A SPATIAL RISK ASSESSMENT OF HEMORRHAGIC DISEASE IN CATTLE IN ILLINOIS FROM 2005 TO 2011

BY

VARUN GOEL

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Geography in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Clinical Associate Professor Marilyn O’Hara
ABSTRACT

Diseases caused by both the bluetongue (BT) and epizootic hemorrhagic disease (EHD) viruses, referred to collectively as hemorrhagic disease (HD), infect both domestic and wild ruminants in the United States. The viruses are transmitted via biting midges, and EHDV in particular causes severe disease in white-tailed deer annually in Illinois. Cattle can also be infected with these viruses; and although clinical illness is rare, the prolonged viremia in cattle after infection may play an important role in disease transmission. The recent discovery in the United States of the genotype EHDV-6 among dead white-tailed deer, an exotic pathogen responsible for cattle morbidity and mortality in Morocco and Algeria, increases the importance of improved vector control strategies in cattle operations and limiting exposure to wild deer.

Utilizing a landscape epidemiology approach, we analyzed the combined effect of rainfall, temperature, forest patterns, deer density and soil on the risk of HD in cattle in Illinois from 2005-2011 at the county level. Data regarding suspected disease in white-tailed deer and areas of potential overlap between cattle and deer were also included in the analysis. Results from a pooled logistic regression for the time period suggest that suspected HD in deer and temperature were the most significant factors in determining the presence HD seroprevalence among cattle in Illinois. It was found that odds of HD seroprevalence in cattle increase by almost 4 times with the occurrence of potential HD in deer in a particular county, while a 1°C increase in temperature increased the risk of HD seroprevalence by 38%. Even though interpretation of subsequently generated risk maps is limited due to data limitations, the approach serves as a framework to enable finer scale analysis and to conduct vector serosurveys. The application of
landscape genetics could further reveal the characteristics associated with the landscape that assist pathogenic evolution, or conversely act as a barrier to it.
ACKNOWLEDGEMENTS

I would first like to thank my adviser, Dr. Marilyn O’Hara, for her valuable guidance in the development of research goals and her support in pursuing these goals. Her thoughtful comments and time spent editing were invaluable to the completion of this thesis. I would also like to thank her and Sara McLafferty for introducing me to the many ways in which geographic methods are valuable to health research. An extended thanks also to Dr. John A. Herrmann and the Center for One Health Illinois for providing research funding, without which this project would have not been possible.

I am also grateful to the University of Illinois Department of Geography and Geographic Information Science for providing me with support through teaching assistantships. The professors and fellow graduate students not only helped me in forming a thesis topic but also contributed and listened to ideas along the way. Special thanks to Bill Brown for providing both data and technical assistance, and to Paul Shelton from the Illinois Department of Natural Resources who supplied data on deer harvests and deer mortality calls. Lastly, I would like to acknowledge the support of my family and friends for supporting me and providing me the strength to complete my graduate degree.
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION .................................................................................................1
  1.1 Hemorrhagic Disease: History, Distribution and Epidemiology .......................................4
  1.2 Hemorrhagic Disease in the United States ......................................................................6
  1.3 Economic Impact of Hemorrhagic Disease .....................................................................8
  1.4 Vector Ecology and Role of Climate .............................................................................10
  1.5 Effect of Other Factors on HD Transmission ..................................................................13
  1.6 A Systems Approach to Transmission of a Vector-borne Disease Agent: Medical Geography, Landscape Epidemiology and GIS ..........................................................15
  1.7 Objectives ..................................................................................................................18

CHAPTER 2: AN ASSESSMENT OF HD RISK FACTORS AND DATA AVAILABILITY ...19
  2.1 BTV and EHDV in Illinois ...........................................................................................20
  2.2 Cattle in Illinois ..........................................................................................................23
  2.3 Deer in Illinois ...........................................................................................................26
  2.4 Deer-Cattle Overlap ....................................................................................................30
  2.5 HD in Deer in Illinois ..................................................................................................32
  2.6 Vector Approximation in Illinois ................................................................................33

CHAPTER 3: DEVELOPMENT OF A RISK MAP AND MODEL OF HD IN ILLINOIS
  CATTLE .........................................................................................................................37
  3.1 Methods ....................................................................................................................38
  3.2 Results .......................................................................................................................44
  3.3 Discussion ..................................................................................................................48
  3.4 Study Limitations and Proposal for Future Research ..................................................52

REFERENCES ..................................................................................................................56
CHAPTER 1
INTRODUCTION

Changes in land use patterns, the foraging behavior of wildlife, and basic characteristics of production methods for domestic livestock can have a profound and long lasting impact on the balance and dynamics of ecosystem function and health (Cooper et al., 2008). Understanding the spatial gradient that characterizes the locations of wildlife and livestock, the pathogens that affect them and factors that shape the interface among them is crucial for assessing the transmission dynamics of infectious diseases. Pathogens that infect wildlife are particularly problematic. Wildlife population density and health status are rarely monitored well, but can be key elements in a disease system. For example, wildlife may act as a reservoir for a pathogen, maintaining it in a sylvatic cycle that can later spill over and infect domestic livestock or humans. Outbreaks of diseases that affect both wildlife and domestic animals, such as foot-and-mouth disease (FMD) and bovine tuberculosis have been associated with places where wildlife receive supplemental feed or are hunted as sport (Knust et al., 2011; Ward et al., 2007). The contact rates between wildlife and livestock are often a function of shared resources such as the distribution of foraging grounds and water, especially under constrained conditions such as drought (Kock et al., 2003). The focus of this thesis is on an examination of the spatial epidemiology of two diseases of moderate epizootic potential: bluetongue and epizootic hemorrhagic disease. These are collectively known as hemorrhagic disease (HD) and the study area is the state of Illinois. The viral pathogens associated with these diseases are known to spread infection from wildlife hosts to domestic livestock in the United States (Bengis et al., 2002). The analysis reported here considered the geographically variable risk for HD in cattle in Illinois, with particular attention to the possible role of deer in the transmission of the virus.
Bluetongue (BT) and epizootic hemorrhagic disease (EHD) are infectious vector-borne viral diseases. The viruses are transmitted by biting midges belonging to the genus *Culicoides* that feed on the blood of a wide variety of hosts, including both wild and domestic ruminants (Thomas et al., 1974; Greiner et al., 1985; Hoff & Hoff, 1976). The causative agents, bluetongue virus (BTV) and epizootic hemorrhagic disease virus (EHDV) belong to the family *Reoviridae*, genus *Orbivirus* and share many morphological and structural characteristics (MacLachlan & Osburn, 2004). In the United States, these viruses primarily affect white-tailed deer (*Odocoileus virginianus*) and mule deer (*Odocoileus hemionus*), and have been observed annually in the United States since 1966. EHDV in particular, has been the cause of large die-offs among wild white-tailed deer populations (Nettles et al., 1992; OIE, 2010). BTV also poses a high risk for clinical disease in sheep populations while EHDV causes clinical disease primarily in white-tailed deer populations. Cattle can be infected by both of these viruses, and although clinical illness from these viruses in cattle has been rare in the United States, the prolonged viremia in cattle after infection may play a very important role in the transmission of the virus (Metcalf et al., 1992).

Recent expansion of the ranges of serotypes and subtypes of hemorrhagic disease viruses raises concern for increased risk of transmission to U.S. livestock. In the United States, five serotypes of BTV (2, 10, 11, 13 & 17) have historically been considered to be endemic while 10 other exotic serotypes have been found in United States since 1999 (USAHA, 2010). Two serotypes of EHDV (1 & 2), have routinely been isolated from white-tailed deer in the United States while an exotic strain, EHDV-6, has recently been found in many states after first being isolated from moribund white-tailed deer in Indiana and Illinois in 2006 (Allison et al., 2010). Even though these serotypes primarily cause clinical illness in deer, and rarely in cattle, isolated
events of suspected clinical illness have occurred in the past. Also, EHDV-6 has been identified as the cause of morbidity and mortality among both beef and milk cattle in Morocco, Algeria and Turkey. Hence, the potency of the strain along with its subsequent spread in other states such as Missouri, Kansas and Texas suggests the virus may be overwintering and is now endemic in a widespread region of the United States.

Geographically, most HD cases in the U.S. have been reported among states located in a diagonal band extending from the southeast to the northwest corner of the United States, with the majority of BTV and EHDV transmission occurring in southeastern and western United States. Knowledge about the distribution of reported cases is primarily based on wildlife surveillance, with seroprevalence of the disease reported based on states. The state of Illinois is located in a transition zone between an area of high transmission in the south and an area of very little transmission in the north, so it is especially difficult to make precise statewide estimates due to the spatial and temporal heterogeneity within the state.

The research presented here builds on the results of a set of previous studies by Boyer et al. (2007, 2008 & 2010) that described the regional seroprevalence of BTV and EHDV among cattle in Illinois from 2000 to 2002 and measured the impact of climatic and environmental variables on the risk of orbivirus seroprevalence in Illinois. The studies found that temperature, rainfall and forest patchiness were associated with orbivirus exposure rates. The authors used Geographic Information System (GIS) and spatial modeling methods to produce risk maps for the state of Illinois. BTV and EHDV seroprevalence has continued to be reported in cattle in the years since these studies, but the introduction of novel strains such as EHDV – 6 and extreme patterns in temperature and precipitation such as drought conditions in 2007 and 2012 have occurred since that time. These events have had an impact on the geographic distribution of
hemorrhagic disease seroprevalence in cattle. Also, the effect of deer population densities and hemorrhagic disease cases in deer populations on disease prevalence in cattle was not considered by Boyer and colleagues, and has not been studied in other similar peer-reviewed literature. In this thesis, I used GIS and spatial statistical models to assess factors such as deer density as well as those factors included by Boyer et al. (2010) with regard to their impact on the occurrence of hemorrhagic disease virus exposure reported in cattle in Illinois from 2005 to 2011. The result was the development of multiple statistical models. These models revealed the potential environmental factors influencing the pattern of HD seroprevalence in cattle and were then used to create predictive risk maps that may be used for improved surveillance and targeted control of HD among livestock in Illinois.

