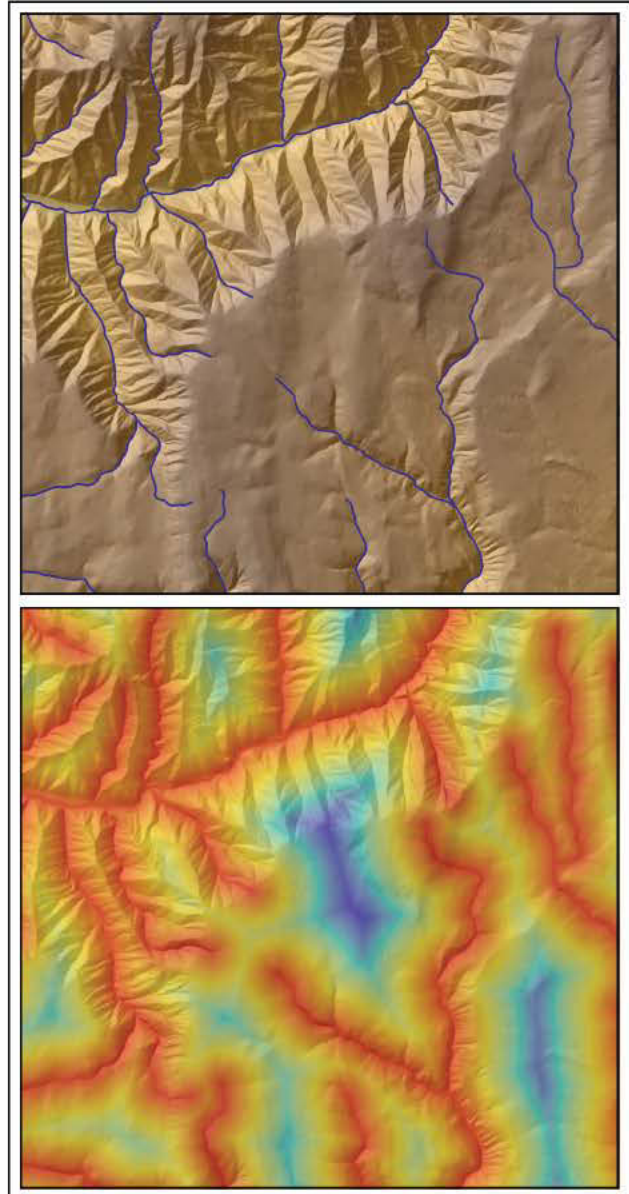
Figure 1.6. Transformation of rivers (top) from a linear feature to a raster data set (bottom). Each cell in the raster contains a value of the linear distance to the nearest river. The distance values increase as the color changes from red to yellow to blue.

ties. Note that for a real data set, risk would be assigned to the predictor variables based on their importance in a spatially corrected regression analysis. Techniques such as hierarchical partitioning (Chevan and Sutherland 1991), partial correlations (Cox 1985), and other measures such as variable importance in projection (Birkner and van der Laan 2006) can be used to determine variable weighting. Once the variables are ranked, the risk model can be

used to forecast risk for areas where disease data have not been collected. Where data on a predictor variable are not available, expert judgment may be used to assign relative variable importance, with a corresponding reduction in confidence in the forecasts.
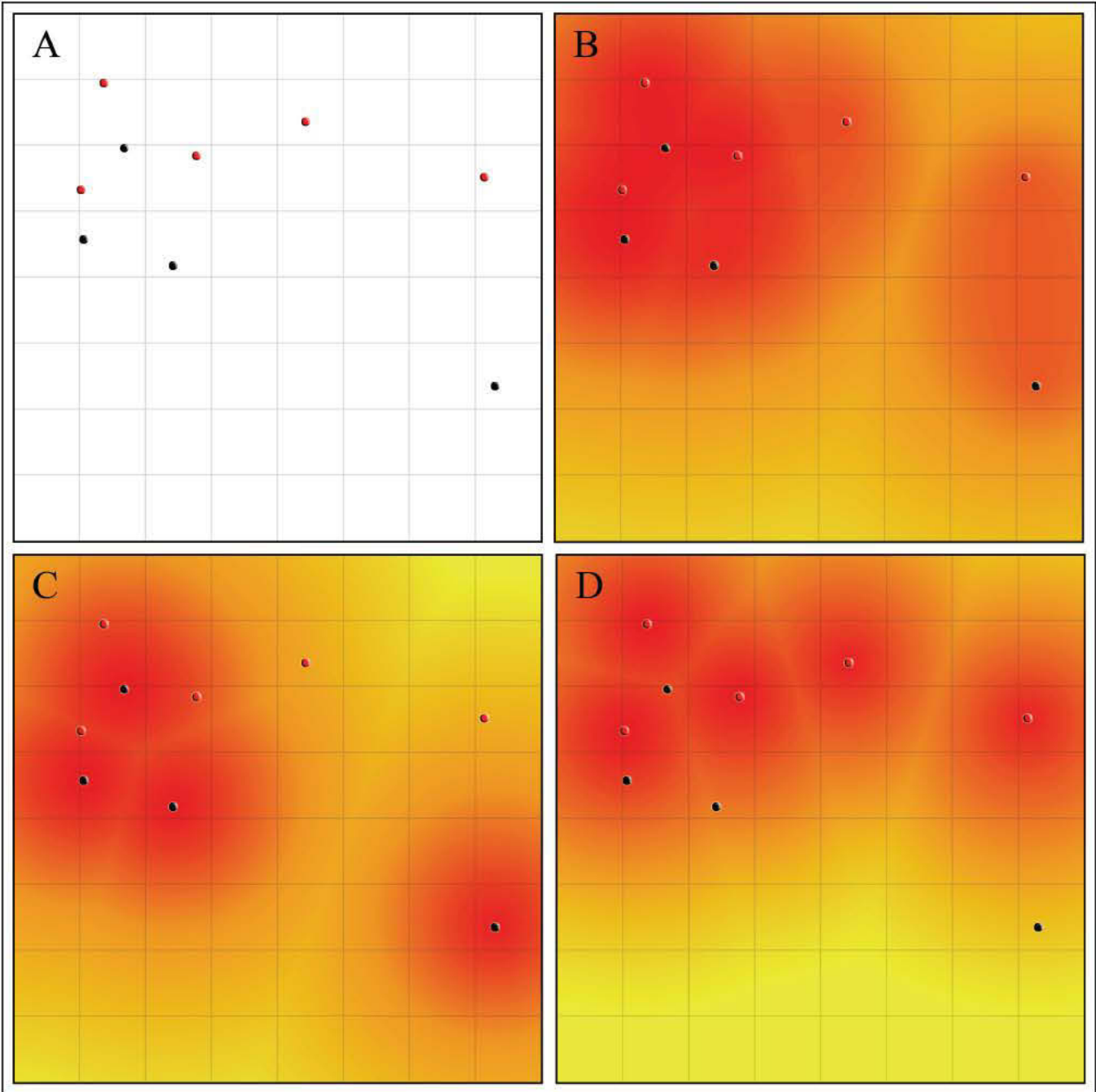
Figure 1.7. *(A) Hypothetical locations of two chronic wasting disease risk factors, positive chronic wasting disease cases (red points) and captive cervid facilities (black points). Risk surfaces when (B) risk is weighted equally between proximity to chronic wasting disease positives and proximity to captive cervid facilities, (C) risk from proximity to captive cervid facilities is 10x that of the risk from proximity to positive cases, and (D) risk from proximity to positive cases is 10x that of the risk from proximity to captive cervid facilities. In (B-D) red represents a relatively high level of risk and yellow represents a relatively low risk. The black grid represents large scale quadrats for visualization, while the surface is modeled at a much finer resolution. The surfaces in this figure were built from a 4000x4000 cell lattice.*

To highlight how important assigning relative risk weight to variables is to the overall risk analysis, we have provided three variations. In the first case, we assign equal risk weights to the proximity to existing cases of CWD (weight = 0.5) and proximity to captive cervid facilities (weight = 0.5). In the second case we assign 10 times the weight to the risk of proximity to captive cervid facilities (weight = 0.91) compared to proximity to existing CWD positives (weight = 0.09). In the final case, we reverse the weighting and assign 10 times the weight to the risk of proximity to existing CWD positives (weight = 0.91) compared to proximity to captive cervid facilities (weight = 0.09).

Step #4 – Displaying Risk Output:

Step 4 involves delimiting areas of relatively high and low risk based on the output layer. Typically isopleths of predetermined levels are shown in different colors allowing for a visual representation of the risk involved (See Figure 1.7). However, threshold levels can also be used to show areas above a set level of risk.

Step #5 – Model Validation:

Model validation is an important part of the risk assessment process. Both internal validation and external model validation should be assessed, especially if management decisions are based on risk models. Internal (e.g., bootstrapping) is the more straight forward procedure as the data already exist. This can be considered a self-consistency check, as any systematic differences between the simulations and the data (upon which the model is based) indicate weaknesses in the model (Gelman et al. 2004). Internal validations tend to be overly optimistic about model performance because the data for modeling and validation come from the same data set and they are thus not independent. External validation provides a more reliable estimate of sensitivity (proportion of false negatives) and specificity (proportion of false positives). External validation compares the fit of model predictions to new data (Gelman et al. 2004). For example, if CWD can be considered stationary and surveillance patterns/intensity are consistent across time, then it may be possible to use temporal external validation, such as building the model from one year of surveillance data and testing it on data from subsequent years. See Gelman et al. (2004) for a detailed explanation of model validation.

GENERAL CONCLUSIONS ABOUT REGIONAL-SCALE APPROACHES FOR CWD MODELING

Although contagious, CWD appears to fall somewhere between the class of slowly spreading diseases, such as rabies, and non-infectious chronic diseases, such as cancer or a pollution/toxin induced illness. CWD is similar to TB in that it spreads among individuals in a wildlife population by direct or indirect contact, and has a long latency period. Existing epidemiological models applied at a regional scale tend to fall into 2 categories: those more useful in representing infectious, relatively quick-spreading diseases or those more useful for non-contagious, spatially static diseases. We know little about the spread of CWD in the wild, and recently detected foci may result from increased surveillance sampling rather than spread. Consequently, methods for non-infectious diseases, such as cluster analysis and risk analysis, seem the most appropriate for spatial modeling and portrayal of CWD at a regional scale.

Thus, we chose risk analysis as the focal approach for large-scale modeling. Cluster analysis is a similarly useful method, and we describe use of cluster analysis as the focal approach at the landscape level in the next section. In general, we recommend micromaps as a first step for describing CWD and communicating patterns at broad scales. For many purposes, the next step is likely to be a risk analysis. The utility of snapshot or occupancy approaches for analyzing CWD spatial epidemiology is uncertain, and these approaches require further testing with CWD or similar diseases. We recommend that simulations (e.g., determining power to detect a specific change

in prevalence given different sample sizes) be conduced to evaluate their viability for CWD data at the regional scale. Finally, we note that epidemic trees are inappropriate for evaluating CWD spread because these were developed for analysis of highly infectious, fast spreading diseases.

The data gaps for these methods are primarily large spatial environmental and ecological data. These data often exist, but usually they need to be compiled and standardized across jurisdictional boundaries, for the region of the analysis. As noted in the focal approach, the quality, type, and spatial resolution of each data layer needs to be evaluated as part of the modeling process. The development of reliable, well-documented, spatial GIS-based layers of relevant biological and ecological factors will strongly promote CWD modeling efforts at the regional scale.

LITERATURE CITED:

Allan, B. F., F. Keesing, and R. S. Ostfeld. 2003. Effect of forest fragmentation on Lyme disease risk. Conservation Biology 17:267-272.

Anderson, D. P., J. D. Forester, M. G. Turner, J. L. Frair, E. H. Merrill, D. Fortin, J. S. Mao, and M. S. Boyce. 2006. Factors influencing female home range sizes in elk (*Cervus elaphus*) in North American landscapes. Landscape Ecology 20:257-271.

Birkner, M. D., and M. J. van der Laan. 2006. Application of a Variable Importance Measure Method," The International Journal of Biostatistics. 2: Article 6.

Chevan, A., and M. Sutherland. 1991. Hierarchical partitioning. The American Statistician. 45:90-96.

Cox, L. A. Jr. 1985. A new measure of attributable risk for public health applications. Management Science 31:800-813.

Briggs, D. J. 2000. Exposure assessment. Pages 335-359 *In* Spatial Epidemiology. P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Oxford University Press, Oxford, England.

Carr, D. B., and S. M. Pierson. 1996. Emphasizing statistical summaries and showing spatial context with micromaps. Statistical Computing and Statistical Graphics Newsletter. 7:16-23.

Carr D. B., A. R. Olsen, J. P. Courbois, S. M. Pierson , and D. A. Carr. 1998. Linked Micromap Plots:

Named and Described. Statistical Computing and Statistical Graphics Newsletter 9: 24-32.

Carr D. B. 2001. Designing Linked Micromap Plots for States with Many Counties. Statistics in Medicine 20:1331-1339.

Dent, B. D. 1993. Cartography: Thematic Map Design (Third Edition). William C. Brown, Dubuque, Iowa, USA.

Dunford, W., and K. Freemark. 2005. Matrix matters: Effects of surrounding land uses on forest birds near Ottawa, Canada. Landscape Ecology 20:497-511.

Freeman, S. N., D. G. Noble, S. E. Newson, and S. R. Baillie. 2007. Modeling population changes using data from different surveys: the common birds census and the breeding bird survey. Bird Study 54:61-72.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2004. Bayesian Data Analysis: Second Edition. CRC Press, Boca Raton, Florida, USA.

Harris, R. L. 1999. Information Graphics – A Comprehensive Illustrated Reference. Oxford University Press, New York, New York, USA.

Haydon, D. T., M. Chase-Topping, D. J. Shaw, L. Matthews, J. K. Friar, J. Wilesmith, and M. E. J. Woolhouse. 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proceedings of the Royal Society of London B 270:121-127.

Hess, G. R., S. E. Randolph, P. Arneberg, C. Chemimi, C. Furlanello, J. Harwood, M. G. Roberts, and J. Swinton. 2002. Spatial aspects of disease dynamics. Pages 102-118 *In* The ecology of wildlife diseases. P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. Oxford University Press, Oxford, England.

Hickling, G. J. 2002. Dynamics of bovine tuberculosis in wild white-tailed deer in Michigan. Michigan Department of Natural Resources Wildlife Division Report No. 3363. Lansing, Michigan, USA. Available at: http://www.michigandnr com/publications/ pdfs/huntingwildlifehabitat/reports/ deer_TB_dynamics.pdf (accessed December 2006).

Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No. 3.

Jenkins, S. R., and W. G. Winkler. 1987. Descriptive epidemiology from an epizootic of raccoon rabies in the Middle Atlantic States, 1982-1983. American Journal of Epidemiology 126:429-437.

Johnson, C. J., K. E. Phillips, P. T. Schramm, D. McKenzie, J. M. Aiken, and J. A. Pedersen. 2006. Prions adhere to soil minerals and remain infectious. Plos Pathogens 2:296-302.

Kallèn, A., P. Arcuri, and J. D. Murry. 1985. A simple model for the spatial spread and control of rabies. Journal of Theoretical Biology 116:377-393.

Keeling, M. J., S. P. Brooks, and C. A. Gilligan. 2004. Using conservation of a pattern to estimate spatial

parameters from a single snapshot. Proceedings of the National Academy of Sciences 101:9155-9160.

Kitron, U. 2000. Risk maps: mapping transmission and burden of vector-borne diseases. Parasitology Today 16:324-325.

Kulldorff, M., and N. Nagarwalla. 1995. Spatial disease clusters: detection and inference. Statistics in Medicine 14:799-810.

Kulldorff, M. 1997 A spatial scan statistic. Communications in Statistics: Theory and Methods 26:1481-1496.

Longley, P.A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2002. Geographic information systems and science. 3rd edition. John Wiley & Sons, Chichester, England.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and. A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200-2207.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2005. Occupancy estimation and modeling. Academic Press, Burlington, Massachusetts, USA.

Merler, S., C. Furlanello, C. Chemini, and G. Nicolini. 1996. Classification tree methods for analysis of mesoscale distribution of *Ixodex ricinus* in Trentino, Italian Alps. Journal of Medical Entomology 33:888-893.

Moore, D. A. 1999. Spatial diffusion of raccoon rabies in Pennsylvania, USA. Preventative Veterinary Medicine 40:19-32.

Morris, S. E., and J. C. Wakefield. 2000. Assessment of disease risk in relation to a pre-specified source. Pages 153-184 *In* Spatial Epidemiology. P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Oxford University Press, Oxford, England.

Murwira, A., and A. K. Skidmore. 2005. The response of elephants to the spatial heterogeneity of vegetation in a Southern African agricultural landscape. Landscape Ecology 20:217-234.

Prins, H. H. T., and F. J. Weyerhauser. 1987. Epidemics in populations of wild ruminants – anthrax and impala, rinderpest and buffalo in Lake Manyora National Park, Tanzania. Oikos 49:28-38.

Rogers, D., S. I. Hay, and M. J. Packer. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Annals of Tropical Medicine and Parasitology 90:225-241.

Royle, J. A. and R. M. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. Journal of Agricultural Biological and Environmental Statistics 11:249-263.

Samuel, M. D., D. O. Joly, M. A. Wild, S. D. Wright, D. L. Otis, R. W. Werge, and M. W. Miller. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk. USGS-National Wildlife Heath Center Report. Madison, Wisconsin, USA. Available at: http://www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf .

SAS Institute Inc., SAS OnlineDoc® 9.1.3, Cary, North Carolina, USA: SAS Institute Inc., 2002-2005.

Schauber, E. M., R. S. Ostfeld, and A. S. Evans Jr. 2005. What is the best predictor of annual Lyme disease incidence: weather, mice, or acorns. Ecological Applications 15:575-586.

Schmitt, S. M., S. D. Fitzgerald, T. M. Cooley, D. S. Bruning-Fann, L. Sullivan, D. Berry, T. Carlson, R. B. Minnis, J. B. Payeur, and J. Sikarskie. 1997. Bovine tuberculosis in free-ranging white-tailed deer from Michigan. Journal of Wildlife Diseases 33:749-758.

Sleeman, J. M. 2005. Disease risk assessment in African great apes using geographic information systems. EcoHealth 2:222-227.

Smith, G. C., and S. Harris. 1991. Rabies in urban foxes (*Vulpes vulpes*) in Britain: the use of a spatial stochastic simulation model to examine the pattern of spread and evaluate the efficacy of different control regimes. Proceedings of the Royal Society of London B 334:459-479.

Symanzik, J. and Carr, D. B. 2007. Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data. In, Handbook of Data Visualization. Chen, Chun-houh; Härdle, Wolfgang; Unwin, Anthony (Eds.). Springer. ISBN-10:3-540-33036-4.

