2018

Spatial Modelling and Wildlife Health Surveillance: A case study of White Nose Syndrome in Ontario

Lauren Yee
yeex9430@mylaurier.ca

Recommended Citation

Yee, Lauren, "Spatial Modelling and Wildlife Health Surveillance: A case study of White Nose Syndrome in Ontario" (2018). *Theses and Dissertations (Comprehensive)*. 2040.
http://scholars.wlu.ca/etd/2040

Spatial Modelling and Wildlife Health Surveillance: A case study of White Nose Syndrome in Ontario

By

Lauren Yee

A Thesis

Submitted to the Department of Geography and Environmental Studies

In partial fulfillment of the requirements for

Master of Science in Geography

Wilfrid Laurier University and the University of Waterloo

AUTHOR'S DECLARATION

Abstract

Wildlife data is often limited by survey effort, small sample sizes, and spatial biases associated with collection and missing data. These factors can create unique challenges from a surveillance perspective when trying to extract spatial patterns of habitat suitability and disease distributions for conservation and management purposes. This thesis examined data quality from a wildlife health database in the context of spatial analysis of wildlife disease. Spatial analysis of the data to predict habitat suitability of bats and white nose syndrome afflicted bats was examined by using the MaxEnt modelling method. Methods to reduce spatial bias were examined and specific settings and parameters were evaluated and compared to the default settings in MaxEnt. This analysis showed that subsampling of occurrences leads to less informative models and increased similarity to accessible locations, whereas restricting background locations increased model performance and the resulting suitability maps showed new suitable areas not present in models with default settings. Alternative data sources of bat occurrence data were used from the Global Biodiversity Information Facility which showed distributions were more related to climatic covariates than CWHC models, which were heavily reliant on hibernacula locations. This suggests that CWHC records may benefit from supplementation of publicly available occurrence data. This thesis provides a basis for data quality analyses that should be performed on surveillance data before undertaking more complex modelling procedures.

Acknowledgements

I would like to start by thanking my supervisor, Dr. Colin Robertson, for his continued encouragement, patience, advice and valuable research opportunities. Dr. Robertson is an approachable advisor whose level of nerdiness and passion for science and music matches my own. I have learned much over the past few years and had many opportunities to attend various conferences and workshops. I am grateful for the life-shaping experience at COPEH-Canada (Montreal, ON). During the conference I met many individuals devoted to improving human, animal and environmental health. My time at COPEH-Canada not only inspired the work I present in this thesis but impacted the way in which I approach research & thinking, as well. I would also like to thank my committee members, Dr. Jane Parmley and Dr. Craig Stephen, for their expert advice regarding wild animals, surveillance and veterinary and epidemiological knowledge, as well as their patience and thought-provoking discussions throughout this process. I would also like to extend my gratitude to the Canadian Wildlife Health Cooperative for providing the main data set for this thesis as well as the Natural Heritage Information Centre. CWHC's friendly, knowledgeable staff answered all my questions with enthusiastic detail, making it clear how dedicated they are to the wildlife they serve. A very special thank you to Alice Grgicak-Mannion for influencing my decision to go to graduate school, and for her steady guidance & strong female leadership throughout my undergraduate studies. Of course, I cannot forget my peers in the Spatial Lab, who helped troubleshoot R and with whom I have shared many laughs, tears and great company on late nights. The Spatial Lab is an environment of collaboration and support and I am very grateful to have been (and continue to be) a part of it. Finally, I must thank my partner in crime, my amigo and favourite dance pal, Layla Hurst, for always believing in me and encouraging me when I needed an extra bit of reminding. My

graduate experience and research was made possible by the financial assistance of MITAC’s Accelerate Internship, the Department of Geography and Environmental Studies (Teaching Assistantships, Graduate Research Scholarships), the Faculty of Graduate studies (Travel Grants and Scholarships), the Canadian Wildlife Health Cooperative, and Dr. Colin Robertson. My express gratitude to these agencies.

Table of Contents

List of Figures and Tables

1 Introduction

The spread of pathogens and adverse health consequences in our environment is inherently spatially motivated, from mapping disease cases to risk/exposure mapping to complex modelling of future pathogen scenarios and dynamics (Walter 2000). Spatial epidemiology is the study of spatial variation in disease risk or incidence (Ostfeld, Glass, and Keesing 2005); where relationships between health outcomes, demographics, environments, socioeconomics, genetics and infection risk factors are examined in a spatial context (Elliott and Wartenberg 2004). Emerging infectious zoonoses (diseases that spread from animals to humans) are particularly sensitive to spatial factors, given the linkages inherent between humans, animal hosts, vectors, and the environments they share. As such, spatial epidemiological analysis has played important roles in many recent zoonoses research such as severe acute respiratory syndrome (SARS) (Lai et al. 2004), Ebola (Judson et al. 2016), Nipah Virus (Peter Daszak et al. 2013), and West Nile Virus (Brownstein et al. 2004; Mostashari et al. 2003). Geographic information systems (GIS) have been used in the surveillance of diseases, health outcomes related to environment or socioeconomics, and in risk assessment for several decades (Elliott and Wartenberg 2004; Ostfeld, Glass, and Keesing 2005; Eisen and Eisen 2010). The sophistication of GIS and spatial analytic tools for spatial epidemiology has shifted greatly in recent years; from simple mapping of existing data to predictive modelling of distributions of hosts, vectors, and their interactions (Openshaw 1990; Rytkönen 2004; Ostfeld, Glass, and Keesing 2005; Goodchild 2010).