1.1 Hemorrhagic Disease: History, Distribution and Epidemiology

Bluetongue was first recorded around 125 years ago during the late 18\textsuperscript{th} century with the introduction of European breeds of sheep in parts of southern Africa (Howell & Verwoerd, 1971; Spreull 1905). This introduction resulted in high morbidity and mortality among the non-native sheep populations (Walton, 2004). The disease was first diagnosed in cattle in 1933 and due to its similarity with foot-and-mouth disease in terms of clinical signs, was referred to as ‘pseudo’ foot-and-mouth disease (Bekker et al., 1934). The name “Bluetongue” was first introduced by Spreull (1905), derived from the African word “blaauwutong” or “bloutong”, after Boer farmers noticed tongue cyanosis in severely diseased sheep (MacLachlan et al., 2009; MacLachlan, 2011). In wildlife, bluetongue has been associated principally with clinical infection in susceptible non-African wildlife (MacLachlan et al., 2009). It has been suggested that Bluetongue may have played a role in the disappearance of desert bighorn sheep from Texas.
(Robinson et al., 1967). In most cases, even though BTV has been isolated in cattle, goats and other wild ruminants, the infection is subclinical or asymptomatic. However, some strains such as the BTV serotype 8 currently circulating in Europe have been shown to induce severe morbidity and mortality in cattle and camelids (Darpel et al., 2007; Toussaint et al., 2007).

According to MacLachlan (2011), there currently exist twenty-four distinct BTV serotypes around the world, with the novel Toggenburg virus considered a twenty-fifth serotype (Hoffmann et al., 2008). Clinical disease is most often associated with sheep. However, recent epidemics in central and western Europe associated with the BTV serotype 8 have resulted in morbidity and mortality among cattle populations (Walton, 2004; Elbers et al., 2008). The isolation of BTV serotype 1 from white-tailed deer in the United States suggests that serious disease can also develop among the deer population (Johnson et al. 2006).

Epizootic hemorrhagic disease (EHD) was known among hunters and woodsmen in southeast United States since the late nineteenth century as “black tongue” (Savini et al., 2011). It was first confirmed among white-tailed deer in New Jersey in 1955 with the isolation of EHDV serotype 1 (Shope et al., 1955; Shope et al., 1960). Although primarily associated with morbidity and mortality in wild animal populations (Savini et al., 2011), diffusion of the virus in other areas such as the Mediterranean Basin has resulted in an increased pathogenicity among cattle, associated with serotypes not previously considered to be pathogenic.

According to Campbell & George (1986), eight serotypes of EHDV (1 – 8) have been identified based on serological reactivity; serotypes 1 and 3 could be considered as one (Anthony et al., 2009). Those types known to cause disease in cattle include a strain similar to EHDV-2 called the Ibaraki Virus (Omori et al., 1969), EHDV-6 (Temizel et al., 2009) and EHDV-7 (Yadin et al., 2008; Kedmi et al., 2010). However, it should be noted that pathogenicity of the
same strain varies spatially (Savini et al., 2011). For example, Ibaraki virus has caused disease in
cattle in Japan while the homologous North American strains are not considered pathogenic.
Similarly, the same serotypes that caused disease in cattle in North Africa and Israel (EHDV-6
and EHDV 7 respectively) have not caused disease in Sudan and Australia (Savini et al, 2011).
In general, according to Anthony et al. (2009), “western” strains (geographic origin in Americas,
Africa and Middle East) are more pathogenic than “eastern” strains (geographic origin in Asia
and Australia). Being a vector-borne disease, the distribution of BTV and EHDV is limited to the
distribution of competent Culicoides vectors. Although both viruses are generally found between
the latitudes 35°S and 50°N, BTV has been reported as far as 60°N in southern Norway (Mellor
et al., 2008). Even though both viruses share similar life cycles, reasons behind their different
geographical range have yet to be explored (MacLachlan, 2011).

1.2 Hemorrhagic Disease in the United States

The geographic distribution of EHDV and BTV in North America is based primarily on
the distribution of the vector C. sonorensis (Savini et al, 2011), with only a few known cases in
New Jersey and western Canada that lie outside the normal range of the vector (Chalmers et al.,
1964; Stallknecht and Howerth, 2004). C. sonorensis is one of the confirmed vectors of BTV,
and is the only confirmed vector of EHDV in North America (Jones et al., 1977; Foster et al.,
1977). C. insignis has also been established as a vector for explaining BTV cases in the
southeastern United States (Gibbs et al., 1988). The cases in places such as New Jersey and
western Canada suggests that other species such as C. debilipalpis, C. stellifer, C. obsoletus, C.
paraensis, and C. spinosus may warrant further investigation as vectors (Mullen et al, 1985;
Smith & Stallknecht, 1996). The BTV-2 serotype is restricted to the southeastern United States,
and is mainly found in Florida. The other endemic serotypes (10, 11, 13 and 17) are coincident with the distribution of *C. sonorensis*, while the restricted distribution of BTV-2 in southeastern United States may be explained by its dependence on the species *C. insignis* for transmission (Tabachnick et al., 2011). Since 1999, ten previously unrecognized serotypes (BTV 1, 3, 5, 6, 9, 12, 14, 19, 22, and 24) have consistently been isolated in the southeastern United States, suggesting that these formerly exotic strains may now be endemic in parts of the country (Ostlund, 2010; MacLachlan, 2010).

EHDV serotypes 1 and 2 are the most common types to have been isolated from white-tailed deer in the United States. In 2006, the serotype EHDV-6 was isolated from moribund white-tailed deer in Indiana and Illinois (Allison et al., 2010). The subsequent discovery of EHDV-6 in Missouri in 2007; Kentucky and Kansas in 2008; and Missouri and Michigan in 2009 (Savini et al., 2011) suggests that the virus may be capable of overwintering and has become endemic in a geographically widespread region in the United States (Allison et al., 2010).

EHDV, like BTV, is also known to infect cattle and other domestic ruminants subclinically. Severe cases of clinical illness have been observed in cattle infected by the Ibaraki strain (Omori et al, 1969), EHDV serotype 6 (Temizel et al., 2009) and the Israeli EHDV serotype 7 (Yadin et al., 2008). In North America, apart from a few cases of morbidity (up to 20%) in cattle associated with EHDV serotype 2 (House et al, 1998), clinical disease has not been observed in cattle. Previous sporadic outbreaks of epizootic hemorrhagic disease were suspected in cattle in Oregon in 1969 and in Tennessee and Colorado in 1972 (Metcalf et al., 1992). Also, during an outbreak investigation in West Virginia in 1993, clinical illness was observed in cattle that tested seropositive for EHDV-2, and strains of EHDV-2 were isolated.
from a die off of approximately 200 deer occurring in the same area (Abdy et al., 1999). For both BTV and EHDV, a transient viraemia in cattle exists, (up to 60 days) (Singer et al., 2001; OIE, 2010) suggesting that cattle may play an important role in disease transmission.

In Illinois, seasonal outbreaks of hemorrhagic disease have consistently occurred among white-tailed deer, with cases reported primarily in late summer and early fall. Serosurvey results from Boyer et al. (2007, 2008, 2010) showed that from 2000-2002, seroprevalence for EHDV and BTV in cattle increased from north to south across Illinois. Although EHDV and BTV seroprevalence typically exhibited the same geographical distribution, some localized observations, such as existence of high EHDV seroprevalence along river corridors, did not exist in the case of BTV seroprevalence for the same cattle population (Boyer et al., 2008). Also, while EHDV was seen in low amounts in northern Illinois, in the case of BTV, northern zones did not report any evidence of the virus.

1.3 Economic Impact of Hemorrhagic Disease

According to the World Organization for Animal Health (OIE, 2012) both bluetongue and epizootic hemorrhagic disease are on the list of diseases that are notifiable, due to their potential for rapid worldwide spread and considerable impact on the health of livestock. Although bluetongue had been on the list of notifiable diseases for decades, the introduction of BTV-8 in Europe and the consequent direct and indirect costs related with clinical disease in cattle intensified efforts to control exposure to the orbivirus (Mellor et al. 2008). The precise costs due to direct and indirect economic impact are difficult to measure, but the loss of productivity in cattle coupled with impositions of non-tariff trade barriers to prevent livestock from moving to bluetongue free areas can be significant. According to Tabachnick et al. (2011), during the 2007 BTV-8 outbreak in Germany, approximately 35,000 cattle, sheep and goats were
infected across 20,000 farms in Germany. The indirect costs associated with the inability to trade internationally in countries such as France and Netherlands, which contain a substantial cattle industry, have been estimated to be approximately $1.4 billion (Tabachnick et al., 2011). In the United States, annual losses of approximately $125 million have been reported due to the restrictions on international trade of cattle to BT free countries and associated testing for bluetongue seroprevalence.