Appraoaches to Landscape-Scale Modeling
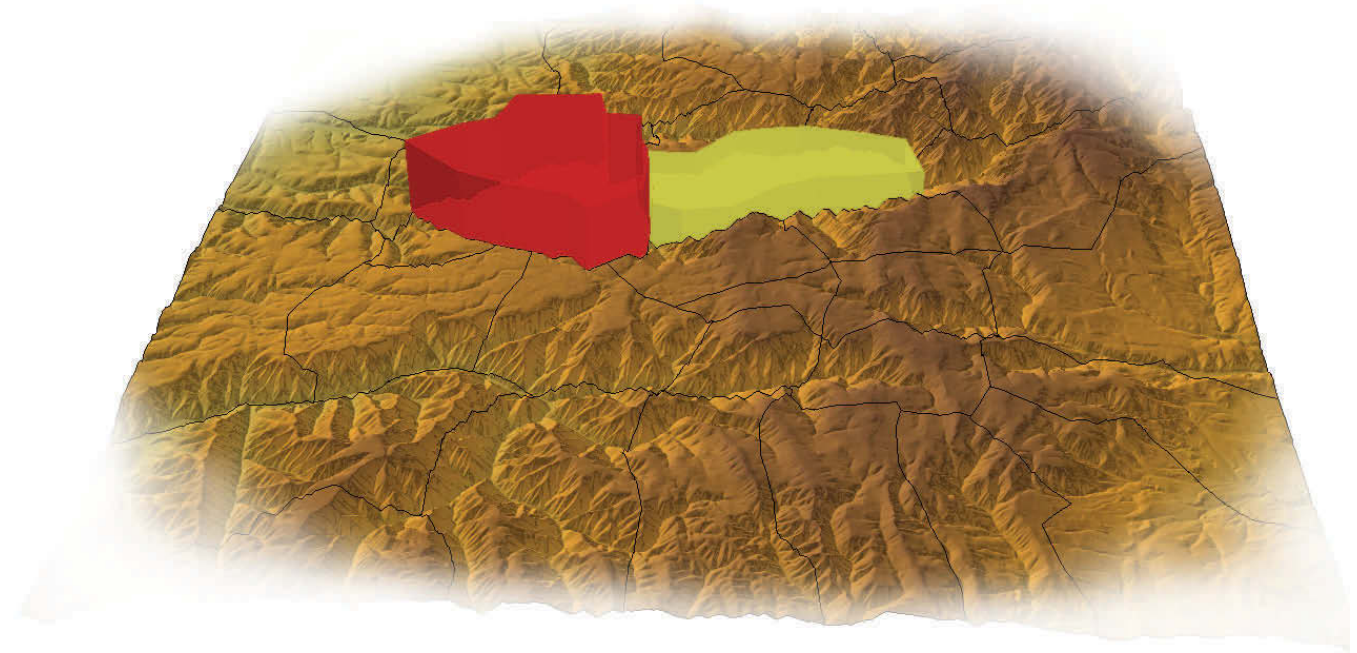
A hypothetical landscape with agency management units (e.g., county, wildlife management unit), over which data is frequently modeled or summarized at the landscape level. The red and yellow units could represent mean prevalence, animal density, or some other feature of the landscape.

The primary goals of landscape-scale modeling are identifying areas of high-risk for spread, as well as general patterns of disease spread. Another modeling goal is to link the spatial structure of wildlife populations and the spatial variability in abiotic and biotic attributes of their environment with disease transmission dynamics. Large-scale modeling often involves averaging over large areas to depict and analyze the patterns and spread of a disease. An inherent shortcoming of averaging is the loss of understanding of some biological processes, such as animal movements, on disease epidemiology. The term landscape epidemiology illustrates the concept by mapping a landscape in terms of spatial risk factors for infection and disease prevalence (Hess et al. 2002).

Features of the landscape and host that may affect disease distribution and transmission can be georeferenced and mapped. In addition to spatial variation, many factors vary in a temporal fashion, such as seasonally or annually. Both spatial and temporal data at the landscape scale can be useful in making predictions based on past conditions, and can be updated as conditions change or new information becomes available.

Statistical approaches seek correlations between environmental conditions and the distribution of disease, while mechanistic approaches attempt to identify biological processes that drive the observed patterns (Lawson 2001). The observed patchiness of a wildlife disease on a landscape could be the product of environmental factors that enhance the exis-

tence or transmission of the disease, or the distribution and movement patterns of hosts or vectors of the disease. At a landscape level, however, other factors that influence this dynamic relationship may emerge, such as effects of management intervention or localized human induced risk areas. Models at a landscape scale may allow us to tease out how these factors interact with the disease to produce the observed patterns.

CLUSTER ANALYSIS

Cluster analysis is a valuable approach for identifying disease patterns at multiple scales, but is especially useful for describing general point patterns across a landscape (Wakefield et al. 2000). From a decision-making standpoint, clustering at the landscape level can help to identify areas where disease incidence is higher than might

be expected by chance alone. In the regional-level section, we presented cluster analysis for multi- state or providence scales where data are likely to have course resolution, or visualization at large extents (i.e., multiple states or providences) requires a reduction in data accuracy and detail. At a regional scale, cluster analysis uses data that are either missing location information for a proportion of the data, or data are summarized into polygon count or prevalence data. Thus, prevalence, or counts of positive and negative cases, could be summarized over areas such as wildlife management units. Here we consider cluster analysis for data with higher spatial resolution, such as data sets where specific location information (e.g., UTM coordinates) are available for most of the samples. For these data, the assignment of a case to a cluster is based primarily on its coordinates. Specifically, CWD positive cases are placed in clusters based on their relative distance to other positive cases while account-
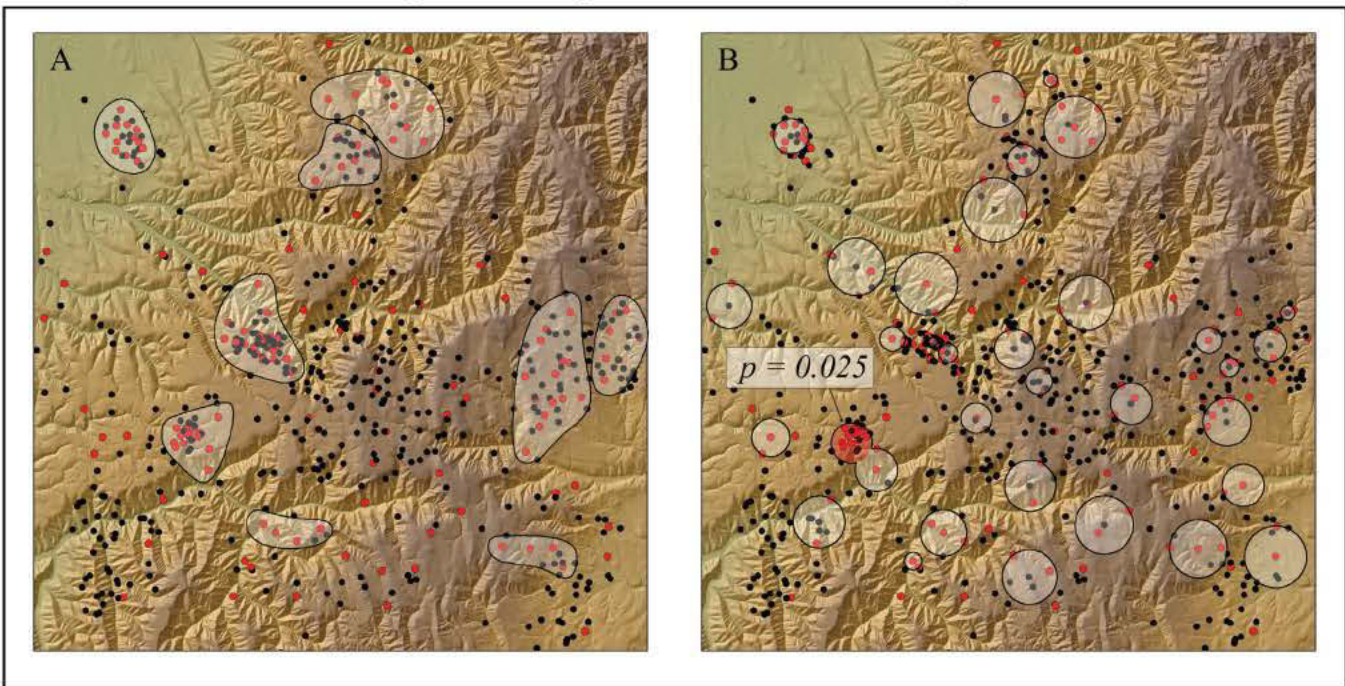
Figure 2.1. Simulated data illustrating the concept of residual spatial variation. Red circles represent positive cases and black circles represent negative cases. (A) A density based clustering algorithm that does not take into account the negative cases, but is well suited to detecting irregular shapes. (B) A spatial scan clustering algorithm which accounts for the underlying structure of the population but uses a circular (or elliptical) window to find clusters. The scan method (B) identifies many possible clusters (white-shaded circles), but only identifies one as statistically significant (red-shaded circle, p = 0.025). The cluster in the upper left was selected as the second most likely cluster, but was not significant.

ing for distances between negative cases. Although we discuss cluster analysis at the landscape scale, it could also be used at the regional scale if most samples have spatial coordinates.

In the context of spatial epidemiology, the definition of a cluster takes on rather specific meaning:

"A spatial aggregation of cases relative to the underlying distribution [locations] of noncases, with a unique structure which differentiates it from other clusters."

or

"the residual spatial variation in locations of disease (Diggle 2000)."

The important point is that both definitions explicitly take into account the spatial heterogeneity of cases relative to non-cases. This idea is particularly relevant to the study of CWD, where hosts or infectious agent may be patchily distributed in space, and data comes from a variety of sources (e.g., agency culling, hunting, captive cervid operations, etc) that also can be patchily distributed. Simple scatterplots of locations of CWD samples clearly show heterogeneity in the density of samples at multiple scales (Farnsworth et al. 2006). Thus, a simple clustering of positive cases (ignoring non-cases) can be misleading as apparent disease clusters may be explained simply by a clustering in sample collection density and represent sampling effort rather than disease clustering (Figure 2.1) (Kulldorff and Nagarwalla 1995). Accordingly, the inference of disease "hot-spots" via cluster analysis should only be considered when they represent, in Diggle's (2000) words, "the residual spatial variation" of positive and negative samples.

One of the most widely used cluster approaches for epidemiological data that accounts for the underlying density of non-cases is the spatial scan statistic. Developed by Kulldorff (1997), it identifies a significant excess
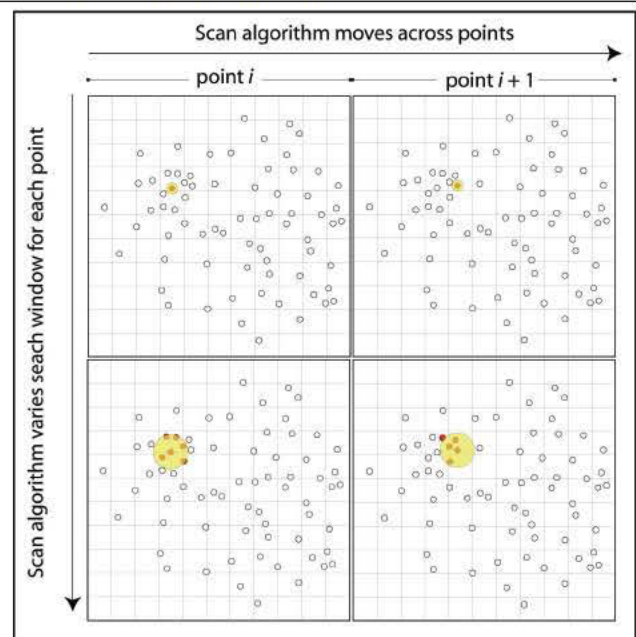
Figure 2.2. Shows an example of how the scan statistic method works. At each data point the algorithm samples the number of members inside a set of circle ranging from 0 to a user specified maximum limit (either an actual radius for the circle or a maximum proportion of the population). The algorithm then moves on to sample the next point in the dataset and continues this process until all points have been sampled.

of cases within a moving window that visits all spatial locations, increasing in size at each location until it reaches a predetermined upper size limit (Figure 2.2). The scan statistic provides a measure of how unlikely it would be to encounter the observed excess of cases in a larger comparison region. A more detailed description of the scan statistic is covered in the focal method of this section. Several variations of the scan statistic method have been applied to a wide variety of epidemiological studies for cluster detection including non-Hodgkin's lymphoma, child mortality, and bovine TB (Perez et al. 2002, Sankoh et al. 2001, Tango and Takahashi 2005, Veil et al. 2000).

Clark and Larson (2002) group clusters into three classes: spatial clusters, temporal clusters, and spatio-temporal clusters. Purely spatial clusters represent risk that remains constant over time. Conversely, temporal clusters exist

across entire regions, but only for relatively brief time periods. Spatio-temporal clusters imply a temporary localized region of elevated disease within a larger space-time context. For example, spatio-temporal clustering methods were used to study outbreaks of acute respiratory disease in cattle herds in Norway (Norstrom et al. 2000).

For highly transmissible and fast moving infectious agents, the incorporation of time is often wise. In the case of the CWD, this is unlikely to be useful due its relatively slow rate of spread. The practical aspects of sample collection also influence the ability to incorporate time into a cluster analysis. Sampling often occurs in periodic "bursts" of relative short periods, such as focal ("hot spot") culling over days, or as annual sustained events such as hunting seasons, which occur over days or weeks. Accordingly, we suggest that addition of time to a cluster analysis for CWD be restricted to annual periods. If little spread is suspected, data from multiple years can be combined for cluster analysis. To do this, data within years (and/or between years) are combined onto a map and considered only as spatial clusters. Attempts to assess changes in prevalence across years should be done carefully, as spatial patterns of culling and hunting may not be consistent across years.

Questions addressed / model predictions:
1. Identifies localized disease high prevalence areas or "hot spots".
2. Identifies spatial disease patterns at different scales.

Data required:
1. Spatial coordinates of positive and negative samples if disease cases do not 'present' themselves – such as for CWD cases.
2. Random and representative samples.
3. Covariates (if applicable).

Output:
1. Assigns positive cases to a particular cluster.
2. Identifies potential spatial covariates to disease pattern.
3. Identifies high-risk areas and cluster maps.

General usefulness:
The main strength of cluster analysis is that it can be used to depict and describe spatial patterns of a disease and identify hot spots of high disease prevalence (at which management intervention could be targeted), as well as to estimate a risk surface. Cluster analyses are exploratory, but can be useful for hypothesis generation. For example, if a risk surface is generated, models with environmental/ecological (abiotic and biotic) covariates, genetic covariates, or other factors suspected to be related to the disease, can be evaluated. Cluster analysis is a valuable initial step in examining the spatial epidemiology of a disease.

Usefulness to CWD modeling and/or management:
Cluster analysis is a valuable descriptive tool for CWD surveillance data. At the landscape scale, we recommend that cluster analysis be conducted as a first step for georeferenced CWD surveillance data. The usefulness of cluster methods for CWD is the same as described above in "General usefulness". We considered the special case of cluster analysis, the use of kernel density estimators, to generate a risk surface, in the regional-level section on Risk Analysis/Assessment.

GEOSTATISTICAL ANALYSIS

Geostatistics are statistics pertaining to the earth, or statistical techniques that emphasize locations with an areal (spatial) distribution. Geostatistics is usually concerned with statistical theory and applications for spatial processes which have a continuous (or nearly continuous) spatial index. In traditional statistical analysis we assume that observations are taken under identical conditions, and independently from one observation to another. However, with spatial

data, the location of each observation gives rise to spatial dependence and heterogeneity (Cressie 1993). Spatial autocorrelation measures the degree of this statistical association between data units for different distances, and if autocorrelation occurs it becomes important to describe the underlying spatial process (Diggle et al. 2003). For diseases, like CWD, positive spatial autocorrelation usually indicates areas of either high or low risk. Negative autocorrelation seems unlikely as it implies areas of higher disease are adjacent to areas of low disease. Such negative correlation patterns might imply landscape boundaries to disease spread.

Geostatistical methods provide an important approach for modeling and correcting the spatial autocorrelation that commonly occurs in disease data. The overall goal of a geostatistical analysis is to assess factors that may be associated with the occurrence of disease, while accounting for spatial dependency. For example, explanatory variables can be incorporated in the analysis to account for factors related to disease risk (males vs. females), disease exposure (age), or type of disease transmission (density-dependent vs. frequency-dependent transmission), as well as to evaluate other factors that might affect disease patterns (animal movement and habitat patterns). These spatial regression models take into account both the importance of dependent variables and spatial dependence. An advantage of the geostatistical approach is that it recognizes both larger scale spatial trends and local spatial correlations. Most CWD affected areas will have substantial small-scale variation, typically exhibiting strong positive correlation between data from nearby spatial locations.

The use of geostatistical methods requires observations of a response variable (0, 1 for CWD infection status) for individual animals or summary of prevalence in a small area, and the spatial locations of these responses. Ideally, locations should be truly continuous in space, but small scale clustering (summary) of data would not likely cause major violations of

this requirement. A geostatistical regression model looks generally as follows: $Z = X\beta + \delta(d)$, with fixed effect dependent variables X modified by β, and $\delta(d)$ is a zero-mean error vector that is spatially correlated (a function of distance) according the model selected through a geostatistical analysis. In the case of disease, the response variable (Z) is typically discrete data (binary for disease status or Poisson for counts of infected animals). In particular, linear models may not always perform well and discrete models may be more appropriate for disease data. Some useful process models (commonly called link functions) include the binary (or logistic) model for CWD status (susceptible or infected), the Poisson for CWD prevalence data using counts of infected and susceptible animals, complementary log-log models for estimation of disease prevalence using age-prevalence data, and more complex hazard rate models (Heisey et al. 2006).

In geostatistical analysis, spatial variance is modeled using a parametric pattern (Cressie 1993) that best describes the spatial correlation (dependency) in the data. This variance can be considered to have both distance and directional properties, and it is evaluated using the variogram (or semi-variogram = variogram/2). The variogram is the cornerstone of geostatistical analysis, and is treated as a random process (variable). The variogram is used to asses the degree of variance between spatial locations as a function of the distance between locations (Figure 2.3).

The basic structural components of the variogram (Figure 2.4) include the parametric model's underlying spatial dependency and the estimated model parameters (nugget, sill, and range). The nugget measures discontinuity at the variogram origin (h=0 or minimum variogram lag distance). Theoretically, we expect the nugget to have a value of 0 because points are perfectly correlated with themselves. Actual data often show differences between data collected at very close locations in space. This variation usually results from random variation at scales below the minimum lag dis-
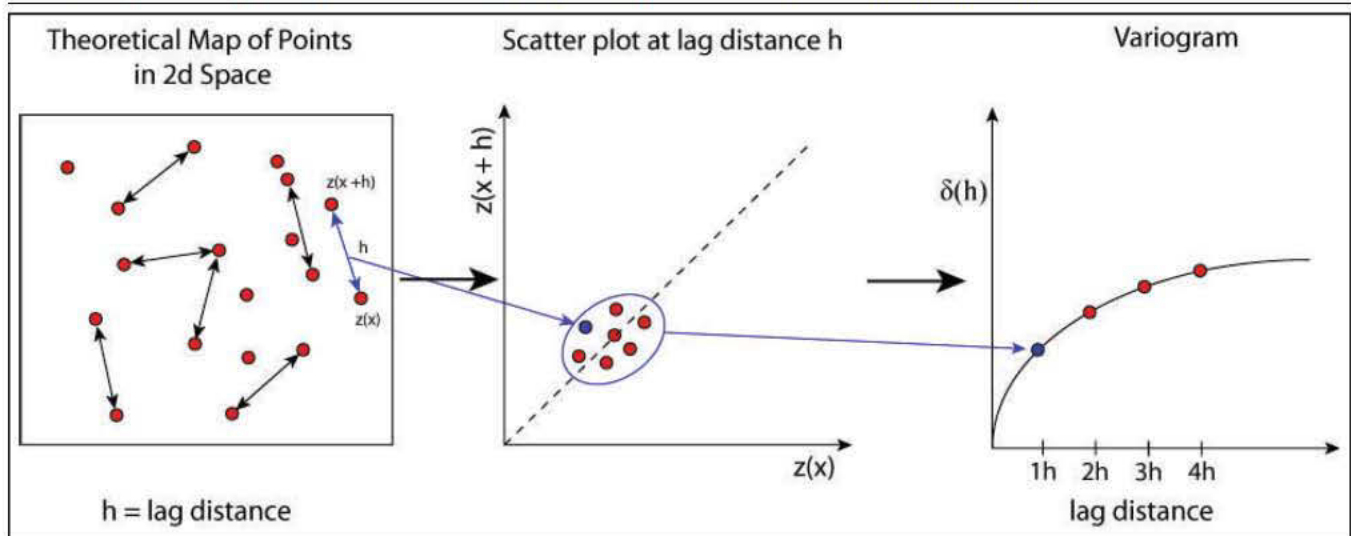
Figure 2.3. *Calculation of the variogram. All pairs of points separated by lag distance h are identified (left). For each pair of points the difference between the values of the points is calculated (middle). For each lag distance the squared differences are summed and divided by twice the number of points for that lag distance (right). The variogram is created by repeating this process at different lag distances (e.g. 1h, 2h, 3h, etc.). Adapted from Shibli 1997.*

tance (microscale variation) or from measurement error. Thus the nugget is an estimate of sampling plus microscale variation. Variogram patterns (i.e., variance) generally increase with lag distances because close points tend to be more similar than those located farther apart. The lag distance at which the variogram values stabilize and the distance beyond which data appear to be independent is called the range. The final structural component, the sill, repre-

sents the value of the variogram at the range and indicates the total amount of correlated spatial variation in the dataset.

When the variogram exhibits different sill or range values in different geometric directions the spatial correlation is a function of both distance and direction, and the random process is considered to be anisotropic. This is usually caused by underlying processes evolving differently in space or underlying landscape features (i.e., valleys, ridges, soil). If the variogram is only a function of distance (i.e. the sills will be the same in all directions) it is considered to be isotropic. In the isotropic case, the pattern (linear, exponential, quadratic, wave, power, etc) of the variogram is used to identify and model the underlying spatial dependency in the data. This model of spatial autocorrelation, $\delta(d)$, is incorporated with the larger process model that accounts for other trends or covariate predictors of disease.
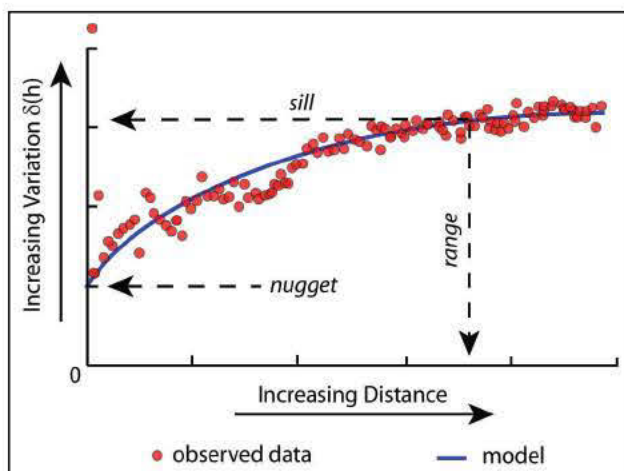
Figure 2.4. *Variogram for an exponential model of spatial variation showing the nugget (random variation), sill (total variation), and range (distance of correlation). Adapted from USDA. F.S.—N.E.F.I.A.*

Joly et al. (2006) conducted a geostatistical analysis of CWD infection data for south-central Wisconsin to evaluate both spatial auto-correlation and ecological factors hypothesized to be related to apparent prevalence. They used data to evaluate the potential effects of

34　　*Approaches to Landscape-Scale Modeling*

deer habitat, age, sex, and distance and direction from a suspected introduction site as factors affecting CWD prevalence, which was aggregated at the section (i.e., 2.6 km^2 or 1 mi^2) scale. Their analysis indicated that CWD prevalence declined over both a broad scale distance from the center of the outbreak area and at a scale reflecting local spatial correlation (i.e., 3.2 km or 2 mi radius). In addition, deer habitat was a significant predictor of CWD prevalence. Joly et al. (2006) used the resulting regression model to produce a map of predicted CWD prevalence.

Additional complexities may also occur in spatial disease data, but these topics are beyond the scope of our review. For example, CWD data may be aggregated into a finite collection of regular (e.g., sections, townships) or irregular (e.g., counties, wildlife management units) spatial sites or cells called lattices. Figure 1.3 illustrates the difference between continuous location of cases (dots) and potential aggregation into an irregular lattice of polygons. Methods for analysis of lattice data is described by Cressie (1993) and traditional applications to human diseases are considered by Elliott et al. (2001) and by Lawson and Williams (2001). It is also possible to consider spatiotemporal analysis, but in most cases this is simplified to a purely spatial process by aggregating over time. Because CWD is typically a slowly transmitted and slowly spreading disease it seems appropriate to aggregate over relatively short time frames (e.g., < 5-10 years). The general goal of geostatistical analysis is to develop models that incorporate disease risk factors and predict CWD infection at known spatial locations.

Questions addressed / model predictions:
1. Evaluates the spatial extent (distance) and direction of autocorrelation found in disease patterns.
2. Potentially evaluates the relationship of biotic and abiotic factors on disease infection or prevalence.

Data required:

1. Spatial coordinates of positive and negative CWD cases.
2. Spatial coordinates of polygon centroids if CWD cases are aggregated. However, for geostatistical analysis, aggregation should occur at a relative small scale compared to the area considered in the analysis.
3. Individual animal covariate data (e.g., age, sex) for factors of interest in predicting disease risk.
4. Spatial environmental or ecological covariate data (habitat, animal density, risk variables) for factors of interest in predicting disease risk.

Output:
1. Estimates the spatial autocorrelation of disease related to distance and/or direction.
2. Estimates prevalence parameters, covariate effect sizes, and other related statistics for factors/variables affecting the probability of disease.
3. Potentially depicts CWD spatial prevalence based on spatial autocorrelation and other significant factors affecting prevalence.

General usefulness:
Geostatistical methods are highly useful for spatial analysis of ecological and geographic processes that are sampled at irregular or random locations. These methods are most useful when the goal of the analysis is prediction at an unobserved spatial location. Geostatistical methods may be additionally useful when there is spatial dependence in the process that generates spatial patterns. Additional explanatory variables can also be included in the geostatistical analysis, leading to an investigation of spatial effects while controlling for explanatory factors or vice versa. Geostatistical methods generally assume a relatively small error in the spatial scale of locating animals (data points). This assumption may be reasonable when animals have small home ranges compared to the area of general analysis; however, for animals with seasonal migrations it may be important to separate analyses based on distinct summer or winter distributions. The georelational database structure of a geographic

Approaches to Landscape-Scale Modeling

information system (GIS) is ideally suited for storing and manipulating data used in geostatistical analyses.

Usefulness to CWD modeling and/or management:

If relevant spatial data are available, a geostatistical analysis has high potential applicability to CWD modeling and management. Based on the slow rate of CWD transmission and spread, potentially irregular spread of disease through heterogeneous habitats, and typically low rates of infection, it seems likely that spatial patterns and dependencies will be important components of CWD spatial distribution. Analysis of CWD data using geostatistical methods can facilitate the evaluation of biotic and abiotic factors affecting the risk of infection while concurrently accounting for likely spatial dependence. Modeling results can be used to produce maps of predicted CWD prevalence or risk. In addition, geostatistical analysis may provide useful insights on the dynamics of CWD spread by describing the extent of spatial correlation on the landscape. However, geostatistical methods may not be useful for all CWD infected areas. In particular, the spatial and temporal scales associated with data collection and CWD case location should receive careful consideration. Finally, aggregation of CWD data over a number of years seems highly likely to improve the distribution and precision of spatial data. Because CWD is a slowly transmitted disease, aggregation over a few years may not be problematic; however, there are currently no specific guidelines for determining the appropriate time frame for aggregation.

CELLULAR AUTOMATA MODELS

To understand what a cellular automata is, consider a chessboard or checker-board. The cellular part is represented by the squares and each cell can have one distinct state, such as color. Thus, a cell could be red or black (2 states), or yellow, white, or orange (3 states). The state must be discreet (i.e., an integer value) and finite; thus in the color example there would be no continuous shading and each cell would be one of a finite number of possible colors. For a three-dimensional problem, the squares would be cubes and the analogy would be a Rubik's cube. For CWD, time could be the third dimension so that x and y represent spatial extent, and z represents temporal pattern. Now we come to the 'automaton' part. As a cellular automaton model runs through time, at each time interval the state of the cells can change, or not, based on a deterministic or probabilistic rule. To enact the rule, each cell looks at the states, or color in this example, of nearby cells, and its own state (color), and then applies the rule to decide its state (color) in the next time step. All the cells change at the same time. This collection of cell-states and rule-based changes is called a *cellular automaton,* or cellular automata model. Two-dimensional systems of grid-cells are also called lattice-systems. Cells need not be blocks but can be any arbitrary shape. Although time must be discrete, it can be at any interval, from sub-seconds to years or longer. Even with very simple rules for each cell, these models can result in complex patterns and dynamics.

Cellular automaton models have been used to study the spatial and temporal rates of disease spread in spatially distributed host populations, as well as to evaluate the effectiveness of vaccination intervention strategies (Rhodes and Anderson 1997). A probabilistic automata network SIS (susceptible-infective-susceptible) model was developed to evaluate the spread of an infectious disease in a population of moving individuals (Boccara and Cheong 1993). When there was high movement, the spatial correlations in infection and recovery disappear and, as expected, the behavior of the system was then correctly predicted by a mean-field model, which assumed that every individual in the population is equally likely to contact every other individual. Results from the mean-field model diverged from the spatial cellular automata models when the neighborhood of interacting grid cells was reduced. At

a fine scale, cellular automata models can be individual-based models where cells can represent individuals with a spatial state. At a larger-scale, cells could represent populations or spatial areas, which would be more appropriate for landscape-level modeling.

Because cellular automaton models are general, they can be used to address a huge range of questions. The level of detail, spatial and temporal resolution, and inputs and outputs can be adapted to the specific questions of interest. Because this class of models includes such a broad range applications, we do not address these categories below.

General usefulness:

Cellular automaton models are particularly useful for developing an understanding how different factors may affect the spatial spread and distribution of a disease at a variety of spatial scales. The main strength of this method is its flexibility and the potential simplicity of rules within a cell. However, realistic data on vital rates, movement, transmission dynamics, population structure and distribution, relevant environmental/ecological (biotic and abiotic) factors, etc. are required to parameterize realistic cellular automaton models. Cellular automaton models are as realistic as the data used in their construction. Still, even with incomplete data, they may offer heuristic insights into the disease system.

Usefulness to CWD modeling and/or management:

The cellular automata method is very general, and could be used for a variety of different purposes. Given the appropriate data, these models could be used to investigate the likelihood of different localities as the point of CWD origin. The usefulness of this approach will depend on how a model is constructed and whether there is adequate data to parameterize it. However, the cellular automata umbrella offers a viable and potentially effective approach to spatial modeling of CWD to researchers with clear questions and appropriate data. Because of its potential usefulness for
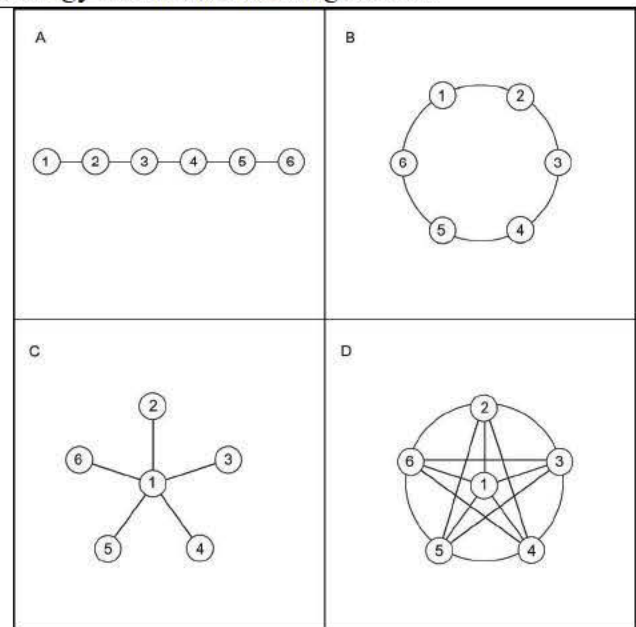
Figure 2.5. Different spatial configurations and continuity used in metapopulation modeling: (A) chain or necklace model, (B) loop model, (C) spider model, and (D) island model. Lines show connectivity between patches via dispersal.

modeling CWD spatial epidemiology, we present a related individual-based model as the focal approach for the fine-scale section.

METAPOPULATION MODELS

 Metapopulation models are based on the premise that there is a set of populations distributed over a number of patches, or areas, which are connected by dispersal (Figure 2.5). Early work on the theory of metapopulations assumed that dispersal among patches was limited and patches would go extinct and/or become recolonized over time (Levins 1969, Hanski and Gilpin 1991). A premise of metapopulation models is that patches "wink on and off" between an occupied and unoccupied state. With diseases, this would mean that the disease would wink on and off in the populations. However, this approach could be adapted for diseases that arrive and either do not wink off (population unoccupied by disease) or wink off very slowly. Thus, we prefer to think of meta-
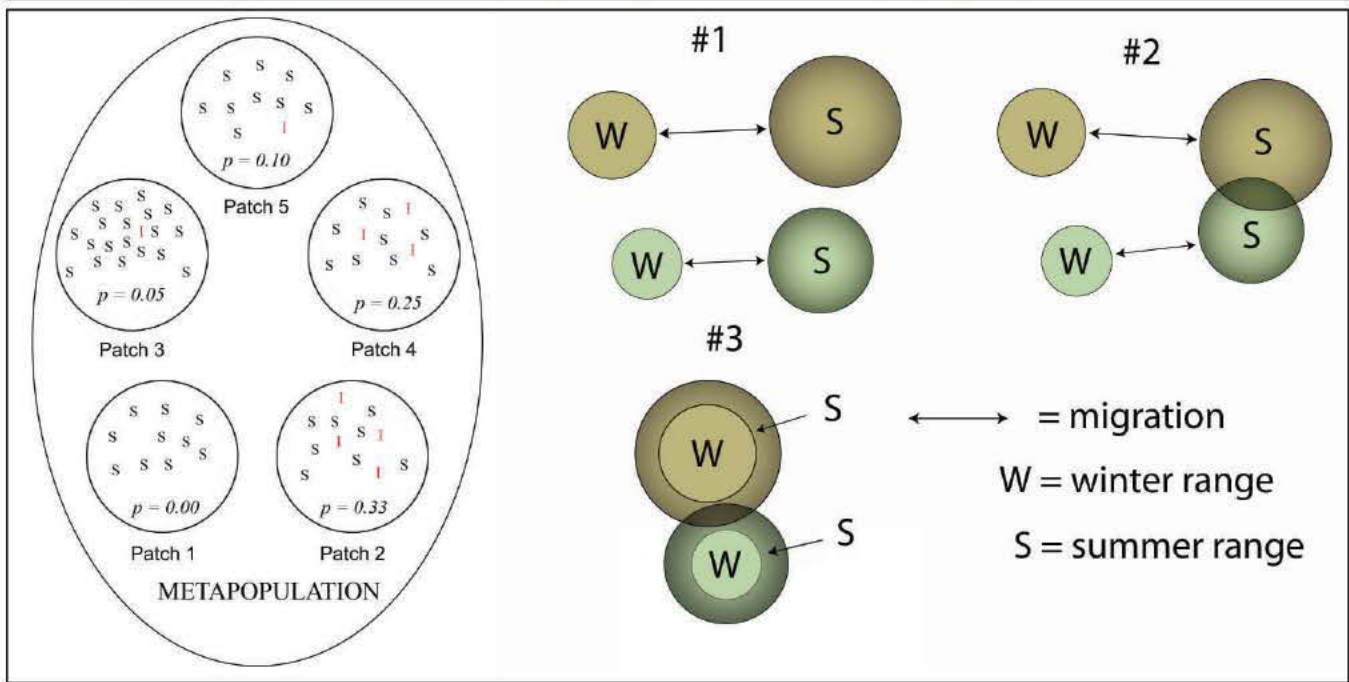
Figure 2.6. *Depiction of infected (I) and susceptible (S) individuals within metapopulations, adapted from Hess (1996), and the adaptation of this model to CWD for mule deer. Traditional metapopulations are connected by dispersal. For mule deer, the mechanism of connection between populations may also be due to either migration to summer range combined with summer range overlap or summer range expansion and overlap. Case # 1 there is no overlap, and hence is connection or metapopulation dynamics .In case # 2 there is migration to summer range combined with summer range overlap. This situation, the product of the probability of migrating × probability of range overlap could represent connection between the populations, which is analogous to the dispersal probability in metapopulation models. In case #3 there is no migration, but there is range expansion in the summer such that there is overlap between the 2 populations. In this situation, the probability of range overlap could represent connection between the populations. Finally, prevalence on winter range can be substituted for the numbers of I and S individuals from the metapopulation to adapt the metapopulation disease model of Hess (1996) to CWD models.*

population models as part of a continuum of spatially (or socially) structured subpopulations. At one end of the continuum dispersal is very rare and subpopulations are relatively independent of one another. While at the other end of the continuum dispersal is frequent and subpopulations are more connected with one another.

Metapopulation models of single species are often used to investigate how dispersal and subdivision affects spatial and temporal population dynamics. However, these models can also be applied to wildlife disease systems (Hess 1996, Swinton 1998, Cross et al 2005) where patches are then defined as single hosts

or a group (herd) of hosts, which are colonized by the parasite (Hess 1996, Hess et al. 2002). Here the 'patch' refers to the population (group) of susceptible and infected animals inhabiting an area that is separate from other populations (groups), but connected by dispersal or movement. The probability a population is 'occupied' by a disease can be modeled following the traditional metapopulation approaches of site occupancy by a species.

Metapopulation models are relevant to modeling diseases when there is spatial (or social) structure in the host population such that host vital rates and disease transmission rates depend upon local conditions within the population (herd). This is often the case for wildlife

diseases that are directly transmitted because hosts are often structured into groups and individuals are most likely to be infected by others within the same group. These models commonly add complexity by explicitly modeling dispersal and connections among subpopulations to more realistically model heterogeneous mixing between populations, which improves prediction of disease spatial spread. Using these models researchers can ask questions about the spread of disease from one population to the next and the likely effectiveness different management strategies, such as quarantine that may be implemented in some subpopulations and not others (Figure 2.6).

Questions addressed / model predictions:

1. Estimates the probability a patch (e.g., population, subpopulation, individual, spatial area such as winter range or wildlife management unit) becomes infected or recovers from infection as a function of within and between population dynamics and movements.
2. Facilitates evaluation of management strategies that may be implemented spatially (e.g., ring vaccination or depopulating areas/populations with high prevalence).

Data required:

1. Dispersal or migration routes and probabilities of connection between patches.
2. For a detailed model, vital rates of population dynamics as well as transmission dynamics within each patch/population.

Output:

1. Estimates disease prevalence for each patch over time.
2. Estimates colonization and extinction probabilities of disease infection for each patch.
3. Estimates the probability that the disease goes extinct through time for the entire metapopulation.

General usefulness:

Metapopulation models could be useful for spatially structured populations to evaluate dif-

ferent management strategies, such as quarantine. This approach is also useful to explicitly model between-population or subpopulation spread of a disease, as opposed to within-population spread, which is modeled by many spatial epidemiology models.

Usefulness to CWD modeling and/or management:

One of the difficulties in applying metapopulation models to wildlife populations is the problem of defining a subpopulation. Subpopulations or herds may vary in size, location and the amount dispersal between herds over time. For example, grouping behavior of elk and deer in many areas of North America are likely to vary between summer and winter months (Conner and Miller 2004). In many cases the amount of movement among groups and the degree of independence among groups is unknown. Metapopulation models may still work where there is overlap between subpopulations as long as there is little mixing. The connection between subpopulations, which is modeled as dispersal probability in traditional metapopulation models, could be modeled by probability of exchange or other biological surrogate for connectivity. Metapopulation models are less useful for modeling the spatial epidemiology of CWD in the more contiguously distributed white-tailed deer populations that also lack seasonal movements between discrete summer and winter areas.

DIFFUSION MODELS

Diffusion models are based on an assumption that the process being modeled can be approximated by random motion. The rationale for using a diffusion model is that, although individuals do not move randomly, the collective behaviors of a large number of individuals cannot be distinguished from predictions of a diffusion approximation (called mean field approximation). This assumption vastly simplifies both the construction and evaluation of models. In the realm of disease ecology, these models pre-

dict the spread of disease over time and, as discussed here, also have a spatial dimension. Diffusion models have a rich history in ecology and they have been applied to an exceptionally wide range of processes, including spatial epidemiology of disease (Okubo 1980).