Epizootic hemorrhagic disease, although considered to be the most important viral disease affecting white-tailed deer in the United States (Nettles and Stallknecht, 1992), has largely been perceived as a disease of relatively minor importance in cattle (Savini et al, 2011). The recent EHDV outbreaks in cattle in the Reunion Islands (Breard et al., 2004), North Africa (OIE, 2010), Turkey (Temizel et al., 2009) and Israel (Yadin et al., 2008; Kedmi et al., 2010) have raised concern over the potential for heavy impact on the cattle industry, due to production losses and morbidity (Savini et al., 2011). According to Kedmi et al. (2010), during the 2006 EHDV–7 outbreak in Israel, the infection caused an average milk loss of 125kg/dairy cow. Combining the costs associated with loss in milk production to those associated with mortality due to infection, the estimated cost was US $2.5 million with an average loss of US $26.5 per dairy cow. Extrapolating the losses in Israeli dairy industry to the number of dairy cattle heads in the United States would result in even more significant losses.

Hemorrhagic disease can also have severe impacts in captive deer populations, especially through translocation between the non-endemic northern areas and endemic southern areas in the United States. This movement opens opportunities to carry the virus from the south to the north, to spread serotypes to new areas and expose immunologically naive deer to new viruses (SCWDS, 2009). According to the European Food Safety Authority (2009), the introduction of
EHDV by infectious wild animals in transit could be substantial depending on the animal origin and the season of movement. Serotypes of both BTV and EHDV have been isolated from captive deer in various parts of the country including Indiana in 2006 (Allison et al., 2010), Colorado in 2007 (Nol, 2010), and Arkansas, Kansas and Texas in 2008 (Allison et al., 2010). Captive deer, which are of great economic value, can also pose significant threats to wildlife or domestic livestock, especially if kept in confined areas in high densities. Captivity may not only cause more stress on deer resulting in reduced immunity and nutritional deficiency, and hence increase risk of disease, but their overlap with wild habitats may even cause transmission of disease to susceptible free ranging wild populations.

### 1.4 Vector Ecology and Role of Climate

As noted before, it is widely established that *Culicoides sonorensis* is the primary vector of BTV in the United States with the exception of Florida, where BTV is associated with the distribution of *Culicoides insignis* (Wilson et al., 2009; Mellor et al. 2008; Mecham & Nunamaker, 1994; Tanya et al., 1992). For EHDV only *C. sonorensis* complies with vector implication criteria defined by the World Health Organization to label an arthropod as a confirmed vector of an arbovirus (WHO, 1967). These criteria are: 1) the virus should be recovered from non-blood-fed wild-caught specimens, 2) the vector should demonstrate the ability to become infected experimentally, 3) there should be some indication of biological transmission and 4) some field evidence needs to be collected to confirm an association between the infected arthropods and the suitable vertebrate host. This suggests that *C. sonorensis* can be implicated as a vector for the EHDV -1 and the EHDV 2 pathogen in North America (Savini et
(al. 2011). Ruder et al. (2012) have confirmed the vector competence of *Culicoides sonorensis* to EHDV – 7.

According to Carpenter et al. (2011), the temperature of the environment is a crucial factor in determining and limiting the rate at which viruses replicate and disseminate in competent vectors, and therefore determining the geographical and seasonal limits to virus transmissions in temperature regions. Differences in specific temperature requirements for the EHD virus replication and the number of degree-days required to attain sustained transmission infection, as compared to BTV may form the basis for observed differences in the regional and global distribution between the two disease case patterns. Hence, a thorough understanding of the distribution of the vector is crucial to understanding the prevalence and risk of disease due to BTV and EHDV.

In the United States, the peak occurrence of HD is also known to coincide with the peak in vector abundance (Gerry and Mullens, 2000; Gerry et al., 2001). According to Mellor et al. (2000), temperature especially, has a profound impact on virus infection and transmission of the Culicoides vector. Elevated temperatures can increase infection rates, cause faster virogenesis, and result in earlier transmission of the disease. At the same time, very high temperatures mean that vector survival rates will significantly decline. In the case of EHDV-1 the median extrinsic incubation period – the time period needed for the virus to become infectious in a vector, decreases significantly with increase in temperature. Infection rates are proportional to the time spent by the vector at permissible temperatures (greater than 15 degree Celsius). At the same time, vector abundance is thought to decrease with increase in latitude. Hence, in general, frequency of HD decreases with latitude, although the probability of mortality in infected
populations is less where the virus has been endemic for longer periods, as evident by numerous animals surviving infection in southern parts of the US (Williams & Barker, 2008).

Hemorrhagic disease occurs primarily in the late summer and early fall with the number of cases being reduced with the arrival of the first frost (Couvillion et al., 1981; Mellor et al., 2000; Purse et al., 2005). It is likely that higher than normal temperatures during winter allow the virus to overwinter and cases may be observed in the subsequent months. Precipitation, although not having as straightforward a relationship as temperature, plays an important role in Culicoides distribution (Mellor et al., 2000). In a study conducted by Boyer et al. (2010), seroconversion rates of orbiviruses in cattle were negatively correlated with the quantity of precipitation occurring weeks before seroconversion. However, these effects have not been consistently observed over other time periods. The coincidental occurrence of drought and high temperatures associated with EHD/BTV infected areas has raised concerns over the role of drought and high temperature in elevating the rate of infection. It has been suggested that during droughts, wild deer tend to concentrate near water bodies, which are also suitable breeding grounds for Culicoides sonorensis (SCWDS, 2011). In a recent example of this, the unprecedented number of EHDV related cases in white-tailed deer in 2012 has preliminarily been associated with the high heat and drought conditions experienced in late summer and early fall across the United States. It should be noted that due to the limited number of Culicoides vector surveys conducted in the United States and the expenses associated with them, the above-mentioned climatic variables such as long-term temperature and precipitation measures have often been used as a proxy for vector distribution.
1.5 Effect of Other Factors on HD Transmission

In addition to climatic factors, many other environmental factors, both at the global and the local scale, can have a profound impact on the increasing interactions between wildlife and domestic ungulates, assisting disease transmission. Global factors include increased human population and density, urbanization, agricultural intensification, and environmental changes such as species overabundance and habitat fragmentation (Martin et al, 2011). For example, while examining the effect of ecological covariates on the transmission of BTV serotype 8 in France, Pioz et al. (2012) found that along with meteorological variables and elevation, the density of dairy cattle was an important predictor of the velocity of the spread of bluetongue virus. The study found that the higher the density of dairy cattle, the lower the velocity of disease transmission, probably due to the management practices associated with dairy cattle, which unlike beef cattle are both more contained and clustered, thus disrupting host availability for the BTV vector. Similarly, according to Mayo et al. (2012) farming practices, such as the use of wastewater from lagoons for extensive irrigation of crops can increase resident population of *C. sonorensis* among sentinel dairy cattle herds in California.

Other factors such as soil texture and soil moisture have also been found to be good indicators of *C. sonorensis* distribution and larvae growth. It has been suggested that wet soil with high organic content facilitates *C. sonorensis* larval growth and sustains the vector (Kettle, 1962). Highly organic soil that is wet but not under water, or soil with high manure loads in swine, sheep and cattle farming operations is also known to support *C. sonorensis* populations. Although not statistically significant in their analysis, pH and soil texture classes were also used by Boyer et al. (2010) to estimate the effect of landscape factors on orbivirus exposure in cattle in Illinois.
Xu et al. (2012), have suggested the use of more biologically relevant variables such as land surface temperature and normalized difference vegetation index (NDVI) to approximate vector abundance in a region. The NDVI measure ranges from approximately -1 to +1, where negative values denote presence of water, values approaching 0 denote a lack of green vegetation and those approaching 1 denote increasing green leaf density for a given pixel. The value is correlated to soil moisture, precipitation and vegetation productivity, and was associated with some vector species such as *Culicoides imicola* in Spain and Italy (Calvete et al., 2009; Conte et al., 2007). In these studies, the NDVI was low in areas with higher Culicoides abundance. This suggests that the vector favored less vegetated shrub and grassland areas. However, suitable vector habitats can differ by species, as demonstrated by *Culicoides obseletus*, which prefers more densely vegetated habitats (Conte et al., 2007).

Since white-tailed deer are suspected to be the main reservoir of BTV and EHDV in the United States, they play an important role in the transmission of hemorrhagic disease in cattle. Boyer et al. (2010) concluded that their proxy for deer density was a significant variable in the estimation of orbivirus exposure in cattle. However, due to limitations of suitable deer density data for that study, forest landscape variables such as edge density and forest patchiness were used as a proxy for deer density in that analysis. House et al. (1998) also suggested that higher deer density in an area containing cattle would result in higher seroprevalence of orbivirus exposure in cattle.

Measures of deer population densities at fine spatial scales are difficult to obtain. The Wisconsin Department of Natural Resources (2012) suggests the number of deer harvested is a good proxy for estimating deer numbers, and these estimates are usually available at the county level in the United States. In Illinois, Roseberry and Woolf (1998) used remote sensing and
landscape data to estimate deer habitat suitability across Illinois at a spatial resolution of 30 meters. Using reclassified land cover classes from the 1996 Illinois Land Cover Database, they developed a model of suitable deer habitat designed to estimate areas of optimum forage for deer during summer and optimum cover for deer during winters. The amount of deer habitat can also serve as a proxy for deer population potential.

1.6 A Systems Approach to Transmission of a Vector-borne Disease Agent: Medical Geography, Landscape Epidemiology and GIS

Although the importance of place and environment for health was recognized by Hippocrates, those ideas were more recently formalized by Jacques May in 1950 and Eugene Pavlovsky in the early 1900s (Levine, 1964). The term “Medical Geography” was defined by May (1950) as the study of relationships between pathological factors termed as “pathogens” and the geographical factors, termed as ”geogens”. While studying tick-borne encephalitis and tick-borne typhus, Pavlovsky proposed that pathogens might be associated with the entire ecosystem. This concept, where rather than existing only within humans, pathogens exist in nature, circulating between arthropods and wild vertebrates, was defined as the natural “nidality” (or focality) of human diseases Pavlovsky (1966). According to Pavlovsky, infection in humans would occur on encountering the nidus either through contact by an infectious vector or a reservoir host. However, often there is a disjointed distribution of infection and residence, as seen for highly clustered pathogens such as those responsible for tick-borne encephalitis.

Also, as is the case with many vector-borne diseases such as trypanosomiasises, Lyme disease, or many arboviral infections, the study of disease involves looking at a “complex matrix of interactions among multiple populations of hosts, vectors and pathogens” (Kitron, 1996). Landscape ecology, defined as the study of dynamic heterogeneous landscapes, interactions
between those landscapes, and the effect of those landscapes on biotic processes (Kitron, 1996), along with the concept of Pavlovksy’s nidality has been studied by many medical geographers (Meade et al., 1988), giving rise to the field of landscape epidemiology.

Figure 1.1. Model illustrating the concept of nidality in a system where vector, host and pathogens exist within a permissible environment to initiate pathogen transmission (Reisen, 2010).

As Kitron (1996) notes, even though landscape epidemiology provides a more holistic view of the dynamics of disease in a given environment, the issue of scale is a major concern. Spatial scale of disease processes may range from the global to the local scale, or from macro through meso to micro scale, prompting multiple empirical considerations in terms of appropriate databases, satellite imagery and statistical methods to be used (Kitron, 1996).

Another issue to be taken into consideration by epidemiologists is spatial dependence and spatial heterogeneity (Getis & Ord, 1992). While spatial heterogeneity refers to the phenomenon where both the data (such as disease cases or vector density) and the spatial units themselves (such as
administrative boundaries or landcover zones) vary from one location to another (Anselin & Getis, 1992), spatial dependence is based on the understanding that variables located near each other (such as heavily infected hosts or vectors), are more strongly related as compared to the variables distant from each other (Tobler, 1970). With advances in computational intensity and power, many statistical methods take spatial characteristics into account, ranging from exploratory spatial data analysis such as the LISA (Anselin, 1995) or Kulldorf’s scan statistic (Kulldorf, 1997), to inferential models such as spatial regression (Anselin et al., 2006) or geographically weighted regression (Brunsdon et al., 1998).

Advances in computing have also aided the development of Geographic Information Systems (GIS), which are very useful for disease mapping and analysis. Even though the idea of using maps in epidemiology can be traced back to at least 1854 when maps were a central part of John Snow’s discovery of the contaminated Broad Street water pump as being associated with a deadly cholera outbreak in London, GIS technology along with advances in remote sensing and global positioning systems (GPS) has been employed successfully to study many emerging vector-borne disease systems such as West Nile Virus, Lyme Disease and Hantavirus. As Auchincloss (2012) notes, there has been a significant increase in the employment of complex computational techniques involving spatial regression in both frequentist and Bayesian frameworks. These statistical methods, coupled with GIS have helped to not only map disease patterns, but also infer possible causes, and create predictive risk maps to aid surveillance and take appropriate preventive measures.
1.7 Objectives

Based on recent searches of peer-reviewed literature, no similar studies have been found that estimate the prevalence of the EHDV/BTV in Illinois since 2002. This study follows on the previous Illinois EHDV risk study conducted by Boyer et al. (2010), which was based on the association of landscape and climate factors with the serotype EHDV-1 and EHDV-2. Since the time of those data, the discovery of EHDV-6 isolates in dead white-tailed deer in Illinois in 2006 suggests that the virus may be capable of overwintering, i.e. the vector is able to pass through the winter season in north temperate regions (Allison et al., 2010). Also, the spatial distribution of seroprevalence of antibodies against EHDV/BTV in cattle has not been explored with respect to white-tailed deer population distributions and seroprevalence of antibodies against EHDV in wild deer.

The main objective of this study is to determine the ecological factors that affect the rate of EHDV and BTV among cattle in Illinois. The study aims to:

1) Assess the combined effect of rainfall, temperature, forest patterns, deer density, soil characteristics and agricultural practices on the risk of hemorrhagic disease in cattle in Illinois.

2) Assess the relationship between the distribution of HD seroprevalence in cattle and the infections in white-tailed deer in Illinois. Since white-tailed deer are one of the most important reservoirs of HD, especially EHDV, it is hypothesized that counties with HD infections in white-tailed deer would have a higher risk of having HD seroprevalence in cattle.

3) Conduct a risk assessment of HD seroprevalence in cattle for all counties in Illinois, and propose a framework to enable finer scale analysis in counties that may warrant further investigation due to increased risk in cattle or captive deer relative to the predicted probabilities of risk in cattle in Illinois.
CHAPTER 2
AN ASSESSMENT OF HD RISK FACTORS AND DATA AVAILABILITY

In order to assess the relationship between HD seroprevalence in cattle and related landscape and climatic variables, we compiled data from different sources to represent exposure to EHDV and BTV and the important factors related to increased exposure to hemorrhagic disease viruses among cattle in Illinois. This chapter describes each of these risk factors and associated data both in terms of the distribution of the factor of interest and in terms of the strengths and weaknesses of representing the factors of interest with existing data. The statistical analysis for which these data were reviewed was carried out at the level of county for the 102 counties in Illinois. Some of the data were reported and available already at the county level, other data were not, thus requiring processing steps to make them useful for the model. The factors of interest included data about: the viruses, deer and cattle hosts, land cover and other conditions related to the vectors, seasonal weather conditions and the climate of the study area. Since the peak transmission activity of the vector, C. sonorensis, occurs primarily between June and October, a study year was defined as going from June 1st to May 31st the following year. Because exposure of cattle to the virus would occur primarily during the period of peak transmission, but seroprevalence tests are conducted year round, this definition groups all positive test cases following the exposure period in the same year. All factors for which data were available to measure change over time were arranged according to this scheme. Each of the factors is described below along with an overview of the main temporal and spatial patterns of each factor.
2.1 BTV and EHDV in Illinois

Test results for HD seroprevalence in cattle were obtained from tests conducted by the Illinois Department of Agriculture animal disease laboratory, with each individual test result containing information about the county of the animal, the date and the type of the diagnostic test performed, and the result of each test. The data were for all tests submitted to the state lab between the years 2005 to 2011. Since Illinois state regulations require cattle imports/exports to be tested for BTV and EHDV exposure, most of the testing was conducted on the request of farmers mandated to get their herds diagnosed before trading or selling livestock across state borders. Hence, the sampling procedure was non-random, but fairly common, because dispersal of cattle herds is an ongoing process. Also, uncertainties related to space and time classification of tests exist. For example, the county of the test was classified based on the address of the premise owner. For large premises located near county lines, the actual herd locations may cross the county border. Similarly, even though the date of the tests was provided, the date of exposure was not known.

For BTV, the presence of antibodies against this virus was checked using two types of laboratory diagnostic tests: the Agar Gel Immunodiffusion (AGID) test and the Competitive Enzyme-Linked Immunosorbent Assay (cELISA) test. For EHDV, the presence of antibodies was checked using only the AGID test. The official OIE manual (2010) suggests that even though AGID is the most widely used test due to its simplicity and commercial availability, one of the major limitations of the test is its lack of specificity in that it cannot differentiate between the BTV and EHDV serogroups due to antigenic similarity between the two groups and problems of cross reactivity. Hence, serogroup-specific assay tests such as the cELISA are used. Even though both AGID and cELISA tests are highly sensitive (are able to detect the presence of the
antibodies) cELISA has much higher relative specificity as compared to AGID. Currently, cELISA is more widely available commercially than in prior years and is now the prescribed test for BTV for international trade, while AGID is an alternative test. For EHDV, as of 2009, although a cELISA had been developed for testing presence of antibodies against EHDV, this test was not yet commercially available. Given this limitation of the AGID test results, the data for both the viruses were combined as a group for analysis and called HD. For the risk model development this served to remove the possible ambiguity in differentiating between BTV and EHDV.

Data for EHDV and BTV test results by county were mapped for the years from 2005 to 2011, with seroprevalence calculated as the percentage of positive tests for each county (figure 2.1). There were 5497 total tests reported during the study period and 347 of those (6.3%) were positive. For EHDV, 85 of 2552 test were positive and for BTV 262 of 2945 tests were positive. The number of tests per county with tests was highly variable (range of 1 to 2584). In addition, the opportunistic nature of the test data made the percentage of positive tests range widely due to the difference in the number of tests conducted in each county. For example, during the period 2005-2011, Livingston County reported 41 positive cases of tests for HD, so the percentage of positive tests would only be 2.6% (41/1552 tests). In contrast, Brown, Lawrence, Pike and Vermilion county had only 2 tests and 2 positives, resulting in 100% seroprevalence. Yearly distribution and the proportion of seroprevalence in BTV compared to EHDV fluctuated across the time period. The highest HD seroprevalence was observed in the year 2007 with 86 and 71 confirmed positive tests for BTV and EHDV respectively (Figure 2.2, Table 2.1). However, in 2008, the following year, while the number of positive BTV cases increased to 97, only 7 cases of EHDV were observed in the state.
The spatial distribution of HD seroprevalence indicates that 22 counties in Illinois had positive tests for BTV antibodies while 8 counties had positive tests for both EHDV and BTV (Figure 2.1). No county had only EHDV positive tests. The spatial pattern of the positive tests is not strong, but it can also be noted that a majority of the counties bordering the Illinois river tested positive for HD seroprevalence. BTV seroprevalence is more evenly spread throughout the state, but with fewer positives in northern Illinois, while EHDV is more concentrated in southern Illinois, with the majority of cases observed in Jackson and Saline counties. Only three counties: Kane, Livingston and Adams county had at least one positive test for HD prevalence each year in the study period (Figure 2.1).