Because of the underlying assumptions, diffusion models are most often applied to geographically widespread diseases and when transportation of animals or disease agents by humans, or dispersal is unimportant. An early application of diffusion theory was Noble's (1974) model for the spread of bubonic plague in Europe. Strict diffusion models may provide insights to broad-scale processes, especially as an alternative comparison with more complex spatial models. Recently, Reluga et al. (2006) constructed models that combined mathematical advantages of a diffusion approximation while permitting the inclusion of spatial structure, including movement of an animal within a home range.

For many diffusion models, the first date an infected animal is reported for a given spatial area, such as a county or wildlife management unit, is the required data. From this data, a differential model or trend surface for rate of spread can be retrospectively fit. In this section, we describe 2 potentially useful forms of the diffusion model. The first, a semi-diffusion model, models disease spread from area to area rather than across continuous differential space and relaxes the assumption of random movement. The second is trend surface analysis, which is based on date of infection, and can retrospectively describe the spread of a disease in both space and time and identify likely corridors or barriers.

SEMI-DIFUSSION MODELS

We use the term 'semi-diffusion' because at a landscape scale transmission takes place across borders of areas, rather than from a point source outward. Semi-diffusion models may allow more realistic modeling of spatial heterogeneity in disease spread because they summarize prevalence within area but not across areas, which requires data at the resolution of the area. In contrast, traditional spatial diffusion models attempt to model prevalence as continuous over space, which requires data at a fine resolution. Semi-diffusion models can be constructed for the spatial scale at which management is enacted. For example, wildlife diseases are often monitored by wildlife management units, and increased harvest or other management intervention actions are enacted for wildlife management units. With a semi-diffusion model, disease spread would be modeled across wildlife management units. The rate of diffusion or spread from one area to the other can be modeled as a function of prevalence in adjacent areas, proportion of border shared with infected areas, amount of connection via migration and dispersal, number of feeding sites in the area, etc. The diffusion rate could also be modeled as a function of any spatial environmental/ecological (biotic or abiotic) factors. A fundamental aspect of a semi-diffusion approach is explicit modeling of spatial dynamics (e.g., proportion of adjacent areas infected, prevalence in adjacent areas, etc.) in the difference equations, rather then modeling the effect through correlation structure. This approach could be framed as a model selection problem, with models representing different hypotheses about the spatial spread of the disease. Semi-diffusion modeling could be developed into a viable approach for any disease for which cases do not present themselves but where surveillance samples of infected and non-infected animals are collected.

Questions addressed / model predictions:
1. Estimates the probability an area becomes infected as a function of observed patterns, management actions, or environmental/ecological (biotic and abiotic) variables.
2. Predicts future spread of a disease from area to area.

Data required:

Approaches to Landscape-Scale Modeling

1. Date of first infection for a given spatial area, such as a wildlife management unit (date first detected is not the same as date of first infected, but if the two are close then date of first detected can be used for date of first infected).
2. Proportion of adjacent spatial areas that are infected for each time step.
3. Proportion of shared borders.
4. Prevalence within each area at each time step.
5. Proximity of spatial area to geographic features that could influence rate of spread, such as rivers (barriers or corridors), major highways (possible barriers), high ridgelines (possible barriers), etc.
6. Covariates for a spatial area expected to influence rate of spread, such as human density, animal density, disease prevalence, or environmental/ecological (abiotic and biotic) variables.

Output:
1. Predicts the probability an area will become infected at a given time step.
2. Estimates spatial and temporal prevalence of each spatial area at each time step.
3. Predicts rate of spread across entire study area.
4. Can provide estimates of covariate effects.

General usefulness:
Semi-diffusion models could be quite useful for diseases that do not present themselves, and it could be useful for fast or slowly spreading diseases. It would also facilitate evaluation of the importance of environmental/ecological (abiotic and biotic) factors in the spread of a disease. The main disadvantage of this method is that it requires several years of adequate (i.e., enough samples to have a high probability of detecting the disease if it is present) surveillance samples from contiguous areas included in the model.

Usefulness to CWD modeling and/or management:
Data collected for CWD surveillance is typically of the type required for a semi-

diffusion model, making this approach potentially viable. Running this type of diffusion model backwards in time may help researchers and managers generate hypotheses about the factors important to CWD spread and potential originating areas. Thus, a semi-diffusion model may provide heuristic insight to understanding the present spatial patterns of CWD. However, a semi-diffusion approach has limited potential for predicting the probability of spread of CWD into uninfected areas. The problem is that to generate a good model from which to predict from, we would have to know how the disease spread. Because most new cases of CWD have not initiated from spread of the disease, but rather from increased surveillance of an area, there are very limited data on the temporal aspect of spread, and this temporal aspect is an essential element of this method or any diffusion method. Thus, we conclude this method is not useful for CWD modeling at the present, but may be in the future.

TREND SURFACE

 Trend surface analysis uses spatial polynomial models that accounts for global effects and local autocorrelation. Models are fit using least-squares regression, except that residual autocorrelation is included in the model. For this discussion based on Figure 2.6, we use months as the time interval and counties as the spatial areas of interest. A trend surface estimates contours for the number of months to the first reported case (Figure 2.6) using the centroid coordinates of wildlife management units or other relevant management area (Moore 1999). From these contour lines, partial differential equations (∂time/∂x, ∂time/∂y) are derived to estimate the slope vectors from contour to contour. These slope vectors represent the rate of spread across the landscape. Large slopes represent high rates of spread which are interpreted as corridors. Similarly, areas with low rates of spread are interpreted as barriers (Figure 2.7).
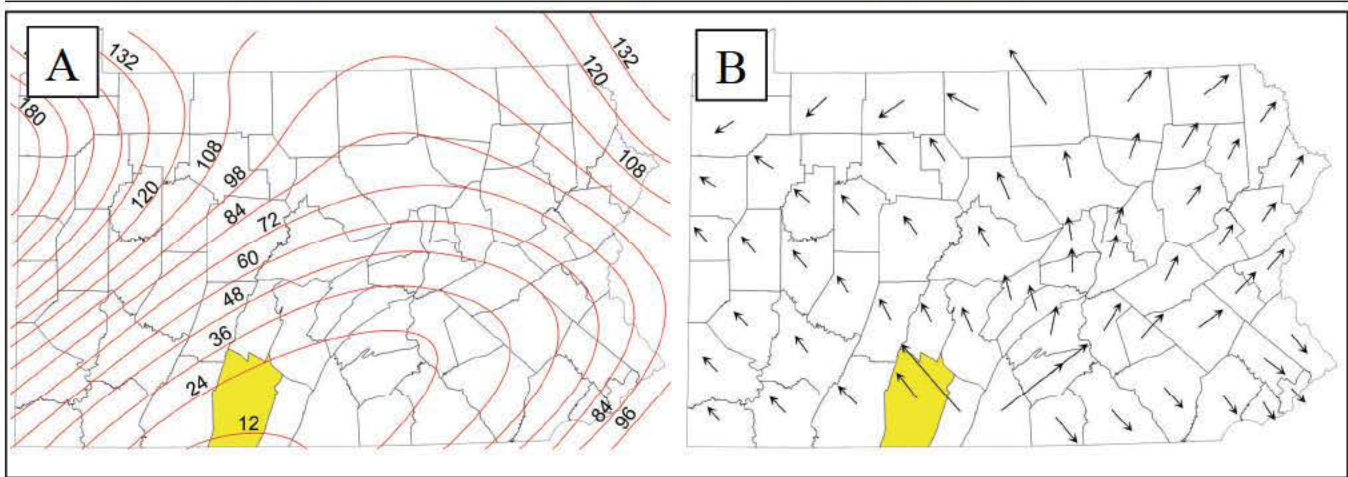
Figure 2.7. *(A) The predicted trend surface for month to the first report of a raccoon rabies case of disease outbreak by county, Pennsylvania, USA, 1982-1996. This figure illustrates the general direction and movement of the diffusion process originating in the county shaded in yellow. The contours represent the predicted number of months to the first reported case. (B) Velocity vectors (derived from the partial differential equations from the trend surface) show the speed and direction of diffusion at each county centroid. Longer arrows represent faster spread. Adapted from Moore 1999.*

Questions addressed / model predictions:
1. Estimates the rate of spread as a function of date that an area was first infected.
2. Identifies areas with fast and slow spread.

Data required:
1. Date of first infection for a given spatial area, such as a wildlife management unit (date first detected is not the same as date of first infected, but if the two are close then date of first detected can be used for date of first infected).
2. Centroid coordinates of the spatial areas.

Output:
1. Predicts contours of months to first reported case of disease.
2. Estimates rate of disease spread.

General usefulness:
Trend surface analysis is useful for infectious, moderate to quickly spreading epidemics in which cases present themselves. It is unlikely that the "first" case will be identified in an area, particularly in wildlife disease systems. Thus the rate of spread will be underestimated. However, as long as the time at which 'first' cases are identified is not biased by area (e.g., bias could arise if cases near ur-

ban areas are identified more quickly than cases in remote areas), trend surface analysis is usable for determining corridors of rapid spread and barrier areas which slow spread. Assigning areas to contain urban and remote areas could be important to avoid bias (e.g., incorrectly estimating spread to be fast around areas where there is a bias toward quick detection of the disease). However, this issue will be problematic when detection probabilities vary over space and time, which is likely for CWD.

Usefulness to CWD modeling and/or management:
Trend surface analysis is probably not useful for most CWD data sets because the date that an infected animal is first reported in an area may be completely unrelated to the date when the disease first occurred in the area. We note that the reason this method is not applicable is because of CWD surveillance and detection issues and not with the inherent usefulness of the method for CWD epidemiology. That is, if surveillance data were colleted randomly and representatively over the entire state or area of interest, trend surface analysis could work well for CWD.

FOCAL APPROACH: CLUSTER ANALYSIS

We selected cluster analysis as the focal approach for landscape level methods. While cluster analysis may not be as complicated or elegant as other approaches in this section, pattern detection with clustering remains a key element for both hypothesis formation and decision-making processes (Jain et al. 1999). State- or providence-wide surveillance programs, especially those involving hunter harvested animals (i.e., hunter harvest check stations) can provide large amounts of data (e.g., >25,000 samples) in a relatively short period of time (e.g., 1-4 months). Such a large influx of data often creates confusion and cluster analysis provides a convenient 'entry-point' into such a complex dataset. Further support for the use of clustering comes from situations where little prior information exists about the data (e.g., when starting a surveillance program, or monitoring new areas). In such cases it is prudent to begin with an exploratory analysis that makes as few assumptions about the data as possible - cluster analysis excels in this situation.

We selected Kulldorff's (1997) spatial scan statistic for our clustering algorithm as it requires fewer *a priori* parameters (e.g., the number of total clusters or the number of individuals in a cluster) than most clustering methods. It has also been shown through power comparisons to be the most powerful method for detecting localized clusters (Tango and Takahashi 2005) and has been used to identify areas of high CWD prevalence from surveillance data on white-tailed deer (Joly et al. 2006).

Scan statistics work by moving a window of variable size across the points in a dataset, counting the number of observed and expected cases falling within the windows (see Figure 2.2). The most likely cluster of high prevalence (i.e., not occurring by chance) is determined using a maximum likelihood ratio statistic, which determines whether there is higher prevalence inside the window compared with outside. By maximizing the likelihood function over all locations and window sizes the most likely cluster is identified. Cluster specific p-values are obtained using Monte Carlo simulations for primary and secondary clusters. Scan statistics are appropriate for detecting clusters

Figure 2.8. Partitioning of a dataset for Bernoulli analysis in program SaTScan^{TM}. The original data array contains a unique identifier, spatial references (x/y) and disease status. This dataset is split into three individual files:

1. *A vector of ID numbers for positive cases (red arrow).*
2. *A vector of ID numbers for negative cases (black) arrow.*
3. *A coordinate array containing all records.*

Approaches to Landscape-Scale Modeling

in space, time, or space-time, however we chose to demonstrate the method in the spatial dimension only. The following analysis was performed using the program *SaTScan*[TM] (Kulldorff 2006), which is available for free from http://www.satscan.org/.

Step #1: Data structure

Our example assumes that we are interested in detecting clusters of high prevalence of CWD by using disease status of individual animals, and therefore makes use of Bernoulli model framework. This type of model in SaTScan[TM] requires the spatial location of animals (x/y coordinates) and their status (0 for CWD negative, 1 for CWD positive). The SaTScan[TM] software requires partitioning of a dataset into three separate files; a positive case file, a negative case file, and a coordinate file covering all points. The positive and negative data files are simply a list of the respective unique identifiers. The coordinate file is linked to the two other files by the unique identifier field (Figure 2.8).

Step #2: Spatial scan & output data

We used two simulated data sets, each consisting of 1000 sampled individuals, to highlight the important aspects of the spatial scan statistic. The first dataset has a 10% prevalence rate randomly distributed across a theoretical landscape (Figure 2.9).

Three tabs are available (INPUT, ANALYSIS, OUTPUT) for the user to specify the model. For the Bernoulli approach we only need to specify the upper limit of the window size. This can either be set as a percentage of the population (both positive and negative events) or as the radius of the circle. For our analysis we used the default of 50% of the population. That is, the scan window will increase from 0 to a size that contains 50% of the population (i.e., 500 individuals) for each point it searches. With very large data sets searching 50% of the population for each point can take substantial time and we suggest setting the upper limit as a fixed radius rather than percentage of the population. The SaTScan[TM] output
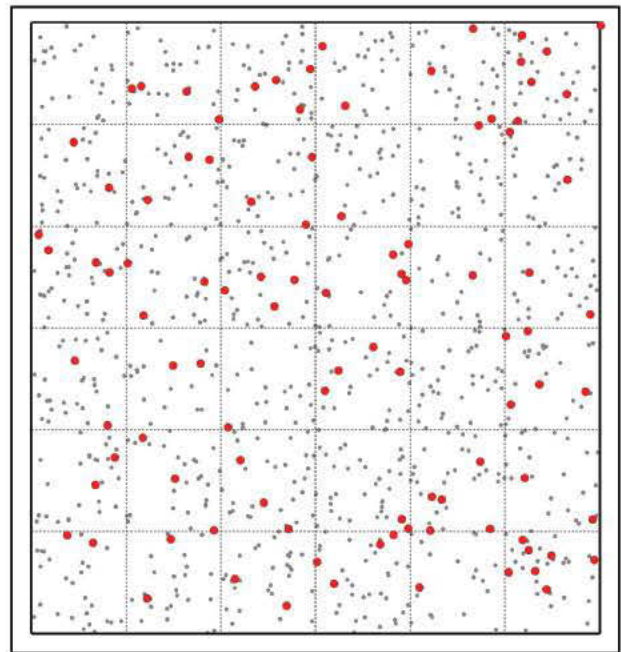
Figure 2.9. *Randomly distributed 10% prevalence data. Red dots indicate positive cases (n=100), gray dots are negative cases (n=900).*

provides various options to facilitate integration with a GIS mapping environment. This allows for visual mapping of the most likely clusters selected from a particular analysis.

It is important to remember that almost all clustering methods will produce clusters (that's their job!), even if they are not biologically relevant. To prevent such "false clustering" the spatial scan statistic tests if the pattern of positives and negatives inside any potential cluster is significantly different from the pattern observed outside that cluster. For our purely random 10% prevalence dataset, the Bernoulli spatial scan results show that even the most likely cluster is nonsignificant ($p = 0.636$) (Figure 2.10). We can infer there is no significant clustering of prevalence in this dataset.

To highlight the scan statistic's ability of detecting true clusters of higher prevalence from clusters caused by increased sampling intensity, we simulated a second dataset. This dataset contains an area of high prevalence and a second distinct area of dense monitoring (many positives coupled with many negatives)

Purely Spatial analysis
scanning for clusters with high rates
using the Bernoulli model.

SUMMARY OF DATA

Number of locations......: 1000
Total population..........: 1000
Total cases..............: 100

MOST LIKELY CLUSTER

1.Location IDs included.: 935, 691, 955, 941, 902
   Coordinates / radius..: (417693,5.07846e+06) / 1912.70
   Population............: 5
   Number of cases.......: 4
   Expected cases........: 0.50
   Observed / expected...: 8.000
   Relative risk.........: 8.292
   Log likelihood ratio..: 6.882812
   Monte Carlo rank......: 636/1000
   P-value...............: 0.636

Figure 2.10. (Top) Cluster map for random 10% prevalence data. (Bottom) Output file from SaTScan™.

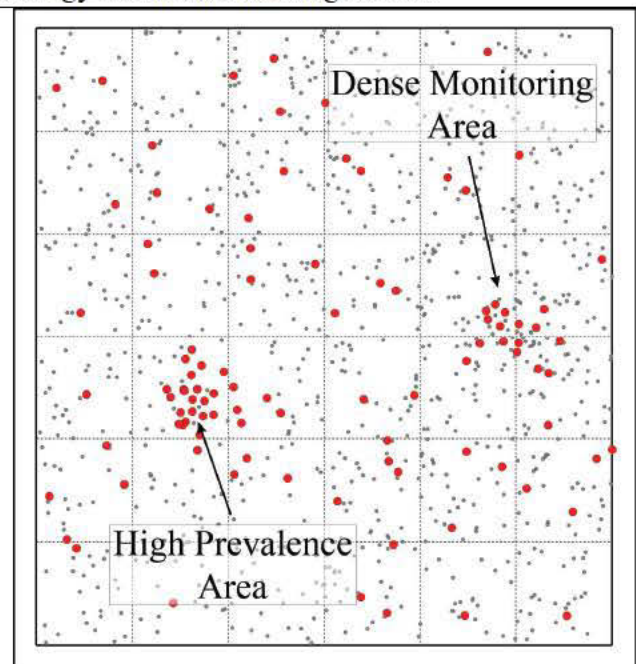*Figure 2.11. Clustered 10% prevalence data with high prevalence and dense monitoring regions. Red dots indicate positive cases (n=100), gray dots are negative cases (n=900).*

(Figure 2.11). Distinguishing between these two potential clusters represents the usual situation where simple spatial clustering often fails (by not accounting for the negative points). Simulating an area of dense monitoring is also relevant to CWD monitoring, because surveillance effort often varies across a landscape.

The scan statistic correctly identified the high prevalence area as significantly different

from the surrounding population ($p = 0.001$). The increased sampling area was identified as a secondary cluster, but this cluster was not statistically significant. Accordingly, we can conclude that this cluster of positives is not significantly different from the surrounding population. It is simply the manifestation of a higher density of samples (both positive and negative) (Figure 2.12).