Figure 2.1. Counties with at least one positive test for EHDV and BTV in cattle in Illinois during the period from 2005 to 2011 (left) and the three counties where at least one positive test was seen for all years during the study period (right).
Figure 2.2. Positive tests/seroprevalence for EHDV and BTV in cattle by year for all counties in Illinois from 2005 to 2011.

<table>
<thead>
<tr>
<th>Virus Type</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTV</td>
<td>230</td>
<td>171</td>
<td>286</td>
<td>407</td>
<td>326</td>
<td>761</td>
<td>764</td>
</tr>
<tr>
<td></td>
<td>(29)</td>
<td>(18)</td>
<td>(86)</td>
<td>(97)</td>
<td>(2)</td>
<td>(9)</td>
<td>(21)</td>
</tr>
<tr>
<td>EHDV</td>
<td>559</td>
<td>470</td>
<td>400</td>
<td>264</td>
<td>235</td>
<td>255</td>
<td>369</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(3)</td>
<td>(71)</td>
<td>(7)</td>
<td>(0)</td>
<td>(1)</td>
<td>(0)</td>
</tr>
</tbody>
</table>

Table 2.1. Summary of tests conducted for seroprevalence of antibodies against BTV and EHDV among cattle in Illinois from 2005-2011. Number of positive tests is represented in brackets.

2.2 Cattle in Illinois

Yearly cattle and calves inventory records for both beef and dairy (as of January 1st of each year) were obtained from the United States Department of Agriculture’s National Agricultural Statistics Service (http://www.nass.usda.gov/). The data were then used to assess the variability of beef and milk cattle distribution and to determine if higher cattle density was related to higher levels of HD at the county level. Illinois primarily contains small beef cattle
operations throughout the state, while dairy cattle operations are observed in more discrete regions of the state (Figure 2.3). The highest concentrations of beef cattle are found in northwestern and western Illinois, with significant numbers also observed in the southernmost counties of Illinois. There are fewer dairy cattle overall, and higher numbers of dairy cattle were located primarily in northwest and southwest Illinois counties, with McLean County, in the center of the state also having notably high numbers. Across the time period, Stephenson County and Clinton County stand out as having the highest cattle inventory in Illinois when both beef and dairy numbers are combined.

During the study period, beef cattle numbers have gone down while dairy cattle numbers have increased. From 2005 to 2012 the average beef cattle inventory was approximately 324,000 heads while the average milk inventory was approximately 83,000 heads (Figure 2.4, Table 2.2). However, during the period from 2006 to 2012, beef cattle inventory estimates dropped significantly, while milk cattle inventory increased.
Figure 2.3. Map of Illinois beef cattle (left) and milk cattle (right) inventory 2005-2011

Figure 2.4. Graph of Illinois beef cattle and milk cattle inventory (2005-2012)
Table 2.2 Illinois beef and milk cattle inventory from 2005-2011 (as of January 1<sup>st</sup> each year)

<table>
<thead>
<tr>
<th>Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>426900</td>
<td>432700</td>
<td>414800</td>
<td>296600</td>
<td>279800</td>
<td>268100</td>
<td>242500</td>
<td>231200</td>
</tr>
<tr>
<td>Milk</td>
<td>68700</td>
<td>65700</td>
<td>64900</td>
<td>97300</td>
<td>96100</td>
<td>95400</td>
<td>87800</td>
<td>88800</td>
</tr>
</tbody>
</table>

2.3 Deer in Illinois

To measure deer distribution, yearly estimates for deer harvest per county were obtained from the Illinois Department of Natural Resources for the time period from 2004 to 2010. It is well established that deer harvest rates are a good indicator for the number of deer in a particular area (Wisconsin Department of Natural Resources, 2012). Hence, white-tailed deer harvest numbers were used as a proxy for the relative number of deer per county. From 2005 to 2011, deer harvest numbers have, on average, been similar across years and within counties with only small increases in the years 2006-2008 compared to the other years during the study period (Figure 2.5 and 2.6). More deer were harvested in the southern and western parts of the state, especially along the river corridors. A noticeably lower number of deer are harvested in the central and northern counties of Illinois. Changes at the county level were somewhat erratic. For example, in both Mercer and Piatt counties, which also contain a relatively large number of deer harvested (greater than 1000), the year 2005-06 witnessed a 25-30% increase in deer harvested, while in subsequent years, deer harvest actually decreased by 5-10% in those counties. Across all counties and the whole time period, there was a 9.8% decrease in the number of deer harvested in Illinois. Having determined that there have been no clear increasing/decreasing trends in deer harvest numbers overall, we have assumed a stable density of deer in Illinois between the years 2005-2011 for the purposes of this analysis.
The distribution of suitable deer habitat as per the model described by Roseberry and Woolf (1998) illustrates the spatial trends in potential for deer populations (Figure 2.7). According to Roseberry and Woolf (1998), areas with optimum forest cover and forage would be most suitable for deer habitat and thus provide information on where deer are likely to be found. Counties in southern Illinois and those bordering the Illinois river contain the largest amount of suitable deer habitat. This model of deer habitat was based on landcover data from 1996. To determine how this estimate of deer habitat was related to deer harvest, we measured the correlation between yearly deer harvests and square kilometers of suitable deer habitat. It was found that for each year from 2005 to 2011, the correlation between deer harvests and suitable deer habitat per county was greater than 0.85. Hence, the strong correlation across the time period suggests that the combination of optimum forage and cover for deer based on Roseberry and Woolf’s data, is strongly related to the number of deer harvested and could be reliably used in the final analysis.

The raising of deer in captivity is a small agricultural activity in Illinois. Data for captive deer herds in Illinois by county for the year 2007 was obtained on a paper map from the Illinois Department of Agriculture and used to create a GIS map (fig 2.8). It can be seen that in 2007, four counties – Jo Daviess, Fulton, Pike and Bond, contained more than 100 heads of captive deer, a significantly higher number as compared to the rest of the counties in the state.
Figure 2.5. Average number of deer harvested in Illinois counties from 2005 to 2010

Figure 2.6. Number of deer harvested by year in Illinois (2005-2010)
Figure 2.7. Suitable Deer Habitat in Illinois

Figure 2.8. Distribution of captive deer herds (2007)
2.4 Deer-Cattle Overlap

The areas of land where deer and cattle have overlapping ranges were considered of particular importance in this study. The addition of deer-cattle suitability data to the number of deer harvests and cattle inventory would help evaluate whether a large percentage of area in a county is suitable for the overlap of the two species. Since it is suspected that wild white-tailed deer may be the reservoir for HD (especially EHDV) in the United States and the source of HD for other animals, the areas where conditions indicate the presence of both cattle and deer may have an increased risk of disease transmission, especially if those areas also happen to be habitats suitable for the vector.

Although a reliable, robust remote sensing based dataset (at a ground resolution of 30 meters) for white-tailed deer habitat was created by Roseberry and Woolf (1998), no comparable models have been published outlining areas with conditions most suitable for cattle production, and the data on cattle numbers are only available aggregated at the county level. In order to approximate areas most likely to contain cattle in Illinois, the ‘grassland herbaceous’ category from the 2007 revised Illinois land cover dataset was used to estimate these areas (figure 2.9). This category not only contained location of areas containing grasslands suitable for grazing, but also the pasture/hay category, shrub/scrub category and the grassland/pasture/non-agricultural category. Since pasture/hay and shrub/scrub are indistinguishable from herbaceous grasslands at the ground resolution of 30 meters for the dataset, all those categories were combined in one. A correlation of 0.54 between the average cattle inventory over the time period and the number of acres of land suitable for raising cattle suggests that the dataset provides some support, if not a strong support for sub-county estimates of cattle locations. Locations of beef cattle are more related to the amount of grassland compared to the locations of dairy cattle in Illinois.
To delineate the area where cattle conditions overlap with deer habitat, the optimum forage and optimum cover classes from the deer habitat estimates were first combined as one class into a separate raster grid and then intersected with the raster grid showing grasslands and associated with cattle conditions. The areal intersection was deemed as the potential area that provides opportunities for cattle and deer interactions or where the species may be close enough to be affected by the same populations of vectors. This intersected raster layer was increased in size by four grid cells (30 m grid size) to account for the uncertainty associated with such remote sensing data, and also to account for the flight range of the vector (figure 2.10).

Figure 2.9. Suitable cattle habitat in Illinois (2007)
2.5 HD in Deer in Illinois

Data on laboratory confirmed cases of HD in deer were not available, but a dataset recording the number of reports of sick or dead deer likely due to EHDV was obtained from the Illinois Department of Natural Resources (IDNR). These data were compiled by state biologists and were based on either citizen reports or those that were observed by the IDNR staff. Notably, since the data included reports or calls from witnesses who happened to encounter moribund or dead deer, counties with large populations such as Cook or neighboring counties were likely to be biased by the large population located in the densely populated Chicago, Illinois region. The calls were included in the dataset based on the subjective judgment of the IDNR biologists, so these data were suggestive of a case of EHDV, and the actual number of deer was difficult to ascertain. Because of this, rather than taking the actual number of deer reports at face value, a 1 was recorded if a county had a deer report/call during the year and 0 if no report/call was recorded. It was observed that during the time period of 2005-2011, about 72% of counties had at
least one call, with the exceptions found in the northwest part of Illinois and some counties in the south. A small majority of these reports (55%) occurred during 2007, which coincided with the year of very high HD seroprevalence in cattle in Illinois.

Figure 2.11. The distribution of likely EHDV-infected deer by county in Illinois

2.6 Vector Approximation In Illinois

Spatial models of *C. sonorensis* abundance based on field data were not available for representing the spatial distribution of the vector in Illinois. Thus, other biotic and abiotic factors as described in published literature, such as temperature, precipitation, soil and wetlands were reviewed in terms of their ability to approximate the relative abundance and distribution of *C. sonorensis* for each county. While it was recognized that the variability of weather would strongly influence the magnitude of the annual, weekly or even daily vector populations, the HD cases were based on exposure to the viruses over a period of time and it was not possible to estimate the time of exposure for any time period of less than one year. The inclusion of
variables related to temperature and precipitation in the HD risk model was thus limited to climatic variables that measure the long-term patterns of these factors.

The availability of spatially modeled weather data from the PRISM climate group in a continuous grid at a 4 km spatial resolution was especially useful (http://www.prism.oregonstate.edu/). The monthly climate normal data for the thirty-year period from 1980 to 2010 were summarized for each county. The maximum temperature normal values and the precipitation average normal values were used to measure the more long-term climatic trends as compared to annual fluctuations that may be caused by extreme weather events (Figure 2.12).

![Figure 2.12. 30 Year annual precipitation normal and annual temperature normal (1980-2010)](image)

The state-level STATSGO data (Soil survey geographic database) provided by the USDA was obtained for Illinois to delimit soil conditions providing suitable conditions for the vector
larvae to thrive. Soil attributes associated with soil texture were mapped at the county level by using an areal weighting scheme whereby the percentage of sand, clay and silt were averaged for each county based on soil mapping units (Figure 2.13). The average pH of the soils was summarized for each county using the same set of mapping units. Soils with the higher levels of clay have the finest particles; silt is of moderate size, while sand in soils results in coarser soil texture. In Illinois, soils with higher clay content are found in the north central counties and in the far southern part of the state. Silt is the dominant component and higher amounts are seen throughout the central part of the state. Higher levels of sand are evident outside of the central counties. Soil pH is higher (less acidic) in the northern and eastern counties.
Figure 2.13. Soil characteristics by county in Illinois.
CHAPTER 3
DEVELOPMENT OF A RISK MAP AND MODEL OF HD IN ILLINOIS CATTLE

To assess the effect of various ecological covariates and the presence of HD cases in deer on HD seroprevalence in cattle, a regression framework was developed. A common statistical assumption while modeling the number of cases of a disease in a region, is that they are drawn from a Poisson distribution (Bivand et al., 2008). Poisson regression models for count data are frequently used in epidemiological studies, with the flexibility of adding an exposure variable as an offset in the model. This exposure variable is used to account for the underlying population at risk to a particular disease outcome, with the expectation that the number of cases is distributed according to the base population. In our study, even though the number of tests and total cattle inventory numbers were available for each county, neither of these provided a good population estimate for the dependent variable: the occurrence of a positive test for HD in cattle. Poisson regression models were not suitable for analysis in this case, as the opportunistic sampling estimates would provide highly unstable numbers.

Hence, the Poisson regression model was discarded in favor of a binary logistic regression model, at the expense of losing the variability in counts of HD seroprevalence cases in cattle. Before finalizing the use of the binary regression technique, machine learning techniques such as Classification and Regression Tree models were explored to achieve potentially better fit for the data. These tree-based models can handle missing covariates, can combine both numerical and categorical data, and do not have assumptions associated with other generalized linear modeling techniques. This approach was primarily used to assist in variable selection for use in the final model, but was not included in the final model as the results did not provide
better predictive power, and were difficult to interpret in relation to the objectives mentioned in the study.

3.1 Methods

The variable for the HD test results was used to create a binary response variable for the analysis with 1 and 0 indicating whether HD seropositive tests among cattle were reported in a particular county or they were not reported in at least one year during the study period. A logistic regression model was then developed to detect factors affecting the seroprevalence of hemorrhagic disease exposure in cattle in Illinois. Logistic regression is a form of generalized linear models, which is suited for modeling a binary response variable. Hence, it is a widely used tool in many epidemiological studies where in many cases, the dependent variable is dichotomous, i.e. it denotes the presence or absence of disease. In the current case, the logistic regression model will help estimate the probability of the presence of HD seropositive cattle in a particular county in Illinois.

Therefore letting Y be the binary response variable, where it is assumed that \( P(Y = 1) \) is possibly dependent on \( \bar{x} \), a vector of independent variables, the goal is to model

\[
p(\bar{x}) = P(Y = 1 | \bar{x})
\]

Even though one could possibly use a linear regression method such as ordinary least squares (OLS), the fitted model may result in estimated probabilities outside [0,1]. Hence the logit model, also known as logistic regression is preferred. Discriminant analysis, a method that can identify factors that discriminate among classes in the response variable, was also considered. The logistic regression model has several advantages over both OLS regression and
discriminant analysis, as it does not make any assumptions of normality, independence and homoscedasticity. The logistic regression equation is as follows,

\[
\log \left( \frac{p(x)}{1 - p(x)} \right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p,
\]

where \( \log \left( \frac{p(x)}{1 - p(x)} \right) \) is called the logit function or just the 'logit'. Hence, the coefficient of each independent variable could be interpreted as the log odds ratio between the odds of each event occurring, with other variables being kept constant (Liao, 1994). In order to calculate predicted probabilities of the dependent variable, the above equation can be exponentiated on both sides and written as,

\[
p(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)},
\]

where \( p(x) \) is the predicted probability of the outcome, and the coefficient of each outcome could then be interpreted as the odds ratio between the odds of each event occurring. This additional equation is useful to create a value that can be mapped for each county to express the risk of HD in that county based on its characteristics.

Because both EHD and BT are infectious diseases whose rate in one area is likely to be related to rates in neighboring areas, parameter estimations in regression models can become skewed due to the presence of spatially autocorrelated error terms or a spatially dependent response variable. Hence, it may be important to account for spatial autocorrelation and include it in the regression models if present. A global Moran’s I test was performed on the county distribution of the binary response variable and on the raw residuals of the model to check whether spatial autocorrelation existed in the model or not (where Moran’s I is significantly positive at p value \( \leq 0.05 \)). The HD variable was not spatially autocorrelated (Moran’s I = 0.018, with k nearest neighbor criteria set to 4). For all our models, Moran’s I of the residuals was
statistically insignificant (p value > 0.05), and hence, no spatial term was included in the final regression models.

Similarly, while dealing with cross-sectional data or panel data over multiple time periods, accounting for autocorrelation due to time can result in less skewed and more stable parameter estimates. For example, in their study of EHDV among white-tailed deer in five states at a county level from 1983 to 2000, Xu et al. (2012) discovered that a county is more likely to have EHDV seroprevalence among white-tailed deer if seroprevalence was present in the previous year. However, in order to incorporate time series in a regression, it is important to have a reasonably balanced dataset, which means that observations stay consistent over the time period. Since most of the sampling for HD seroprevalence in Illinois was opportunistic, many counties that were sampled in one year were not repeated across other years. Thus, the data are unbalanced, making it unsuitable for measuring the effect of time in the regression model. To overcome the instability in the data, sampled counties within the time period 2005-2011 were pooled together as one single set of observations. Hence, although the variation within years was lost, it ensured that there was a reasonable sample size and stability available for regression analysis.

Overall, 56 out of 102 counties in Illinois were tested during the time period (2005-2011). Out of these, 30 counties tested positive for HD seroprevalence while 26 counties were found to have no seroprevalence of HD in cattle. The 56 counties with test results were included in the regression analysis. A set of 16 predictor variables were used in the initial logistic regression model (Table 3.1). All predictor variables were aggregated to a county level.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum temperature</td>
<td>The maximum temperature normal for the time-period 1981-2010. (°C)</td>
</tr>
<tr>
<td>Precipitation</td>
<td>Average precipitation normal for the time-period 1981-2010 (inches)</td>
</tr>
<tr>
<td>pH</td>
<td>Average soil pH</td>
</tr>
<tr>
<td>Sand</td>
<td>Percentage of sand in soils</td>
</tr>
<tr>
<td>Silt</td>
<td>Percentage of silt in soils</td>
</tr>
<tr>
<td>Clay</td>
<td>Percentage of clay in soils</td>
</tr>
<tr>
<td>Deer habitat</td>
<td>Percentage of suitable deer habitat</td>
</tr>
<tr>
<td>Cattle habitat</td>
<td>Percentage of land suitable for cattle production</td>
</tr>
<tr>
<td>Deer-Cattle Overlap</td>
<td>Percentage of potential overlap and interaction between deer and cattle</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>Average number of beef cattle per county (2005-2011)</td>
</tr>
<tr>
<td>Milk cattle</td>
<td>Average number of milk cattle per county (2005-2011)</td>
</tr>
<tr>
<td>Beef cattle density</td>
<td>Average density of beef cattle per county (no. of beef cattle/suitable cattle habitat)</td>
</tr>
<tr>
<td>Milk cattle density</td>
<td>Average density of milk cattle per county (no. of milk cattle/suitable deer habitat)</td>
</tr>
<tr>
<td>wetlands</td>
<td>Percentage of area covered by wetlands</td>
</tr>
<tr>
<td>NDVI</td>
<td>Average Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>Deer calls</td>
<td>Variable indicating whether a potential case of EHDV occurred in deer in a particular county between 2005-2011 (0 : No EHDV reported, 1: &gt;=1 case of EHDV reported)</td>
</tr>
</tbody>
</table>

Table 3.1. Logistic Regression covariates. All variables were summarized or aggregated at the county level.