GENERAL CONCLUSIONS ABOUT LANDSCAPE-SCALE APPROACHES FOR CWD MODELING

As at the regional scale, we recommend cluster analysis as the first step of describing CWD patterns. Although there are many approaches to mechanistically modeling the spatial epidemiology of disease at this scale, their utility in the context of CWD is limited by several factors. The most important factor is lack of data. That is, we cannot reliably predict the spatial spread of CWD because sampling is usually insufficient to determine whether the disease is absent from some areas, or present but not detected. Diffusion models are inappro-
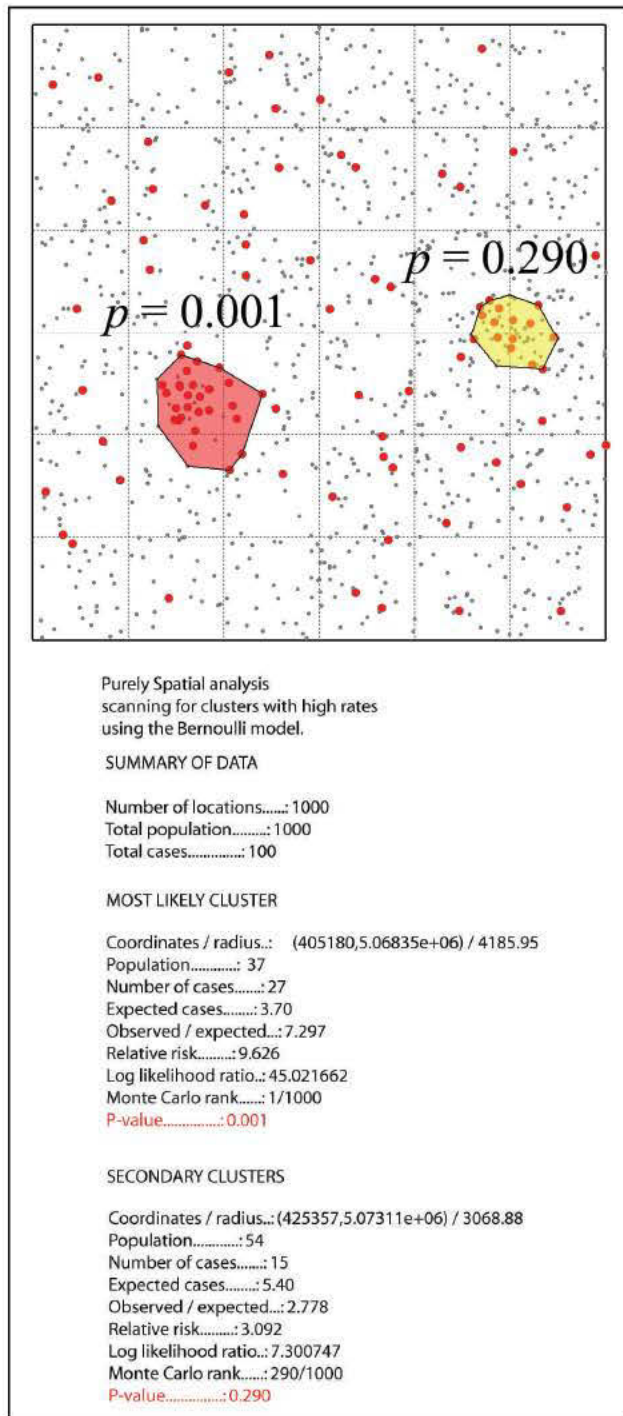
Purely Spatial analysis
scanning for clusters with high rates
using the Bernoulli model.

SUMMARY OF DATA

Number of locations......: 1000
Total population.........: 1000
Total cases.............: 100

MOST LIKELY CLUSTER

Coordinates / radius..:   (405180,5.06835e+06) / 4185.95
Population...........: 37
Number of cases......: 27
Expected cases.......: 3.70
Observed / expected..: 7.297
Relative risk........: 9.626
Log likelihood ratio..: 45.021662
Monte Carlo rank......: 1/1000
P-value.............: 0.001

SECONDARY CLUSTERS

Coordinates / radius..: (425357,5.07311e+06) / 3068.88
Population...........: 54
Number of cases......: 15
Expected cases.......: 5.40
Observed / expected..: 2.778
Relative risk........: 3.092
Log likelihood ratio..: 7.300747
Monte Carlo rank......: 290/1000
P-value.............: 0.290

Figure 2.12. Spatial scan cluster map and SaTScan™ output table.

priate for additional reasons. For example, diffusion models are typically predicated on random mixing of individuals and, for some models, on the fact that the date associated with the first infected sample in an area represents the first occurrence of the disease in the area. The

random mixing assumption can be relaxed if factors causing differences in the rate of spread are explicitly included in the model. However, the second assumption is more problematic for CWD. The date when CWD is first detected in a new area usually does not represent a first infection, but rather a first detection. Diffusion methods typically assume that disease cases "present themselves" because they were developed for human diseases, where sick humans "present" themselves to doctors for treatment. However, deer with CWD rarely present themselves for testing. Thus, when we detect CWD for the first time in an area it likely that the disease was already there, perhaps for some time, but the area was not adequately sampled. Thus, until all areas are adequately sampled, surveillance data cannot be used in diffusion methods to model the spread of CWD in a meaningful way.

In addition, due to the apparently slow spatial spread of CWD, we believe that diffusion, cellular automata, and metapopulation models are unlikely to be particularly useful for addressing many CWD questions. In particular, we expect that the ecological system could change dramatically (e.g., hunting pressure, land use, water distributions, animal translocations [legal or otherwise], etc.) over the amount of time that it is likely to take CWD to spread, and that these factors might play a much larger role than the diffusion process. We have one caveat to our skepticism. If surveillance data over larger areas are adequately sampled, the semi-diffusion approach may be useful for predicting potential spread, or to evaluate management actions occurring at an appropriate scale, such as county or wildlife management unit. However, given that CWD appears to spread slowly, even if the semi-diffusion approach is viable it may not yield helpful results. That is, it may be years before CWD is predicted to move from populations in one large area to populations in another large area.

The largest data gap at the landscape scale is the lack of adequate sampling to detect

newly infected areas. Enough samples to ensure a 99% probability of detecting a prevalence of <1% should to be collected over all areas of interest in order to predict any future spread of CWD. Sample sizes and designs to achieve this were thoroughly discussed in the previous workshop (Samuel et al. 2003). Because of inadequate samples, the first time CWD is detected in an area often represents the first time there are adequate samples and power to detect low prevalence CWD, not the first time it occurs or "spread" there. Modeling spread of disease based on observed patterns will not be valid until there are adequate samples in the relevant study areas.

In areas where sampling is powerful enough to describe present spatial patterns of CWD epidemiology, running models backwards in time may be a fruitful line of future investigation. This approach could reveal likely originating locations and times, as well as potential patterns of spread, that led to present patterns. Hypotheses of originating locations, times, and patterns of spread, including corridors and barriers, could be constructed. The forward projection of the outcomes of these hypotheses could be compared to the observed patterns via model fit statistics. The endemic area of Colorado and Wyoming may be an area with adequate samples to attempt a backwards time approach.

LITERATURE CITED:

Boccara, N., and K. Cheong. 1993. Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. Journal of Physics A: Mathematical and General. 26:3707-3717.

Clark, A. B., and A. B. Lawson. 2002. Spatio-temporal clustering of small area health data. Pages 235-256. In A. B. Lawson, and D. Denison, Editors. Spatial cluster modeling. Chapman & Hall, London, England.

Conner, M. M., and M. W. Miller. 2004. Movement patterns and spatial epidemiology of a prion disease in mule deer population units. Ecological Applications 14:1870-1881.

Cressie, N. A. C. 1993. Statistics for spatial data. John Wiley & Sons, Inc., New York, New York, USA.

Cross, P. C., J. O. Lloyd-Smith, and W. M. Getz. 2005. Disentangling association patterns in fission-fusion societies using African buffalo as an example. Animal Behaviour 9:499-506.

Diggle, P. J. 2000. Overview of statistical methods for disease mapping and its relationship to cluster detection. Pages 87-103 In P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Spatial epidemiology. Oxford University Press, Oxford, England.

Diggle, P., P. Riberio, Jr., and O. Christensen. 2003. An introduction to model-based geostatistics. Pages 43-86 In spatial statistics and computational methods. J. Moller, Editor. Springer, New York, New York, USA.

Elliott, P. J. Wakefield, N. Best, and D. Briggs, editors. 2001. Spatial epidemiology methods and applications. Oxford University Press, Oxford, England.

Farnsworth, M. A., J. A. Hoeting, N. T. Hobbs, and M. W. Miller. 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. Ecological Applications 16:1026-1036.

Hanski, I., and M. Gilpin. 1991. Metapopulation Dynamics: brief history and conceptual domain. Pages 3-16 In M. Gilpin and I. Hanski, Editors. Metapopulation dynamics: empirical and theoretical investigations. Academic Press, London, England.

Heisey, D. M., D. O. Joly, and F. Messier. 2006. The fitting of general force-of-infection models to wildlife disease prevalence data. Ecology 87:2356-2365

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

Hess, G. R., S. E. Randolph, P. Arneberg, C. Chemimi, C. Furlanello, J. Harwood, M. G. Roberts, and J. Swinton. 2002. Spatial aspects of disease dynamics. Pages 102-118 In P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. The ecology of wildlife diseases. Oxford University Press, Oxford, England.

Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No. 3.

Joly, D. O., M. D. Samuel, J.A. Langenberg, J. A. Blanchong, C. A. Batha, R. E. Rolley, D. P. Keane, and C. A. Ribic. 2006. Spatial-epidemiology of chronic wasting disease in Wisconsin white-tailed deer. Journal of Wildlife Diseases 42:578-588.

Kulldorff, M., and N. Nagarwalla. 1995. Spatial disease clusters: detection and inference. Statistics in Medicine 14:799-810.

Kulldorff, M. 1997 A spatial scan statistic. Communications in Statistics: Theory and Methods 26:1481-1496.

Kulldorff, M., and Information Management Services,

Inc. 2006. SaTScan™ v7.0: Software for the spatial and space-time scan statistics. htpp://www.satscan.org/.

Lawson, A. B. 2001. Statistical methods in spatial epidemiology. John Wiley & Sons, Chichester, England.

Lawson, A. B., and F. L. R. Williams. 2001. An introductory guide to disease mapping. John Wiley & Sons, West Sussex, England.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America. 15:237-240.

Moore, D. A. 1999. Spatial diffusion of raccoon rabies in Pennsylvania, USA. Preventative Veterinary Medicine 40:19-32.

Noble, J. V. 1974. Geographic and temporal development of plagues. Nature 250:726-728.

Norstrom, M., Pfeiffer, D. U., and Jarp, J. 2000. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds. Preventive Veterinary Medicine 47:107-119

Okubo, A. 1980. Diffusion and Ecological Problems: Mathematical Models. Springer-Verlag, Berlin, Germany.

Perez A.M., M.P. Ward, P. Torres, and V. Ritacco. 2002. Use of spatial scan statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. Preventive Veterinary Medicine 56:63-74.

Reluga, T., J. Medlock, and A. Galvani. 2006. A model of spatial epidemic spread when individuals move within overlapping home ranges. Bulletin of Mathematical Biology 68:401-416.

Rhodes, C. J., and R. M. Anderson. 1997. For disease spread, epidemic thresholds and vaccination in a lattice model of disease spread. Theoretical Population Biology 52:101-118.

Samuel, M. D., D. O. Joly, M. A. Wild, S. D. Wright, D. L. Otis, R. W. Werge, and M. W. Miller. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk. USGS-National Wildlife Heath Center Report. Madison, Wisconsin, USA. Available at: http://www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf .

Sankoh O.A., Y. Ye, R. Sauerborn, O. Muller, and H. Becher. 2001. Clustering of childhood mortality in rural Burkina Faso. International Journal of Epidemiology 30:485-492

Shibli, S.A.R. 1997. Geostatistics FAQ. Available at: http://www.ai-geostats.org/fileadmin/Documents/FAQ/FAQ_Geostatistics_01.pdf

Swinton, J. 1998. Extinction times and phase transitions for spatially structured closed epidemics. Bulletin of Mathematical Biology 60:215-230.

Tango, T., and K. Takahashi. 2005. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geographics 4:11.

Viel, J. F., P. Arveux, J. Baverel, and J.Y. Cahn. 2000. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. American Journal of Epidemiology. 152: 13–19.

Wakefield, J. C., J. E. Kelsall, and S. E. Morris. 2000. Clustering, cluster detection, and spatial variation in risk. Pages 128-152 *In* P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. Spatial epidemiology. Oxford University Press, Oxford, England.

Approaches to Fine-Scale Modeling
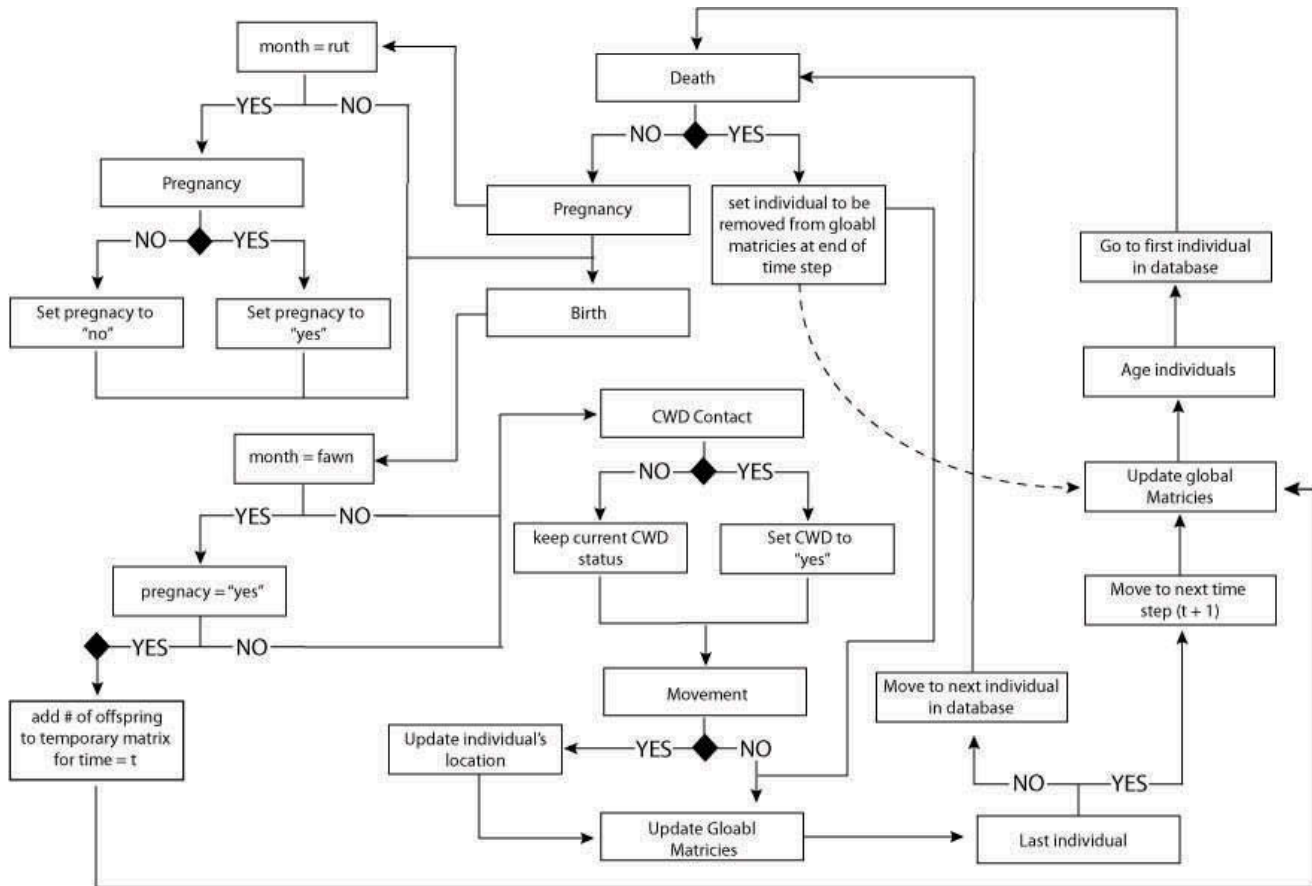
A hypothetical conceptual model for an individual-based spatially explicit model for chronic wasting disease. Black diamonds represent more complex decision rules based on the current conditions of the model and model parameters (see focal approach for more details on IBMs).

We consider the term 'fine scale' to describe small areas occupied by one population or sub-population. Models at this scale may attempt to incorporate a high degree of detail in the processes that lead to transmission of disease, including parameters that define both direct and indirect transmission, seasonal or age-specific effects on vital rates, and a detailed representation of movement and aggregation patterns. At a fine scale, the processes that determine CWD transmission may be revealed by comparing results from models that differ in transmission function, spatial connection, or social structure to observed data. For CWD, most data that can be used to infer mechanisms of transmission come from observations of disease dynamics of captive herds (Miller et al.

2006), which represents a very fine scale and resolution (detail) of data.

Although mathematical models contain unambiguous assumptions, subjective decisions the appropriate scale and level of detail may still be required. For example, researchers must decide how to model transmission (e.g., direct, indirect, horizontal, vertical, etc.) as well as the appropriate functional relationship between population density and transmission (i.e., density or frequency dependent). Decisions involving the level of detail for modeling population and demographic processes are also required. At fine scales, models are frequently age or stage structured, transmission is explicitly represented, and population processes

(e.g., birth, mortality, etc.) are often included, particularly for diseases where on a similar timescale as birth and death. In addition to the effect of disease on population performance, the model may explicitly limit population size via density dependence, harvest, and/or density independent processes. Data required to reliably estimate model parameters are more often available at the level of a single population or finer scale and are rarely available at broader spatial scales.

COMPARTMENT MODELS

At a variety of scales, many models of disease dynamics will divide the host population into categories of susceptible, infected, and recovered (SIR) (e.g., Anderson and May 1979, Anderson and May 1991, Hudson et al., 2001), where recovered can indicate removal from the susceptible pool through acquired immunity (Figure 3.1A). We note that for CWD, the appropriate compartment model is a SI (Figure 3.1B) because animals do not recover. SIR-type models have led to a broad range of important insights to disease dynamics and control strategies during the last 80 years (Kermack and McKendrick 1927, Bartlett 1957, May and Anderson 1978, Anderson 1979, Hudson et al. 2001). The basic SIR model structure has been expanded to accommodate many complex details, including latent periods between infection and infectiousness, age and sex structure, individual variation in susceptibility and infectiousness, and spatial/social structure (Figure 3.2). Moreover, for modeling CWD, an environmental reservoir can easily be included in these SIR-type models (Figure 3.1B). Compartmental
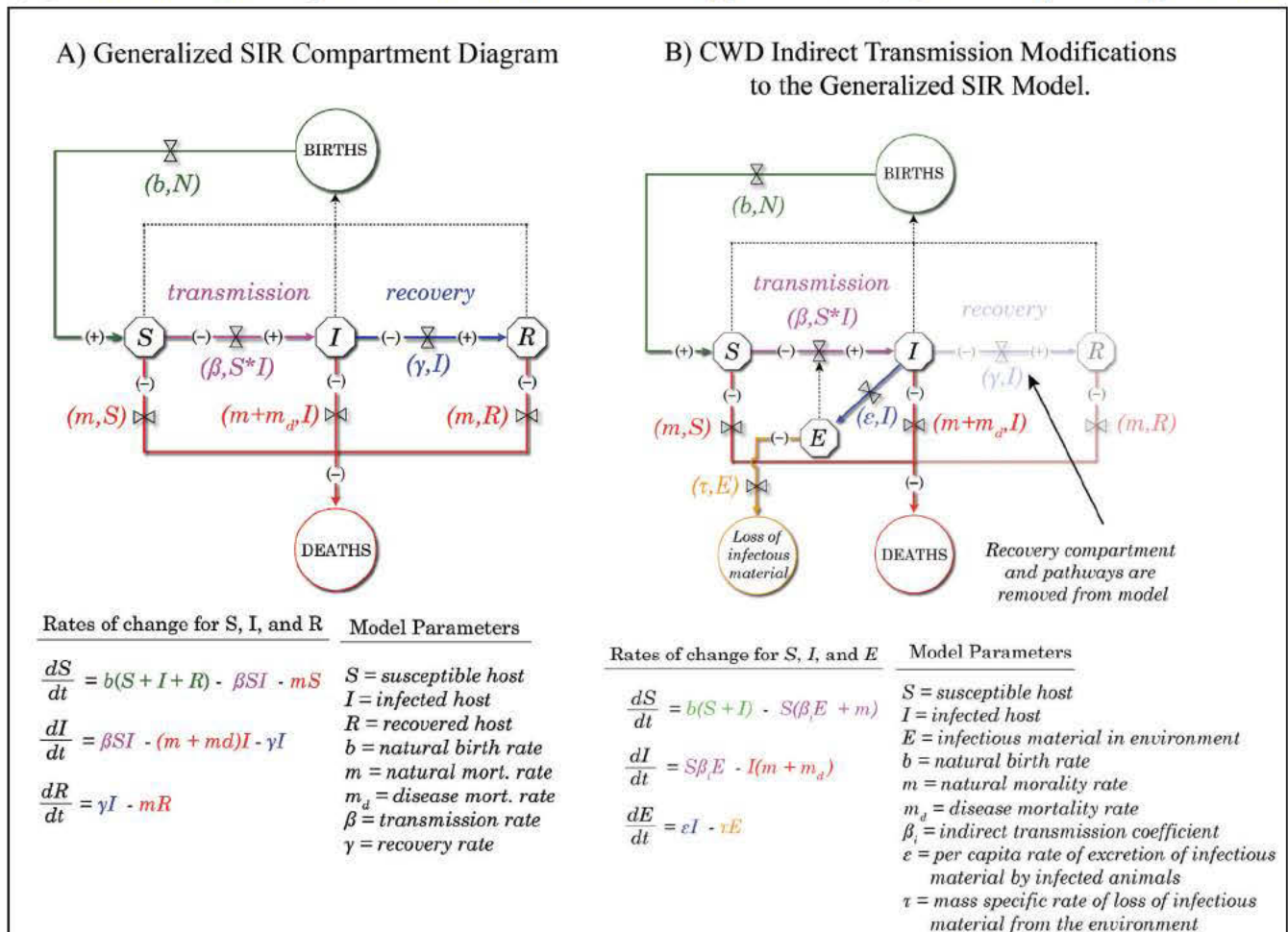
Figure 3.1 *Compartments and traditional differential equations for (A) generalized SIR model and (B) CWD adaptation.*
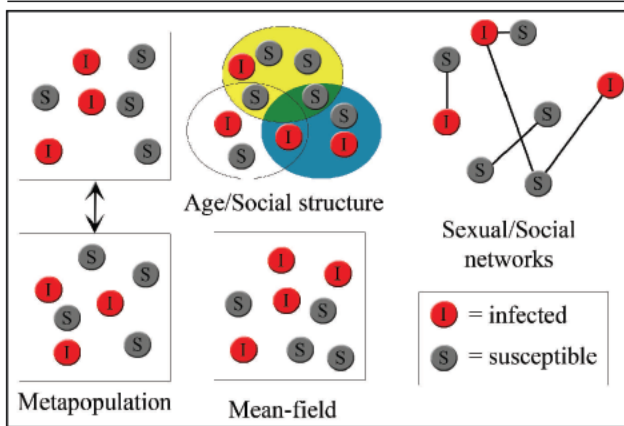
Figure 3.2. Elaborations of traditional SIR models of disease transmission. Adapted from Ferguson et al. (2003).