Scatter plots of each predictor variable and their logits were plotted to check for violations of linearity and to assess the need to conduct appropriate power transformations. None of the variables in logit form were found to violate the assumptions of linearity and hence were not transformed. The predictor variables were further checked for multicollinearity, a statistical phenomena in which one or more predictor variables are highly correlated with each other. Presence of multicollinearity can cause highly unstable and erratic coefficient estimates. Hence, a correlation matrix was created to check and eliminate all variables that had a high collinearity.
(r >= 0.8) in the model. From this assessment, it was found that the soil texture classes (silt, clay and sand) were all highly correlated to each other. Hence, only one class (silt in this case) was included in the model, since it explained the effect of the other two classes as well. Model comparison was evaluated using Akaike’s information criterion (AIC), a widely used model evaluation statistic favoring parsimonious models by penalizing models with more parameters. A lower AIC score indicates a relatively better model as compared to another model, calculated by the following formula

$$AIC = 2k - 2[ln(L)],$$

where k equals the number of parameters in the model and L is the maximum likelihood for the estimated model.

Stepwise logistic regression and ROC (receiver operation curve) analysis were used to determine the set of factors that best predict HD presence in cattle. Overall, the aim was to minimize the AIC score and maximize sensitivity to detecting HD seroprevalence in Illinois. A bi-directional AIC stepwise algorithm was used in the R statistical package (http://www.r-project.org/) to estimate the variables that minimize AIC score for the whole model. This was followed by using a bootstrap technique to validate the prediction results from our regression model. Bootstrap is an internal validation technique, which looks at the performance on the population underlying the sample, as compared to an apparent or external validation technique. Even though internal validation is more difficult to calculate, it gives a more honest estimate as compared to other forms of validation. As an alternative to bootstrapping, the split-sample and the cross-validation approaches are also available to validate prediction results. However, due to the relatively small sample size the bootstrap approach was chosen as the validation procedure for prediction based on our regression results.
In a bootstrapping approach, a random sample is first drawn with replacement from the sample used in the regression model. Although the bootstrap sample size is the same as the number of observations (n) in our regression procedure, some observations may be excluded, while others may be repeated more than once. The regression model is then run on the bootstrap sample, and the goodness of fit is compared to the original sample. The difference in model performance is then used to indicate an ‘optimism’ score. This procedure is then run at least 100 times to obtain a stable estimate.

Another diagnostic test commonly used to determine the accuracy of screening tests to discriminate diseased cases from normal cases is the Receiver Operating Characteristic (ROC) curve analysis (Zweig & Campbell, 1993). The ROC curve is a graphical plot in which the true positive rate (Sensitivity) is plotted against the false positive rate (1 – Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular threshold. A perfect test on the graph (upper left corner) would indicate 100% Sensitivity and 100% Specificity, which means that the test was able to accurately distinguish diseased and normal cases from a sample population. Hence, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test. This accuracy of the test is determined by calculating area under the curve (AUC). Once the AUC is calculated, the optimism score calculated from the previous bootstrap validation method is subtracted from the AUC to attain an adjusted AUC score.

After the best model was determined using the sampled counties, the probabilities of HD seroprevalence in cattle for all 102 counties were calculated based on that model and displayed as risk maps using ArcGIS 10.1 software. The maps were then compared to areas containing high numbers of beef and milk cattle, and high captive deer populations to identify areas where the
potential for HD in domestic animals was highest. In addition, this evaluation was used to suggest areas especially suitable for further analysis to study the dynamics of HD in cattle or to initiate surveys that would reveal more precise dynamics of interactions among vectors and hosts and the variable risk of HD.

### 3.2 Results

Preliminary univariate logistic regression models were run for each individual predictor variable against the response variable (presence/absence of HD in cattle in a county) and the odds-ratio exponential of each individual coefficient was calculated along with their confidence intervals (α=0.1) and corresponding p values (Table 3.2).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds-Ratio</th>
<th>Confidence Interval</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>1.38</td>
<td>1.02</td>
<td>1.91</td>
</tr>
<tr>
<td>Precipitation</td>
<td>1.09</td>
<td>0.95</td>
<td>1.25</td>
</tr>
<tr>
<td>pH</td>
<td>0.70</td>
<td>0.08</td>
<td>5.69</td>
</tr>
<tr>
<td>Sand</td>
<td>1.05</td>
<td>0.96</td>
<td>1.21</td>
</tr>
<tr>
<td>Silt</td>
<td>0.90</td>
<td>0.76</td>
<td>1.01</td>
</tr>
<tr>
<td>Clay</td>
<td>1.09</td>
<td>0.91</td>
<td>1.32</td>
</tr>
<tr>
<td>Deer habitat</td>
<td>1.01</td>
<td>0.98</td>
<td>1.04</td>
</tr>
<tr>
<td>Cattle habitat</td>
<td>0.97</td>
<td>0.91</td>
<td>1.03</td>
</tr>
<tr>
<td>Deer-Cattle Overlap</td>
<td>1.01</td>
<td>0.98</td>
<td>1.03</td>
</tr>
<tr>
<td>Beef cattle</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Milk cattle</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beef cattle density</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Milk cattle density</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Wetlands</td>
<td>1.13</td>
<td>0.91</td>
<td>1.42</td>
</tr>
<tr>
<td>NDVI</td>
<td>1.02</td>
<td>0.98</td>
<td>1.06</td>
</tr>
</tbody>
</table>
These initial results showed that out of the 16 possible predictors to explain the presence of HD seroprevalence in cattle, only 2 are statistically significant at the 0.1 significant level: the maximum temperature normal and the occurrence of a deer reported potentially due to HD during the time period in Illinois. The results suggest that, with a 1°C increase in maximum temperature normal, the odds of having presence of HD in cattle in a particular county would increase by 38%. The deer call variable had the strongest magnitude of effect suggesting that the presence of a sick deer report in a particular county in Illinois, would increase the odds of having HD seroprevalence in a particular county by 3.3 times. It should also be noted that even though the deer call variable was first recorded as a 0, 1, 2, 3, 4 or 5 depending upon the number of years in which a county had a deer call recorded, this was later converted to 0 and 1 only since the multinomial variable did not provide more information to the model.

In the first multivariate model, both of these initially significant variables were included together in a logistic regression model, but in this model, neither of the variables exhibit statistical significance (Table 3.3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds-Ratio</th>
<th>Confidence Interval</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>1.14</td>
<td>0.76</td>
<td>1.72</td>
</tr>
<tr>
<td>Deer calls</td>
<td>2.53</td>
<td>0.74</td>
<td>9.1</td>
</tr>
</tbody>
</table>

Table 3.3. Coefficients of logistic regression with confounding variable

On further exploration of the relationship between the two predictor variables, it was noticed that maximum temperature normal acts as a confounder because it is related to both the
response variable and to the other indicator/exposure variable (i.e. deer calls). Hence, not only is the magnitude of the odds ratio lessened for both variables, the statistical significance disappears. However, it should be noticed that both variables, especially deer calls exhibit a positive relationship with risk for HD seroprevalence in cattle in Illinois.

In order to find the best model fit for prediction, for the second multivariate model, a stepwise algorithm was run to choose a model with the lowest AIC score. The results of the model are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds-Ratio</th>
<th>Confidence Interval</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>Silt</td>
<td>0.85</td>
<td>0.7</td>
<td>0.98</td>
</tr>
<tr>
<td>Deer calls</td>
<td>4.25</td>
<td>1.5</td>
<td>12.9</td>
</tr>
</tbody>
</table>

Null AIC score: 79.63  
Model AIC score: 75.53

Table 3.4. Coefficients from stepwise logistic regression with lowest AIC score

Deer calls emerged again as a statistically significant risk factor, while more silt in soil was a protective factor. This suggests that while controlling for percentage of silt in the county, the possible presence of deer infected with HD would increase the chance of HD seroprevalence in cattle by almost 4 times. Even though silt does not turn out to be a statistically significant factor, it does contribute to improving the model, and its role as a protective factor suggests that every unit increase in the percentage of silt per county would reduce the odds of having HD seroprevalence in cattle by 15%. However, the contribution of silt to orbivirus exposure is not well established from other studies, suggesting that such a result may not be substantively significant.
The lower AIC score suggests that the above model provided considerably better support as compared to the null model. Hence, this model was used to calculate the prediction probabilities for mapping the risk of the presence of HD seroprevalence in cattle for all 102 counties in Illinois. Using a cutoff probability of 0.5, all predicted probabilities >=0.5 were grouped as indicating potential HD seroprevalence (1) and those below 0.5 were grouped as not containing HD seroprevalence (0). Based on this rule, 25 out of the 30 counties from the sample dataset that had at least one positive test were correctly classified as having HD seroprevalence in cattle while 5 were misclassified. However, out of the 26 counties not containing any seroprevalence of cattle in Illinois, only 12 were classified as negative for HD seroprevalence, while 14 were classified as positive for HD seroprevalence in cattle in Illinois. Given the current lack of testing in these counties, and the possibility of high seroprevalence, especially in the southern counties, it is not possible to determine the true status of seroprevalence of HD in cattle. However, some of these counties may be, in fact, positive, but lack testing.

<table>
<thead>
<tr>
<th>Predicted HD</th>
<th>Actual HD (cutoff = 0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HD</td>
</tr>
<tr>
<td>HD</td>
<td>14</td>
</tr>
<tr>
<td>No HD</td>
<td>25</td>
</tr>
</tbody>
</table>

Sensitivity: 25/30 = 83.3 %
Specificity: 1 – 14/26 = 46%

Table 3.5. Accuracy assessment of predicted results to sample

Finally, an ROC curve was produced to illustrate the performance of the prediction. The raw area under the curve was 0.695 suggesting that the prediction model is slightly better than a
classification system based on chance alone. A bootstrap validation approach with 1000 repetitions was carried out and an optimism score of 0.05 was produced. Hence, the bias corrected area under the curve was 0.69. Since the optimism score for predicted probabilities was very low, no adjustments in the prediction model were made and the predicted probabilities computed from the prediction model were used to create risk maps for the state of Illinois.

\[ AUC = 0.69 \]

Fig 3.1. ROC curve based on prediction model

3.3 Discussion

Based on the predicted probabilities from the logistic regression models, a map depicting the probability of risk of HD seroprevalence in cattle for each county in Illinois was produced (Figure 3.2).
Figure 3.2. Predicted probability of HD seroprevalence in Cattle in Illinois. The black outlines denote ten counties with highest populations of a) beef cattle, b) milk cattle and c) captive deer.

The map in general follows the distribution of HD reports in deer in Illinois where most counties in northwestern Illinois and some counties in central Illinois have a less than 50% chance of having HD seroprevalence in cattle. Hence, it conforms to our hypothesis that counties
with HD infections in white-tailed deer would have a higher risk of having HD seroprevalence in cattle.

Even though the percentage of silt soils in a county was not a statistically significant variable in the prediction model, it acted as a small barrier for HD seroprevalence in cattle. Hence, counties such as Jo Daviess or Stephenson in the northwest, or Knox or Warren county further south, all have the lowest predicted probabilities of risk of HD seroprevalence associated with the highest percentages of silt in the state. Similarly, counties in the northeast such as Cook or Will County, or even Alexander county in the south, lie in the highest risk quartile for HD seroprevalence in cattle due to having both the occurrence of reports of EHD infected deer and relatively low percentages of silt as compared to other counties. The dependence on those two variables, deer calls and silt, however results in misclassification in counties such as Rock Island and Kane county, where even though HD seroprevalence was detected, predicted probabilities were very low due to no presence of likely EHD deer reports or low percentage of silt. This simple model at the county level is not capturing all of the variability that would help to explain the HD distribution.

Although the predicted risk map has limited interpretation due to data limitations, the map, and the described landscape epidemiological approach can be used to conduct further investigations of areas that, for example, not only have high risk of HD seroprevalence in cattle according to the model, but also coincide with high numbers of cattle or captive deer herds. Also, narrowing down a focus to particular counties based on the predictive map can help in conducting targeted serosurveys, analyzing vector distribution and assessing deer-cattle interactions at a finer scale for future research which otherwise would not only be very expensive but also be time and labor intensive. For example, to better analyze these factors at a more
detailed level, Pike or Effingham county could possibly be further studied as they are not only areas at high risk of HD seroprevalence in cattle and have likely EHD deer reports, they also have one of the highest cattle inventories (Pike for beef, and Effingham for dairy) along with a higher numbers of captive deer herds in the state. Given the unprecedented number of HD cases reported in deer in 2012 in Illinois and in neighboring states, minimizing interaction between cattle or captive deer and wild deer by targeting specific counties could help reduce economic and cattle losses.

Results suggest that at the county level, even though none of the climatic or environmental variables exhibited any significant associations with the presence of HD seroprevalence in cattle in Illinois in a multivariate analysis, an increase in temperature does appear to escalate the risk of HD seroprevalence by around 40%. Although the temperature in this study is defined on the scale of a 30 year normal, it does conform to previous studies in the literature suggesting the increase in ambient temperature is a possible reason for not only the changing patterns of HD and its occurrence in places northwards where the disease was never observed before, but also vector-borne disease in general. For example, according to SCWDS (2012), the widespread outbreak of EHDV in places such as Cook county, where the epizootic was never confirmed before, has been attributed to the dry and hot summer preceding the Culicoides vector season. However, as shown by the results, temperature acts as a confounding variable in the model due to its high positive association with likely EHD deer reports. Hence, other climatic and environmental variables such as wind speed or humidity need to be examined at possibly finer spatial and temporal scales to be able explain the patterns of HD seroprevalence in cattle in Illinois, especially in areas where the disease has recently been detected.
3.4 Study Limitations and Proposal for Future Research

The data and the methodological approach in the thesis were undertaken under several limitations and assumptions. First, since the given data on HD seroprevalence was aggregated at a county level, accurate point locations of cattle testing seropositive for either BTV or EHDV were not known. As shown by Boyer et al. (2010), even though both EHDV and BTV seroprevalence in Illinois have comparatively similar distributions in cattle, micro-scale heterogeneities have emerged, for example across river corridors. Hence, seroprevalence rates of both EHDV and BTV would be variable across counties. Also, since the tests for HD seroprevalence in cattle were conducted opportunistically, the uncertain nature of tests does not rule out the seroprevalence of HD in untested areas. Missing information on which cattle herds were tested could also play a role in sampling bias, especially if a majority of tests in one county were conducted from a single herd. The age of the cattle was not included in the test report. Without this information, it was not possible to determine the likely time of exposure of the cattle, even for young cattle.

In a comprehensive study conducted by Xu et al. (2012) on the seroprevalence of EHDV in white-tailed deer in the southern US over a span of two decades, it was found that at the county level, spatial and temporal autocorrelation both played a role in explaining the risk of EHDV seroprevalence in a particular county. A county having an occurrence of EHDV seroprevalence in the previous year, and surrounded by an EHDV positive neighbor, would have a much higher risk of EHDV. Also, as explained by Boyer et al. (2010), even though spatial autocorrelation was not a significant factor in estimating orbivirus seroprevalence risk in Illinois among cattle after adjusting for covariates, a more concentrated and less dispersed sample of cattle herds would probably yield some degree of spatial autocorrelation. The spatial clustering
of the *Culicoides* (Diptera: Ceratopogonidae) vector on a local scale in northern Europe (Kirkeby et al., 2013), also suggests that diseases caused by some biting midges may be correlated in space. Hence, a comprehensive vector survey in Illinois along with a targeted cattle herd sampling design over multiple seasons would be useful in determining the extent of spatio-temporal autocorrelation of EHDV and BTV among cattle in Illinois.

The use of more robust variables may explain more variation in patterns of HD seroprevalence. For example, the independent variable of infected deer reports is based on field sightings by bystanders or biologists, and not confirmed clinical studies. Also, counties with higher human population are more likely to witness a moribund deer and report it, while presence of clinical infection in deer in some counties may have gone unnoticed. Similarly, other anthropogenic factors such as irrigation management practices and information about types of cow and calf operations may explain the variation in disease patterns among beef cattle and milk cattle (Mayo et al., 2012). For example, different cattle habitat and movement patterns may influence disease transmission. Cattle farm density for beef cattle was found to be positively correlated with the spread of BTV–4 in central and southern Spain, while milk cattle farm density was negatively correlated, probably due to the confined, intensive spatial structure of dairy farms as compared to the more extensive beef farming systems. The inclusion of such information could help estimate suitable cattle habitat more accurately as compared to the current landcover classification.

The performance and predictive power of the regression was also limited by the small sample size and the unbalanced panel data, where counties tested for one year were not necessarily tested for other remaining years, causing gaps for temporal analysis. Also, since some soil texture classes and climatic variables such as precipitation were highly correlated with
other similar variables, they were removed from the analysis. However, a factor analysis or principal component analysis may be used in the future to aggregate related variables in categories to enhance the explanatory power of the prediction models. Availability of suitably sampled HD seroprevalence data along with fine scale ecological variables would also enable the use of more advanced regression techniques such as Poisson or negative binomial generalized linear models supporting count data. Additionally, decision tree learning techniques such as Classification and Regression Tree could be then used to develop robust models, given large and complex datasets. This would also help develop continuous risk maps in comparison to the ones that are based on county boundaries.

This study provides a preliminary framework to conduct a local scale risk assessment of EHDV and BTV in cattle in Illinois. Utilizing the high risk counties as a starting point for a comprehensive vector serosurvey and sampling from cattle herds over multiple seasons, and making finer scale estimates of deer locations, not only can the spatio-temporal effects be measured, but other ecological variables that are known to affect Culicoides abundance at a micro scale, such as wind speed and humidity, could also be taken into account. This would also enable the use of more advanced techniques such as the combination of landscape genetics and molecular epidemiology. Such techniques would not only help explain local scale spatial variations between BTV and EHDV but also genetic variations among the various serotypes of the two groups, and how they may be correlated with specific characteristics of the landscape (Carrel & Emch, 2013). As the introduction of exotic virulent strains such as EHDV-6 causes greater concern among agriculture communities in Illinois, especially due to the state’s location as a transition zone, such an approach may reveal what interacting population and environmental dynamics allow a virulent strain to circulate and evolve in a landscape. Simultaneously, it may
also reveal spatial characteristics that are barriers to pathogenic evolution and persistence, thus highlighting human, animal and environmental interactions that are beneficial in reducing both current and future risks of orbiviruses-related disease among cattle.
REFERENCES


Nol, P., Kato, C., Reeves, W. K., Rhyian, J., Spraker, T., Gidleski, T., ... & Salman, M. (2010). Epizootic hemorrhagic disease outbreak in a captive facility housing white-tailed deer (Odocoileus virginianus), bison (Bison bison), elk (Cervus elaphus), cattle (Bos taurus), and goats (Capra hircus) in Colorado, USA. Journal of zoo and wildlife medicine, 41(3), 510-515.


United States Animal Health Association (USAHA), (2010). 


Wisconsin Department of Natural Resources, (2012). 

http://whqlibdoc.who.int/trs/WHO_TRS_369.pdf