SIR models can be deterministic or stochastic, spatial or non-spatial, and composed of difference or differential equations. Statistical estimation via likelihood theory can be used to estimate model parameters, while model selection methods, such as Akaike's Information Criterion or Bayesian Information Criterion, can be used to compare and evaluate support for competing SIR model structures. SIR models that are composed of a relatively small number of differential equations may be solved using analytical tools. When additional details are added to SIR models, it quickly becomes more difficult to find analytical solutions, estimate model parameters, and evaluate the level of support for different model structures. As a result, researchers usually explore more complicated model structures via computer simulation. Simulation and randomization techniques can be used to model more complicated structures or to evaluate the effects of stochasticity in various model parameters.

Compartment models have been used to evaluate potential control and transmission of CWD. Hobbs (2006) constructed a relatively simple compartment model to explore the potential of a predator (e.g., large carnivore) that selectively fed on CWD-infected elk to control or eradicate the disease. The model showed that under circumstances thought to be within the bounds of realistic parameter estimates, a small positive selection for infected elk would have a large influence on prevalence of CWD.

In another example, Miller et al. (2006) constructed a set of six compartment models that varied in complexity and in potential routes of transmission. Beyond the basic SIR structure, model complexity varied by including (or not including) a latent period, indirect transmission, and an incubation period. They fitted model parameters to observations from two epidemics in captive herds of mule deer, using information criteria to identify the models that best matched observation. Model results for the two epidemics in captive herds best supported a model that included only indirect transmission, substantiating empirical evidence for environmental transmission of CWD in mule deer (Miller et al. 2006). Transmission rates estimated by Miller et al. (2006) are likely much greater than those in free-ranging deer, and they thus provide an upper bound for modeling CWD transmission and spread over larger spatial scales. Results from this and several other studies suggest the role of an environmental reservoir of infection. This environmental reservoir should be considered in the construction of SIR-type models of CWD as well as several of the other modeling methods we discuss in this text.

Compartmental models cover an exceptionally broad range of model types, as evidenced by Anderson and May's (1991) 700 page book, which focuses on models based on SIR-type structures. It is thus neither possible nor useful to describe all the kinds of questions that can be addressed with these models. In general, compartmental models are most suited to large populations, where aggregate behaviors adequately account for disease and population dynamics. Problems with demographic and/or disease stochasticity may arise when host populations are small or disease is uncommon. Compartmental models that are not individual-based and assume that all individuals are equal within their particular disease category are often not suited to simulating dynamics where the attributes or behavior of individuals are important (e.g., where there are socially dominant animals, or where movement patterns are highly heterogeneous). Compartment models

51

may need to be individual-based where the number of infectious individuals is small, the spatial scale is small, or where there is considerable and important heterogeneity between individuals.

Questions addressed / model predictions:

1. Predicts R_0, (the average number of secondary cases that arise from a single case at the start of a disease outbreak) and associated disease dynamics, including rates of flux between groups of susceptible, infected, and recovered (or dead).
2. Depending on model detail, compartmental models can address dynamics of disease with latent periods.
3. Estimates rates of spatial spread of disease.
4. Facilitates evaluation of types and relative importance of models and mechanisms of transmission.
5. Estimates threshold population size for persistence of disease.

Data required:

1. Data requirements are highly dependent on model structure and level of detail. Minimal requirements would include data on the proportion of the population in each class (susceptible, infected, or recovered) over time.
2. For highly detailed compartment models, additional data may be required on movement rates, sex and age composition, disease state, on social contacts, effects of infection on vital rates, factors related to disease resistance, effects of environmental contamination levels, and population demographic processes.

Output:

1. A minimal set of outputs would be the number of individuals in each disease class of susceptible, infected, and recovered or dead at each time step. With further embellishment, parameters can be fit to data to estimate such things as latency period, number of infectious contacts, mode of transmission, a threshold population size below which the disease cannot persist (if

any), rate of spread, and many other attributes.

General usefulness:

Compartment models provide a versatile and well understood approach to modeling diseases, especially at a fine scale. Mathematical techniques for estimating parameters and analyzing model behaviors are generally known, and this knowledge greatly facilitates model construction and evaluation. The ability to use analytical mathematical techniques to fully understand model dynamics makes these models particularly suitable for exploring the potential effects of management actions.

Usefulness to CWD modeling and/or management:

Compartment models can be extremely useful for modeling transmission and dynamics of CWD. In particular, simple models can be quickly and easily constructed to simulate and evaluate the effects of assumptions such as transmission mode and rate, control or eradication strategies, and population processes. Anderson and May (1991) provide a compendium of compartment model structures and a wide range of applications.

INDIVIDUAL-BASED MODELS
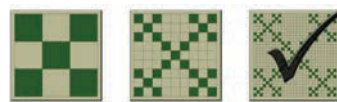
 Individual based models (IBM) explicitly represent each individual in one or more populations. In an IBM, individuals are typically characterized by their sex, age, disease status, and other relevant characteristics that can include physiological state, genetic constitution, reproductive condition, resistance to disease, membership in a social group, propensity to migrate, etc. Bonabeau (2002) noted that individual, or agent-based, models are likely to be appropriate when:

- Individual attributes likely to affect disease dynamics are highly heterogeneous.

- Transitions are non-linear and may be characterized by threshold of behavior (e.g., sudden long-range jumps).
- The focus is on initial stages of disease invasion, or when the disease is at low prevalence such that the discrete nature of individuals and stochasticity are important to the ultimate dynamics of the disease.
- Interactions between individuals are heterogeneous (e.g., via social or mating structure) and these interactions result in large deviations from a predicted aggregate behavior.
- Averages are inappropriate and exceptional or rare events are important (e.g., a rare infection that leads to an epidemic).

These traits are characteristic of most natural animal populations, and they may be very important at some spatial scales. A key advantage of IBMs over many state-variable models (i.e., models that aggregate individual into large, homogeneous classes such as females and males) is the potential ability to model the attributes of individuals and the mechanisms by which individuals interact with their environment. By so doing, IBMs do not require simplifying assumptions that we know are false. By contrast, many state-variable models require estimation of parameters that operate over broad spatial and temporal scales – measurements that are frequently difficult and expensive, to obtain, and that are estimated with wide confidence intervals. Model structure and model parameters in IBMs are generally easy to interpret, and to explain to non-technical audiences. Huston et al. (1988), DeAngelis and Gross (1992), and Grimm and Railsback (2005) provide more comprehensive descriptions of IBMs and their applications.

By concept, IBMs can be very simple and require only a few easily-measured parameters. However, it is very easy for modelers to construct highly detailed IBMs and there is often a tendency to do so. Highly detailed IBMs of CWD may be useful for scenario analyses, but they may also be impossible to validate because they will likely require estimating a large number of poorly know parameters. With complex IBMs, interactions between functions and individuals can lead to substantial difficulties in attempts to directly relate changes inputs to changes in model behavior. IBMs are generally not suitable for analytical analyses, and a key step in model development is to conduct a comprehensive sensitivity analysis.

As both the spatial scale and number of animals increase, simpler models may adequately mimic system dynamics. Recent research, however, has shown the importance of individual variation in disease dynamics (Lloyd-Smith et al. 2005a). Many disease models, particularly those of microparasitic infections (e.g. bacteria and viruses), assume that all individuals are the same with respect to their infectiousness and susceptibility. For sexually-transmitted and vector-borne infections there have been many studies illustrating wide variation in individual contact rates (Kretzschmar 2000, Liljeros et al. 2001, Eames and Keeling 2004).

This led to the concept of a general 80-20 rule, whereby 80% of infections are likely to be caused by only 20% of the infectious individuals (Woolhouse et al. 1997, Woolhouse et al. 2005). Lloyd-Smith et al. (2005b) showed that for human microparasitic diseases, a large skew in the number of infections caused by different individuals was common and even more skewed than what would be expected from the 80-20 rule. These highly infectious individuals, the superspreaders, are likely to play a large role in the disease dynamics, and this individual heterogeneity is easily incorporated into IBMs. Theoretical modeling suggests that disease systems with a large degree of heterogeneity in individual infectiousness are more likely to go extinct, but if they do persist they tend to have more explosive dynamics. Furthermore, control efforts focused on superspreaders are much more effective than control measures that are broadly applied to the entire host population (Lloyd-Smith et al. 2005b). At this point, there are no data on contact rates and the infectiousness of different

individuals for CWD. However, the variation in prevalence among different sex and age groups (Miller et al. 2000, Miller and Conner 2005), as well as potential differences in genetic susceptibility (Jewel et al. 2005), suggest that substantial individual variation may also exist in CWD systems.

Gross and Miller (2001) and Cary (2004) constructed IBMs to explore dynamics of CWD in deer populations, the former in Colorado mule deer and the later in Wisconsin white-tailed deer. The Colorado model was non-spatial and simulated CWD dynamics in a single, closed population, whereas the Wisconsin model included a high degree of detail on small-scale movements of deer in an agricultural landscape. These differences in model detail reflected the relative availability of data from the two regions and the types of questions the models were designed to address. Both models were developed to evaluate the effects of a range of potential management options to control or eradicate CWD.

A comparison of the IBMs developed by Gross and Miller (2001; hereafter G-M) and Cary (2004; hereafter Cary) is a useful illustration of alternative approaches to model development. The G-M model was specifically developed to examine potential impacts of CWD on mule deer populations in the endemic areas of Colorado. Relatively good data on the individual epidemiology of CWD were available from captive animal studies, but similar to many wildlife disease systems few data were available on naturally infected populations and individuals. Model construction and parameter estimation and evaluation reflected the paucity of data and the need to broadly explore model behavior. The non-spatial IBM simulated a single population, and incorporated a simple frequency-dependent, random-mixing social structure for disease transmission, to broadly explore model behavior. Results were presented for a wide, but realistic, range of parameter values, and only general (versus specific) model dynamics were discussed. Simulations showed that all realistic sets of parameters eventually caused dramatic declines in deer populations, and that all disease control strategies would require intensive, long-term commitments and resource investments.

By contrast, the Cary model included a highly detailed spatial representation of the study area, and estimates of model parameters were based on a broader range of studies of deer biology, harvest data, and a very detailed land classification map. Nonetheless, the level of detail in this model required estimating many parameters for which there was relatively little data. The spatial extent of the model was explicit and consisted of 20736 grid cells, each representing 0.65 km^2 (i.e., 0.25 mi^2 or 160 acres). During simulations, the position (grid cell) of each individual was tracked, and deer were anchored to specific home ranges, which could shift in response to winter feeding. Cary's model was constructed to evaluate a series of specific management actions, on a very specific population inhabiting a well-defined landscape. Cary examined a variety of alternative transmission functions, and showed that "... many combinations of transmission functions, latency time, and transmission coefficient were successful in reproducing the details of a cluster of CWD cases …".

Based on existing data and assumptions on disease transmission and animal movements, the Cary model estimated the time of establishment of CWD prior to observation (7 to 15 years), and projected specific rates of spatial spread of the disease (1.6 to 3.7 miles per year). Under a range of model assumptions, the Cary model concluded that harvest of sufficient intensity to remove the majority of infected animals prior to death by disease could effectively stem the spread of CWD, and perhaps eventually result in disease eradication. Such specific conclusions could not be derived from the more general structure of the G-M model, but these conclusions also required assumptions on animal and disease behavior that still need to be verified.

Questions addressed / model predictions:

Box 3.1 Density Dependent vs. Frequency Dependent Transmission

A major challenge in modeling disease is determining the functional form of the equations that most appropriately represent disease transmission. Until fairly recently, most disease models represented interactions between hosts and pathogens as random encounters, where the likelihood of contact was directly proportional to host density (McCallum et al. 2001), also called mass-action transmission (de Jong et al. 1995 propose alternate terminology). This 'random mixing' model is generally described as a density-dependent (DD) transmission, and the rate of disease transmission increases with host density. By contrast, another large class of models treats disease transmission as a function of the proportion of infected individuals in the population (the disease prevalence) rather than host density. This mode of transmission has often been described as frequency-dependent transmission (FD). This distinction has implications for disease control (Anderson and May 1991; Lloyd-Smith et al. 2005a, b).

The choice of DD or FD disease transmission can lead to key differences in the behavior of models under some conditions.. A particularly important difference is that simple models with DD transmission exhibit a population threshold density, below which a disease cannot invade or persist (Anderson and May 1979). This model prediction is the theoretical basis for using population reductions be used to control or eradicate disease.

In contrast, the efficiency of disease transmission in FD models can remain high even when population densities are low, and simple FD models generally do not exhibit a lower population threshold below which disease fails to persist (Getz and Pickering 1983). FD models better represent disease transmission in social animals, where group size and contact rate between individuals is determined more by social behavior than by random encounters between individuals. Among published studies, FD models were most often used to model sexually transmitted diseases.

It seems likely that many diseases, including CWD, will exhibit behaviors consistent with DD transmission, at least at extremely high and low densities. However, as Swinton et al. (2002) noted, predictions from DD and FD models are identical when host densities remain the same. In addition, McCallum et al. (2001) concluded that there is little support for DD transmission among wildlife studies. Field observations of CWD prevalence, and estimates of host density, are currently too imprecise to distinguish predictions from models with DD or FD transmission.

Nonetheless, Schauber and Woolf (2003) criticized Gross and Miller's (2001) CWD model and model interpretations, focusing on the sole representation of FD disease transmission by Gross and Miller. Under conditions more relevant to management of cervid populations (i.e., moderate to high population densities and low to moderate disease prevalence), the behavior of models with DD and FD transmission will be indistinguishable when compared to field data. We constructed a set of simple disease models suitable for simulating CWD to demonstrate this point. Following Anderson and May (1979), we represented a population as consisting of susceptible (S) and infected (I) individuals. In this case, a S individual has not been infected and is neither influenced by disease nor can they transmit disease. Animals infected (I) with disease can transmit the disease and exhibit a higher death rate.

Following Anderson and May (1979, 1991), both population dynamics and disease dynamics can be represented by a simple set of equations:

$$\frac{dS}{dt} = b(S + (1-e)I) - \alpha S - mS$$

$$\frac{dI}{dt} = \alpha S - (m + \mu)I + beI$$

Box 3.1 Density Dependent vs. Frequency Dependent Transmission
(continued from last page)

with the variable definitions and initial values as shown in the table below.

Parameter	Initial value	Description
b	0.4	Offspring per individual per year
α		Transmission rate (see functions below)
m	0.10	Proportion of population dying each year
μ	(fit)	Proportion of infected individuals dying from disease each year
β	(fit)	Transmission parameter (see functions below)
S	800	Number of susceptible animals
I	10	Number of infected animals
N	810	Total number of animals ($S + I$)
K	2000	Number of animals where birth rate is zero (density-dependent population parameter)

α = transmission rate, where:

$\alpha_{DD} = \beta I$ for density-dependent transmission,

$\alpha_{FD} = \beta I / N$ for basic frequency-dependent transmission, and a form (α_{MM}) proposed by McCarty and Miller (1998) and used by Gross and Miller (2001) in their CWD model.

$$\alpha_{MM} = 1 - \left(1 - \frac{1}{N}\right)^{\beta I}$$

For the purposes of the simulations, parameters are per year and as above.

To control population size, density dependence in reproduction was represented in the form of $1-(N/K)$, thus the final equations are:

$$\frac{dS}{dt} = b\left(1 - \frac{S+I}{K}\right)(S+I) - \alpha S - mS$$

$$\frac{dI}{dt} = \alpha S - (m + \mu)I$$

* Note: vertical transmission rate (e) was set to zero and hence eliminated from the above equations.

No field studies have unambiguously documented the existence of host population thresholds (Lloyd-Smith et al. 2005a). While the existence of a lower population threshold is of theoretical interest, many models of wildlife disease exhibit very similar behavior over a broad range of host densities. Field observations from free-
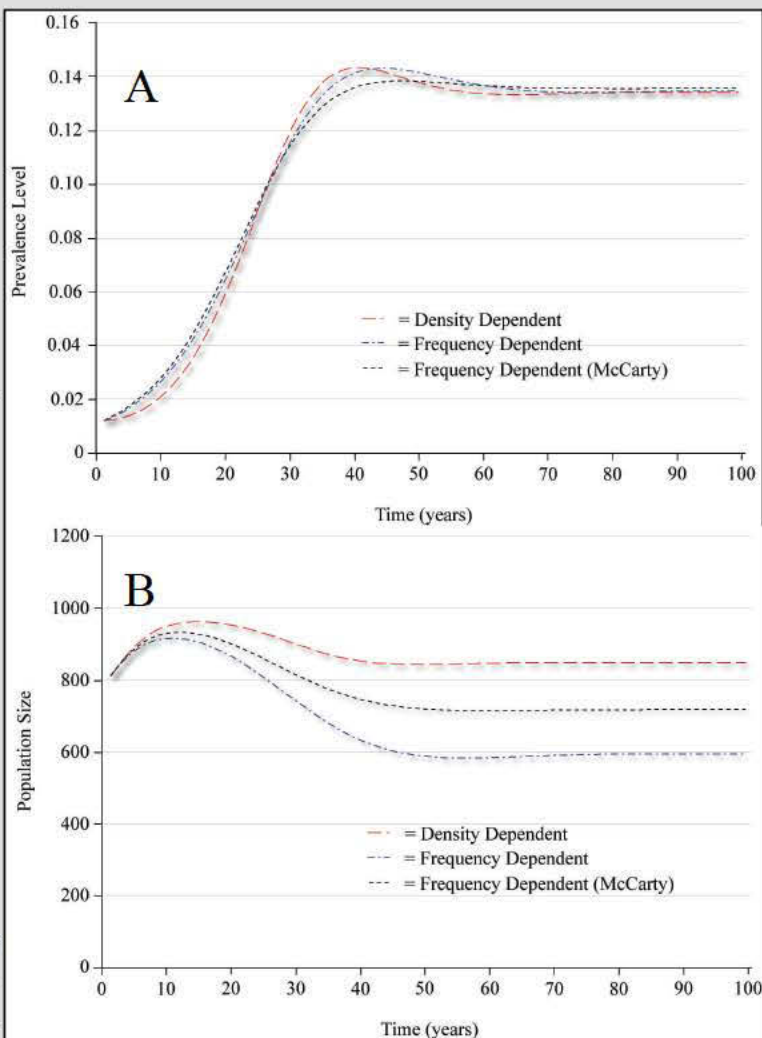
Comparison of (A) prevalence and (B) population size for density dependent, density independent, and McCarty and Miller (1998) density dependent functions of CWD transmission rates.

56   *Approaches to Fine-Scale Modeling*

Box 3.1 Density Dependent vs. Frequency Dependent Transmission
(continued from last page)

ranging populations are simply not adequate to distinguish dynamics produced by frequency- or density-dependent transmission over broad (realistic) ranges of host density. Key differences in disease dynamics related to mode of transmission occur when host population density declines below the threshold in a DD model, and these dynamics have very important implications on the ability of a control strategy (especially population reduction) to eradicate disease. Additional data on transmission dynamics are required to realistically evaluate the likely effectiveness of intensive culling as a management strategy for CWD (Gross and Miller 2001, Schauber and Woolf 2003). We encourage future modeling efforts evaluate the sensitivity of their results and conclusions to this assumption.

1. Predicts disease dynamics, (e.g., rates of change in infection and duration of epidemic) within a population through time.
2. Predicts effects of population sex and age structure on disease dynamics.
3. Predicts effects of control strategies, which may include test and cull, population reduction, habitat manipulation, harvest strategies, and/or vaccination.
4. Predicts effect of individual variation on factors such as genetic resistance, transmission rates, and movement.
5. Predicts effects of spatial structure of the environment on disease transmission and/or persistence.
6. Predicts effects of social structure on disease dynamics and the effectiveness of control strategies.

Data required:

The data required varies with level of detail and intent of the model. For a theoretical model, existing observations of population structure (e.g., proportion in observable age-sex categories) and disease prevalence may all that is required. For a highly complex model with population and spatial structure, detailed data on population age and sex composition, disease prevalence, and on animal movement and contact rates may be necessary before there is sufficient confidence in model results to influence management decisions. It is very easy to construct overly-complex IBMs. Considerable thought should be directed toward constructing simple, tractable models; that is, models with the smallest possible number of estimated parameters.

Output:

Model outputs vary with model structure and level of detail, but virtually all IBMs will simulate population structure and disease state (e.g., susceptible, infected, infectious) through time for defined sex and age classes. Model outputs will typically include harvest and treatment variables, such as the number of animals vaccinated, tested and culled, or harvested. Outputs of spatially explicit models will include the location of all animals, which are usually used to estimate animal densities across the landscape. Any other simulated variables of interest can also be produced, including physiological state, genetic composition, and number of offspring. These variables permit calculation of many other factors of ecological interest such as generation time, indices of genetic diversity, gene flow rates, etc.

General usefulness:

IBMs have proven to be highly useful for simulating a wide variety of situations in animal ecology. They are routinely used in applications that share many characteristics that apply to CWD: where migration or dispersal of individuals is important to establishing new populations or transmitting disease, for simulating changes in genetic composition, and for evaluating the consequences of behavioral differences of individuals and species. Grimm and Railsback (2005) provide many other examples. Although run separately here, IBMs have often been run on top of a grid, where the environment is described by the attributes of the grid cells as is discussed in the Spatial Stochastic Model section below.

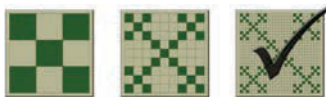***Usefulness to CWD modeling and/or management:***

This approach has great potential and the first CWD models constructed were IBMs (Gross and Miller 2001; Cary 2004). The level of detail is easily varied to accommodate species or site-specific characteristics. A drawback to using an IBM is that the models must be evaluated using numerical rather than analytical techniques, which can be quite time consuming. Model evaluation should include a carefully conducted sensitivity analysis. Complex IBMs can generate very large quantities of output – sometimes measured in giga-bytes – in which case data reduction, analysis, interpretation, and communication can be significant challenges.

There are a number of CWD-specific questions that individual-based model are particularly well suited to addressing. For example,

- How does individual variation in propensity to disperse affect the efficiency of management activities to control CWD?
- How do individual social behaviors (e.g., fidelity to a family or other social group) affect disease dynamics and CWD control strategies?
- How does genetic variation in resistance to CWD affect disease and population processes, including changes in gene frequencies, disease dynamics and population growth and persistence?

NETWORK MODELS

 Early models of disease often assumed that the host population was homogeneously mixed (Anderson and May 1991). In other words, each individual was equally likely to contact every other individual within a single unit of time. Because this assumption obviously does not hold for many human or wildlife situations, many studies have used different methods of accounting for spatial or social structure (e.g., Swinton 1998, Keeling 1999, Keeling and Gilligan 2000a, b, Thrall et al. 2000, Park et al. 2001,

Fulford et al. 2002, Hess et al. 2002, Keeling and Rohani 2002, Cross et al. 2004, Hagenaars et al. 2004). Network models represent a very flexible method of capturing different social/ spatial structures (Keeling 1999, Watts 1999, Newman 2002, Cross et al. 2004, Ferrari et al. 2006). Traditional models typically assume that an individual's risk of infection depends upon the prevalence or density of infectious individuals in the local (or global) population. Network models, on the other hand, explicitly incorporate information about who is connected to whom and then assess each individual's infection risk according to the number of contacts they have with infectious individuals. These models have been primarily used for sexually-transmitted infections where the contacts among individuals may be limited and variable. The strength of the network modeling approach is its flexibility to represent a wide range of social or spatial structures. Contact networks may change over time, but due to the lack of empirical data on network structure and how it changes over time, most network models have been static (Keeling 1999, Watts 1999, Read and Keeling, 2003). Ferrari et al. (2006), however, illustrated how the contact network could evolve over time as individuals become infected and removed by a disease. In particular, the most well-connected individuals are infected first, leaving a much more sparsely connected network of susceptible individuals that are less likely to be contacted and infected.

A matrix of pairwise contact probabilities often underlies these models. This association matrix is filled with association indices (a_{ij}), which describe the amount of contact between individual i and individual j. These association indices can then be multiplied by infection rates or probabilities to simulate the disease dynamics. Keeling (1999), and Keeling and Grenfell (2000) used contact networks to extend SIR models to structured populations. They found that inclusion of spatial heterogeneity and social structure provided predictions of R_0 that were more concordant with empirically derived estimates than models that excluded these factors.
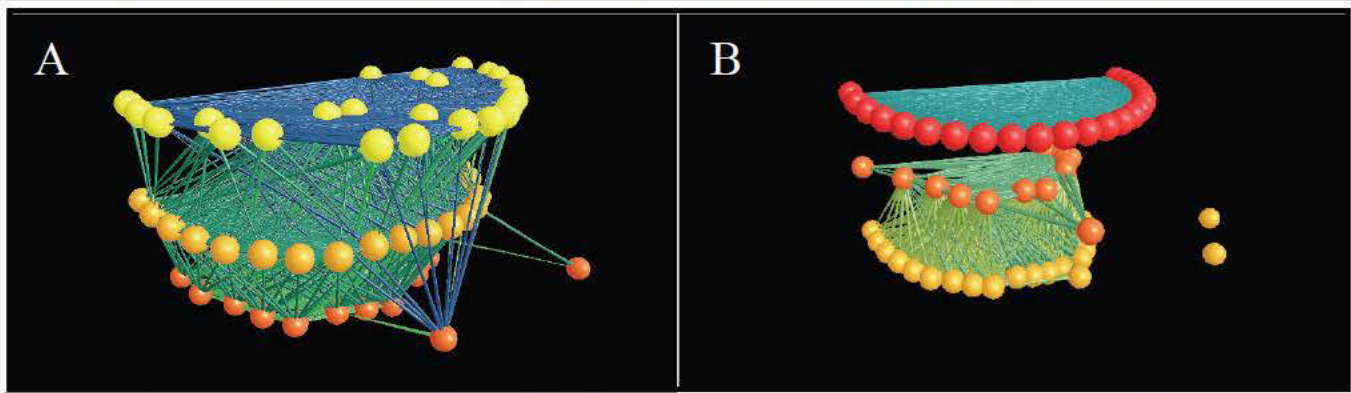
Figure 3.3. *Network graphs of African buffalo association data for (A) November 2001 through October 2003 and (B) May 2002. Balls represent radio-collared buffalo and lines indicate that two individuals spent some portion of the time interval within the same group. Note that the data collected over a short window of time indicates that one herd was completely separated from the others (B). Over the course of the study, however, movement among herds creates a well-connected network (A). Figure from Cross et al. (2004).*

The study by Cross et al. (2004) was primarily for heuristic purposes, but to our knowledge, is the only study in a wildlife system to use a dynamic network modeling approach based on associations between individuals from different population groups (Figure 3.3). The study showed that the dynamic properties of the network were particularly important for acute infections where the disease may go extinct within a local group prior to any connections forming between groups. For chronic diseases like CWD, the network structure connecting different groups may be of minor importance because disease persists for a long time relative to the frequency of new connections developing between groups. Consequently, disease could readily move from one group to another regardless of the network structure. The major hurdle to applying this approach is the difficulty of estimating associations between individuals and then scaling those estimates up to create an appropriate network structure that accurately reflects the entire population of interest.

Spatial heterogeneity or social structure in a population will reduce the spread of a disease when the number of long-distance connections is relatively low. The existence of either factor violates the assumption of homogeneous mixing and this invalidates the estimation of R_0 by many epidemiology models. If populations are not homogeneous, then theoretical estimates of R_0 exceed, often greatly, the observed R_0 (Keeling 1999). R_0 is often a very poor predictor of disease invasion in spatially or socially structured populations where the local group size is small (Figure 3.4) (Ball et al 1997, Cross et al 2005a). Even if R_0 is high and the disease easily invades the local group of individuals, R_0 does not inform us about the likelihood of continued spread of the disease to other groups. Group-to-group transmission of a disease depends upon the movement rate of hosts and parasites and the persistence of the parasite within the local group (Cross et al 2005a, b), assuming no environmental transmission.

Questions addressed / model predictions:
1. Predicts R_0 and R_t and ensuing disease dynamics (speed and duration of epidemic) within a population through time.
2. Estimates effects of population structure on disease dynamics during the duration of the infection.
3. Indicates "core groups" that are likely to harbor disease and where management efforts may be focused.

Data required:
1. Estimates of interconnectedness or association of individuals or small groups, such as territory members, (i.e., population struc-

Approaches to Fine-Scale Modeling
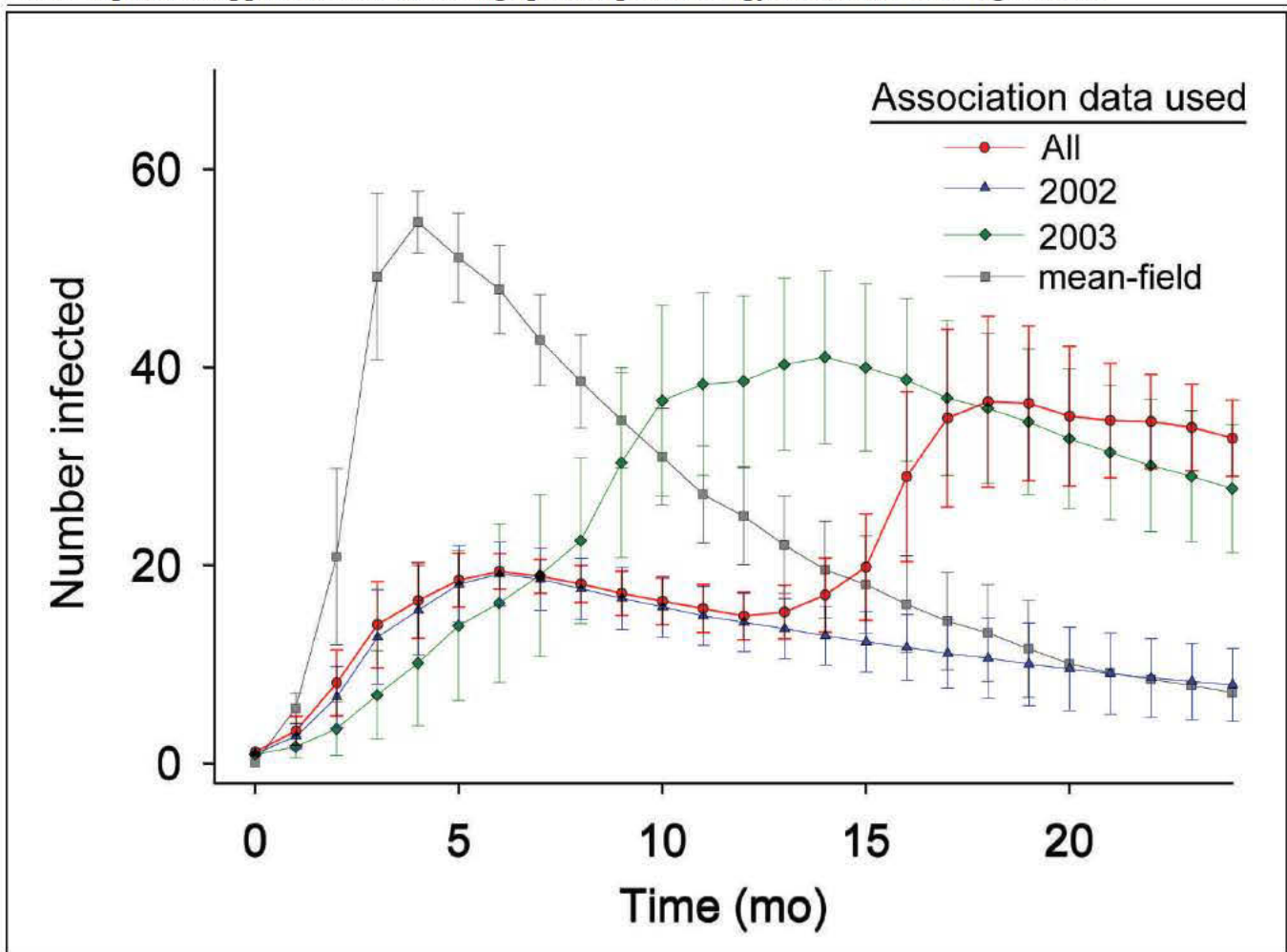
Figure 3.4. *Mean and standard deviations of the number of infected individuals over 50 runs of the disease model using monthly association data from the entire study period (red circles), 2002 (black triangles), 2003 (white squares), or a mean-field model (blue circles) where all individuals were connected but the monthly force of infection was the same. All simulations used a transmission coefficient β of 0.03 and a recovery probability γ of 0.1. Figure from Cross et al. (2004).*

ture), preferably at each time step of interest.

2. Infection status of study individuals (susceptible, infected, or recovered) at each time step of interest.

Output:

1. Predicts total number of individuals that will become infected during the course of an epidemic and number of individuals infected through time (time trace of the epidemic).

2. If data is collected through time, estimates of variance in population structure.

General usefulness:

This approach has great potential because of its flexibility to simulate many different spatial/social structures. However, its utility is likely to be limited in many wildlife disease systems by the lack of individual data on associations between individuals. Continued improvements in radio-tracking and GPS technology will make these data more available, but several questions remain that limit the general utility of contact network modeling. This approach can uniquely address questions such as: "How do we efficiently sample a network?" Then, given that sampling, "How do we scale up the sample so that it represents the entire network of interest?" To the authors' knowledge, these questions, crucial to network mod-

eling, have not yet been answered for any human or wildlife disease system.

Usefulness to CWD management:

Given the potential for environmental transmission and chronic nature of CWD, network modeling may be of limited utility for CWD management. In the case of CWD it will be very difficult to define who contacts whom, particularly when the infectious agent may persist in the environment for several years (Pálsson 1979, Miller et al. 2004). When environmental contamination is significant, the network of contacts between live individuals may be far less important in determining disease dynamics.

SPATIAL STOCHASTIC MODELS
(FOLLOWING SMITH ET AL. 2002)

Spatial stochastic models are called spatial because each individual is explicitly located in space, usually on a grid cell representing a small (e.g., territory) to large (e.g., county) area, and stochastic because rules of movement and vital rates are chosen randomly from a distribution or bootstrapped from the data. Spatial stochastic models can operate at a variety of scales; at a landscape scale space may represent the summary of disease cases or prevalence throughout an area, such as a wildlife management unit, or for an entire population (for details, see the landscape-level section on cellular automata models). However, at a fine scale, grid-based spatial stochastic models are used to model disease epidemiology within a single population, or in a relatively small area, such as winter range. Grid-based spatial stochastic models can incorporate characteristics of other model types, and the size of grid cells can be defined to represent an area that might contain one or a few animals (perhaps a family or social group), or an entire population. At the fine scale the data is usually more intensive and at finer resolution. Birth and death rates, gene frequencies, and rules for movements between cells are often required,

while at larger scales data such as time of first infection in an area can suffice. The question of interest and biology and ecology of the relevant animals dictate the scale of the model.

More complicated grid-based models can be designed to directly ingest information from a GIS to characterize cell properties that can include elevation, vegetation type, food availability, or cover. These attributes can then be used to estimate habitat quality or the ability of areas, represented by cells, to support growth or persistence of organisms. Grid cells can be organized into multi-cell units to represent territories, and population densities based on areas (e.g., contiguous grid cells or 'patches') that contain a suitable mix of habitat types. Models to simulate population processes (e.g., birth, death, movement, etc.) and disease dynamics are run over the grid of cells, and suitable metrics can be extracted at any desired level – by cell, groups of cells, across a given area, or for the entire population. Model results can be interpreted as non-spatial (e.g., total number of individuals), but the strength of the approach is to investigate the effects and consequences of spatial patterns or heterogeneity.

Cary's (2004) CWD model is an example of a spatial stochastic model. However, because it was described in detail in the IBM section we do not discuss it here. Instead, we discuss 2 other spatial stochastic models used to model disease spatial epidemiology. Smith and Harris (1991) used this approach to evaluate the efficacy of different control strategies on the spread of rabies in urban foxes in a city in southern England. They subsequently applied their model to several other cites in southern England. They did not use underlying habitat or environmental factors to predict population density, but rather modeled a range of typical fox densities based on data from similar areas in southern England. Although this model was spatial, it was not explicit in that fox densities/territories were not related to particular physical locations. Fox territories were represented by an appropriate number of grid cells; territo-

Approaches to Fine-Scale Modeling

ries were smaller at high densities and larger at low densities. Similar to a cellular automata approach, for every time step the density of foxes in and near each cell determined dispersal rules, probability of encounter, and home range size at the next time step. These model outputs were then used to calculate the number of foxes infected with rabies and ultimately to depict the spatiotemporal dynamics of the disease under various control strategies. Note that rules were not static, but they varied with fox biological season.

In their model of the spatial dynamics of parapoxvirus disease in red and grey squirrels, Rushton et al. (2000) provide a good example of the nexus of an explicit spatial and individual based model approach. In this spatially stochastic model, the landscape was represented by 25 m^2 cells, where each cell was classified by proportion of different habitat type that was relevant to squirrels. Remote sensing data was used to define the habitat type of each cell, for a particular location in England, making this an explicitly spatial model. From the amount of contiguous habitats, potential densities of red and grey squirrels were calculated, and then dispersal and competition rules determined the relative densities of the two species for each cell or group of cells. An individual based epidemiologic model was run on top of this spatially explicit population model in which population density determined rates of encounter with infected individuals and the likelihood of becoming infected. These dynamics predicted the number of infected individuals of each species, for each cell and time step.

Because the class of spatial stochastic models includes models that are very general to those that are highly detailed and complex, models of this type can be used to address a huge range of questions. The level of detail, spatial and temporal resolution, and inputs and outputs can be adapted to the specific questions of interest. Because this class of models includes such a broad range applications, we do not address these categories below.

General usefulness:

Spatial stochastic models are useful for evaluating population dynamics where spatial heterogeneity is important. In general, they are used for scenario, or 'what-if', analyses because the amount of data required to accurately estimate model parameters usually exceeds what is available. Consequently, error propagation is a serious issue, and confidence intervals on outputs may be so large that estimates are not useful in themselves. The primary value of highly detailed spatial models is usually the ability to compare the relative value of various management scenarios. Any model validation that does occur is usually at the scale of "was the disease present in this group of cells or not", with observed values compared to predicted values. Finally, it is relatively easy to construct grid-based spatial models using off-the-shelf software.

Usefulness to CWD management:

Spatial stochastic models are useful for modeling CWD and evaluating management strategies where adequate data exists. Cary (2004) used a grid-based representation, over which individuals moved, to simulate CWD in a Wisconsin landscape. For CWD, different movement rules or transmission functions could be included in a spatial stochastic model and results compared to observed patterns of prevalence. This type of approach may provide insight into the function and influence of these types of factors at the scale at which adequate data could be collected.

FOCAL APPROACH:
INDIVIDUAL-BASED MODELS

Individual-based models (IBMs) will clearly contribute to our understanding of the dynamics of CWD and they will likely play an increasingly important role in modeling a wide variety of diseases. We thus present this as a focal method for fine-scale modeling of CWD. As described above, the range of problems that can be addressed by IBMs is vast, and this translates into a similarly large range in the

level of detail and complexity that can be included in any particular model.

We describe the general stages or tasks that a 'typical' IBM project will require. While these steps are described as if they are accomplished sequentially, model development is rarely a linear process. One needs to simultaneously consider model objectives, the types and quality of data available, and there is usually a need to continuously evaluate model objectives, model structure, and model performance.

Step #1: Define model objectives

The first stages of model development are the same for virtually all models, including IBMs. Step one is to clearly define the objectives for the modeling exercise. Model objectives need to articulate the questions that must be addressed, features that are desirable, and the scales of space and time that are relevant to the questions. Will the model be used to support decisions in a specific management area, or is the primary use of the model to understand more general system behaviors? What data are available to estimate model parameters, and to compare to model results? The answers to these questions will help determine model structure and the required types of model outputs. At this stage, it is usually important to consider the tradeoff between model parsimony and realism, and the position along this gradient will likely be constrained by the availability of data.

Common uses for IBMs are to compare the relative consequences of competing management actions, which can include factors that might affect disease transmission or prevalence. One may wish to examine the potential effects of supplemental feeding in harsh years, habitat manipulations, harvest regimes, or test-and-cull of diseased animals. The main purpose of modeling may be to determine the likelihood of achieving a specific management goal or target, or to project population changes or prevalence rates over time and compare results to those in the absence of disease man-

agement. If the intent is to evaluate management actions, the best objectives are quantitative, specific, time-bound, and results are reflected by variables that can reasonably be simulated by an IBM and measured and compared to field observations. In the case of CWD, models can be constructed with specific objectives (1) to evaluate whether our hypotheses about the epidemiology of CWD, as codified in mathematical equations, were consistent with observed disease dynamics (Miller et al. 2006), and (2) to investigate the likely consequences of typical actions to control disease (e.g., Gross and Miller 2001; Cary 2004).

Step #2: Define model experiments

After the key objectives for the model are identified, a related set of model simulations should be defined. For most IBMs, these model 'experiments' will consist of scenarios, based on input variables that define the initial model conditions and the 'treatments' that are to be applied. The universe of potential model scenarios for any IBM is huge and one must define a limited number of experiments that are to be conducted, and the analyses that will be used to evaluate results. For IBMs of CWD, model experiment scenarios might include the proportion of a population examined each year in a test-and-cull program (say, 10%, 25%, 50%, 75%, and 100%), or the harvest rates of adult does and bucks. Because 'treatments' are usually nested and crossed, the potential number of experiments can rapidly become unmanageable. Thus, one should start with clearly defined and listed scenarios.

Step #3: Develop conceptual model

As early as possible, the modelers should develop a conceptual model of the entire system to be simulated (Jackson et al. 2000). A conceptual model generally consists of one or more diagrams of the system, and a narrative that describes key processes and functions. Development of the conceptual model usually helps all involved to more fully identify, articulate, and understand what processes and functions the model needs to include, and, more importantly, what can be left out. In the

process of constructing a conceptual model, knowledge gaps are almost always identified as well as parts of the IBM most likely to be problematic. When developing the conceptual model, it is helpful to very carefully document potential sources of information that can inform model construction and evaluation, from published and other sources.

Many problems with simulation models can be traced to errors in the scheduling of model events. A common error is to produce outputs at an inappropriate time for comparing to field observations. For example, errors in model evaluation can occur when observed prevalence rates of CWD are estimated from animals harvested in the fall, but the model produces prevalence estimates just after birth, a time equivalent to late spring. A well-constructed and detailed conceptual model can help avoid these sorts of errors.

In general, the model development process is to first implement a very simple host demographic population model that includes simple functions for birth and death, as well as an appropriate level of detail on the individuals in the model (typically, the sex and age of each individual). For CWD, approximate vital rates can readily be obtained from the literature for deer (and elk and moose), and population performance of the IBM can be compared and calibrated to observations. Once the basic population model is functioning, more detailed processes can be implemented. Harvest and/or density-dependent reproduction (and perhaps mortality) is typically added next to restrict population size. After this, disease control treatments, genetic inheritance, movement, infection dynamics, or other more complex functions may be incorporated. Additional species may be added so that predation (selective or random) can be simulated, or animal-habitat interactions may be incorporated. Regardless of which features are implemented, it is critical to very carefully examine model performance as each new function is added.

Step #4: Estimate model parameters

Throughout the process of model development, the process of parameter estimation will usually be going on simultaneously. The science and art of parameter estimation is well beyond the scope of this handbook; Hilborn and Mangel (1997) provide an outstanding introduction to the subject.

Step #5: Validate model

Once the model is running, robust, and appears to be operating correctly, it is important to conduct a thorough verification process before proceeding with what are likely to be time-consuming model experiments. Because most IBMs incorporate both stochasticity and complex interactions, model verification can be a difficult and time-consuming process. Interested readers should refer to examples of IBM and more comprehensive treatises (e.g., DeAngelis and Gross 1992; Grimm and Railsback 2005).

Step #6: Run model experiments

For most models, an almost infinite number of model experiments could be conducted. It is necessary to carefully prioritize a limited number of model scenarios that will effectively address the management or heuristic questions. Even with a limited number of scenarios, IBMs are usually capable of producing huge quantities of model output. A core challenge is to reduce and summarize model outputs, and to develop graphics or other summaries that concisely and effectively communicate results to key audiences. The analysis and communication challenges posed by output from IBMs are usually underestimated.

GENERAL CONCLUSIONS ABOUT FINE-SCALE APPROACHES FOR CWD MODELING

There are several key assumptions often made in modeling analyses at the fine scale that, in part, determine model results and conclusions. Of particular importance to disease control and management is the relationship between transmission and host density and/or populations size (McCallum et al. 2001,

Schauber and Woolf 2003). There are few data available to estimate the relationship between host density and CWD transmission, or other factors thought to significantly influence transmission of CWD in free-ranging populations. Data from Caley et al. (1998) and Joly et al. (2006) provide data sets suitable for estimating or inferring changes in contact with variation in host density. Because data on transmission rates are typically sparse, modeling analyses are often forced to assume a particular relationship between host density and transmission rates (Box 3.1). With respect to CWD, key differences in disease dynamics result from the assumed mode of transmission when host population density declines below the threshold, and these dynamics have very important implications on the ability of a control strategy (especially population reduction) to eradicate a disease. Our current understanding of CWD transmission in free-ranging populations is not adequate to unambiguously distinguish dynamics produced by frequency- or density-dependent transmission over broad (realistic) ranges of host density and disease prevalence. Recent studies (Joly et al. 2006) observed patterns of CWD prevalence consistent with density-dependent disease transmission, but the relative roles of different transmission modes are unknown. Additional data on this relationship are critical to determining the likely effectiveness of management strategies for CWD (Gross and Miller 2001, Schauber and Woolf 2003), and we encourage future modeling and field efforts to better understand the epidemiology of CWD and to evaluate the sensitivity of model results to transmission functions.

In addition, several studies suggest a strong role of an environmental reservoir of infection for CWD (Miller and Williams 2003, Miller et al. 2004, Miller et al. 2006). This environmental reservoir should be considered in SIR-type models of CWD as well as several of the other modeling methods we discuss in this text. To do so requires adding an addition compartment or variables to track the amount of infectious material, which could be increased by the presence and death of infectious individuals,

and decreased by the degradation of the prion proteins over time. The inclusion of an environmental reservoir of CWD can have important implications for the effectiveness of different management strategies and the duration required to achieve management objectives.

Finally and most importantly, demographic data for CWD infected versus uninfected free-ranging deer are needed for all methods operating at the fine scale. It is at a fine scale that the basic biology of CWD transmission and its true effects on the vital rates and dynamics of deer populations will be revealed. Consequently, field studies designed to estimate survival and fecundity rates of CWD infected and uninfected deer are needed to ultimately determine the effect of CWD on population growth rate. This, along with field studies of transmission dynamics, including effects of environmental contamination and social structure, are needed to determine the spatial epidemiology and functions of transmission of CWD within and between deer populations. Although we chose IBMs for the focal approach, data needs outlined here will support most of the methods in this section. Used individually or together, compartment models, IBMs, and spatial stochastic models are all needed to fully understand the nature of the spread of CWD and its ultimate effects on deer population dynamics.

LITERATURE CITED:

Anderson, R. M. 1979. Parasite pathogenicity and the depression of host population equilibria. Nature 279:150-152.

Anderson, R. M. and R. M. May. 1979. Population biology of infectious diseases: Part I. Nature 280:361-367.

Anderson, R. M. and R. M. May. 1991. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford.

Ball, F., D. Mollison, and G. Scalia-Tomba. 1997. Epidemics with two levels of mixing. Annals of Applied Probability 7:46-89.

Bartlett, M. S. 1957. Measles periodicity and community size. Journal of the Royal Statistical Society 120:48-71.

Approaches to Fine-Scale Modeling

Bonabeau, E. 2002. Agent-based modeling: methods and techniques for simulating human systems. Proceedings from the National Academy of Sciences. 99:7280-7287.

Caley, P., N. J. Spencer, R. A. Cole, and M. G. Efford. 1998. The effect of manipulating population density on the probability of den-sharing among common brushtail possums, and the implications for transmission of bovine tuberculosis. Wildlife Research 25:383-392.

Cary, J. 2004. Modeling chronic wasting disease in white-tailed deer in Wisconsin. Unpublished manuscript, University of Wisconsin, Department of Wildlife Ecology, Madison, Wisconsin, USA.

Cross, P. C., J. O. Lloyd-Smith, J. A. Bowers, C. T. Hay, M. Hofmeyr, and W. M. Getz. 2004. Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. Annales Zoologici Fennici 41:879-892.

Cross, P. C., J. O. Lloyd-Smith, and W. M. Getz. 2005a. Disentangling association patterns in fission-fusion societies using African buffalo as an example. Animal Behaviour 9:499-506.

Cross, P. C., J. O. Lloyd-Smith, P. L. F. Johnson, and W. M. Getz. 2005b. Dueling timescales of host mixing and disease recovery determine disease invasion in structured populations. Ecology Letters 8:587-595

de Jong, M., O. Diekmann, and H. Heesterbeek. 1995. How does transmission of infection depend on population size? Pages 84-94 *In* D. Mollison, Editor. Epidemic models: their structure and relation to data. Cambridge University Press, Cambridge, England.

DeAngelis, D. L., and L. J. Gross. 1992. Individual-based models and approaches in ecology: populations, communities, and ecosystems. D. L. DeAngelis and L. J. Gross, Editors. Chapman & Hall, New York, New York, USA.

Eames, K. T. D., and M. J. Keeling. 2004. Monogamous networks and the spread of sexually transmitted diseases. Mathematical Biosciences 189:115-130.

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

Ferrari, M. J., S. Bansal, L. A. Meyers, and O. N. Bjornstad. 2006. Network frailty and the geometry of herd immunity. Proceedings of the Royal Society of London B 273:2743-2748.

Fulford, G. R., M. G. Roberts, and J. A. P. Heesterbeek. 2002. The metapopulation dynamics of an infectious disease: tuberculosis in possums. Theoretical Population Biology 61:15-29.

Getz, W. M., and J. Pickering. 1983. Epidemic models: thresholds and population regulation. American Naturalist 121:892-898.

Grimm, V., and S. F. Railsback. 2005. Individual-based modeling and ecology. Princeton University Press, Princeton, New Jersey, USA.

Gross, J. E., and M. W. Miller. 2001. Chronic wasting disease in mule deer: disease dynamics and control. Journal of Wildlife Management 65:205-215.

Hagenaars, T. J., C. A. Donnelly, and N. M. Ferguson. 2004. Spatial heterogeneity and the persistence of infectious diseases. Journal of Theoretical Biology 229:349-359.

Hess, G. R., S. E. Randolph, P. Arneberg, C. Chemini, C. Furlanello, J. Harwood, M. G. Roberts, and J. Swinton. 2002. Spatial aspects of disease dynamics. Pages 102-118 *In* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. The ecology of wildlife diseases. Oxford University Press, Oxford, England.

Hilborn, R., and M. Mangel. 1997. The ecological detective. Princeton University Press, Princeton, New Jersey, USA.

Hobbs, N. T. 2006. A model analysis of effects of wolf predation on prevalence of chronic wasting disease in elk populations of Rocky Mountain National Park. Unpublished report to National Park Service, Fort Collins, Colorado, USA. 9 pages.

Hudson, P.J., A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson. 2001. The ecology of wildlife diseases. Oxford University Press, Oxford, England.

Huston, M., D. DeAngelis, and W. Post. 1988. New computer models unify ecological theory. Bioscience 38: 682-691.

Jackson, L. J., A. S. Trebitz, and K. L. Cottingham. 2000. An introduction to the practice of ecological modeling. BioScience 50:694-705.

Jewell, J. E., M. M. Conner, L. L. Wolfe, M. W. Miller, and E. S. Williams. 2005. Low frequency of PrP genotype 225SF among free-ranging mule deer (*Odocoileus hemionus*) with chronic wasting disease. Journal of General Virology 86:2127-2134.

Joly, D. O., M. D. Samuel, J. A. Langenberg, J. A. Blanchong, C. A. Batha, R. E. Rolley, D. P. Keane, and C. A. Ribic. 2006. Spatial epidemiology of chronic wasting disease in Wisconsin white-tailed deer. Journal of Wildlife Diseases 42:578-588.

Keeling, M. J. 1999. The effects of local spatial structure on epidemiological invasions. Proceedings of the Royal Society of London B 266:859-867.

Keeling, M. J., and B. T. Grenfell. 2000. Individual-based perspectives on R_0. Journal of Theoretical Biology 203:51-61.

Keeling, M. J., and C. A. Gilligan. 2000a. Bubonic plague: a metapopulation model of a zoonosis. Proceedings of the Royal Society of London B 267:2219-2230.

Approaches to Fine-Scale Modeling

Keeling, M. J., and C. A. Gilligan. 2000b. Metapopulation dynamics of bubonic plague. Nature 407:903-906.

Keeling, M. J., and P. Rohani. 2002. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecology Letters 5:20-29.

Kermack, W. O., and A. G. McKendrick. 1927. Contributions to the mathematical theory of epidemics. Proceedings of the Royal Society of Edinburgh 115:700-721.

Kretzschmar, M. 2000. Sexual network structure and sexually transmitted disease prevention – a modeling perspective. Sexually Transmitted Diseases 27:627-635.

Liljeros, F., C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg. 2001. The web of human sexual contacts. Nature 411:907-908.

Lloyd-Smith, J. O., P. C. Cross, C. J. Briggs, M. Daugherty, W. M. Getz, J. Latto, M. S. Sanchez, A. B. Smith, and A. Swei. 2005a. Should we expect population thresholds for wildlife disease? Trends in Ecology and Evolution 20:511-519.

Lloyd-Smith, J.O., S. J. Schreiber, P. E. Kopp, and W. M. Getz 2005b. Superspreading and the effect of individual variation on disease emergence. Nature 438:355-359.

May, R. M., and R. M. Anderson. 1978. Regulation and stability of host-parasite population interactions II. Destabilizing processes. Journal of Animal Ecology 47:249-267.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modeled? Trends in Ecology and Evolution 16:295-300.

McCarty, E. W., and M. W. Miller. 1998. A versatile model of disease transmission applied to forecasting bovine tuberculosis dynamics in white-tailed deer populations. Journal of Wildlife Diseases 34: 722-730.

Miller, M. W., E. S. Williams, C. W. McCarty, T. R. Spraker, T. J. Kreeger, C. T. Larsen, and E. T. Thorne. 2000. Epizootiology of chronic wasting disease in free-ranging cervids in Colorado and Wyoming. Journal of Wildlife Diseases 36:676–690.

Miller, M. W., and E. S. Williams. 2003. Horizontal prion transmission in mule deer. Nature 425:35-36.

Miller, M.W., E. S. Williams, N. T. Hobbs, and L. L. Wolfe. 2004. Environmental sources of prion transmission in mule deer. Emerging Infectious Diseases 10:1003-1006.

Miller, M. W., and M. M. Conner. 2005. Epidemiology of chronic wasting disease in free-ranging mule deer: spatial, temporal, and demographic influences on observed prevalence patterns. Journal of Wildlife Diseases 41:275-290.

Miller, M. W., N. T. Hobbs, and S. J. Tavener. 2006. Dynamics of prion disease transmission in mule deer. Ecological Applications 16:2208-2214.

Newman, M. E. J. 2002. Spread of epidemic disease on networks. Physical Review E, 66.

Pálsson, P. A. 1979. Rida (scrapie) in Iceland and its epidemiology. Pages 357-366 In Slow transmissible diseases of the nervous system: Volume I. S. B. Prusiner and W. J. Hadlow, Editors. Academic Press, New York, New York, USA.

Park, A. W., S. Gubbins, and C. A. Gilligan. 2001. Invasion and persistence of plant parasites in a spatially structured host population. Oikos 94:162-174.

Read, J. M., and M. J. Keeling. 2003. Disease evolution on networks: the role of contact structure. Proceedings of the Royal Society of London B 270:699-708.

Rushton, S. P., P. W. W. Lurz, J. Gurnell, and R. Fuller. 2000. Modeling the spatial dynamics of parapoxvirus disease in red and grey squirrels: a possible cause of the decline of red squirrels in the UK? Journal of Applied Ecology 37:997-1012.

Schauber, E. M., and A. Woolf. 2003. Chronic wasting disease in deer and elk: a critique of current models and their application. Wildlife Society Bulletin 31:610-616.

Smith, G. C., and S. Harris. 1991. Rabies in urban foxes (*Vulpes vulpes*) in Britain: the use of a spatial stochastic simulation model to examine the pattern of spread and evaluate the efficacy of different control regimes. Proceedings of the Royal Society of London B 334:459-479.

Swinton, J. 1998. Extinction times and phase transitions for spatially structured closed epidemics. Bulletin of Mathematical Biology 60:215-230.

Swinton, J., M. E. J. Woolhouse, M. E. Begon, A. P. Dobson, E. Ferroglio, B. T. Grenfell, V. Guberti, R. S. Hails, J. A. P. Heesterbeek, A. Lavazza, M. G. Roberts, P. J. White, and K. Wilson. 2002. Microparasite transmission and persistence. Pages 83-101 In P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. The ecology of wildlife diseases. Oxford University Press, Oxford, England.

Thrall, P. H., J. Antonovics, and A. P. Dobson. 2000. Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. Proceedings of the Royal Society of London B267:1555-1563.

Watts, D. J. 1999. Small worlds: the dynamics of networks between order and randomness. Princeton University Press, Princeton, USA.

Woolhouse, M. E. J., C. Dye, J. F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana, and R. M. Anderson. 1997. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proceedings of the National Academy of Sciences of the United States of America 94:338-342.

Woolhouse, M. E. J., D. J. Shaw, L. Matthews, W. C.

Liu, D. J. Mellor, and M. R. Thomas. 2005. Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule. Biology Letters 1:350-352.