Recommendations from the Centre of Diseases Control and Prevention (CDC) and World Health Organization (WHO) have included GIS as a low-cost tool for detection of high-risk areas, predicting potential outbreaks, and for surveillance purposes (Kaiser et al. 2003; WHO 2006).

Wildlife health surveillance is a difficult task in practice that aims to quantify the health of wildlife in nature, however, most wild animals live and die in anonymity without ever being documented through census data, without baseline health information and without the causes of morbidity and mortality being determined (McAloose and Newton 2009). Disease surveillance is more often quantified through post-mortem examinations of wildlife. Figure 1 shows a general conceptual framework guiding this thesis and the analysis conducted. The pyramid in Figure 1 indicates that wildlife in nature can be never be fully captured by a surveillance system or GIS system – they are only abstractions of nature with the best available data. Section 2.2 will provide definitions, an overview of surveillance activities and spatial bias in more detail for surveillance of wildlife. At each step along the pyramid in Figure 1 biases and errors can accumulate further complicating analysis from this data. Surveillance data for wildlife encompasses spatial locations of where a dead animal was found by an actor along with a diagnosis from a trained pathologist (where possible). A portion of these data will be classified on the disease status of the wildlife and reports will be generated to share with wildlife managers, government and other relevant stakeholders. There are many possible derived outputs of spatial information that can be developed from surveillance data. One simple output are maps of observations; more complex outputs would include: risk maps or species distribution models. This thesis will evaluate how surveillance data can be used for spatial outputs such as: species distribution modelling. The spatial nature of infectious disease and the spatial heterogeneities in transmission spread make risk and species distribution models valuable tools to assess the proportion of humans or the number of individuals that are potentially at risk for exposure to vectors and vector-borne pathogens or for practical prevention and control activities (Ostfeld, Glass, and Keesing 2005; Eisen and Eisen 2010; Linard and Tatem 2012). Risk mapping characterizes the spatial variation to a static ecological risk of infection and potential causes for that variation. Risk maps can examine distributions of arthropod vectors, vertebrate reservoirs and actual cases of disease in a host and many other health related measures (Ostfeld, Glass, and Keesing 2005). This process involves gathering remote-sensing data and other spatial datasets that characterize the environment and climate, or future climates. Distribution maps of the vector, reservoir or disease can then be created to examine the influence of the environmental variables and to project the distribution to other un-sampled areas or future scenarios (Ostfeld, Glass, and Keesing 2005).

For this thesis, the application of species distribution models from surveillance data will be explored. To apply species distribution models to wildlife, we need precise locations for where the specie(s) of interest occur in the landscape otherwise known as "presences". There are many methods for tracking and recording where wildlife occurs, however this thesis will focus on wildlife mortality data recorded through surveillance systems and wildlife observations recorded in citizen science data. Wildlife surveillance systems are one such example where spatial data are being incorporated by recording the site where a mortality event occurs. The benefit of using species distribution models is that they can make predictions on where a species is likely to occur in similar environments *without* having recorded presences in that area. This is especially valuable when wildlife data is sparse or limited. The model will examine relationships to where a species is known to occur within a set of environmental and climatic conditions and predict a surface of suitability of similar conditions across the study area.

These models are powerful tools, however there are additive consequences at each step of the modelling process, starting with data collection & recording. The choice of the modelling method relies on the data available and how the data is collected. Elith et al. (2006) reviewed the performance of species distribution models as well as their assumptions. Since surveillance data is often based on observing a dead animal (presences) and not the absence of a dead animal and the sampling design is usually unknown, this limits the methodological choice to presence only methods. For presence only methods, MaxEnt is the most popular and recommended method with 1000 published uses from 2006-2013 (Merow, Smith, and Silander 2013). To complicate matters further, with new systems and disciplines incorporating GIS into their toolkits, there is additional concern with data misuse or misinterpretation, as geospatial data and tools are being utilized by non-spatial disciplines which may have limited expertise on appropriate methodology and interpretation (Unwin 2005; Van Oort and Bregt 2005; Devillers et al. 2007; Pierkot et al. 2011). Both types of wildlife data, mortality and occurrence data, are biased by accessibility. For this paper, accessibility refers to the uneven patterns in data that arise from variability in the ease with which observations are made.

Samples biased by accessibility may affect the prediction accuracy of models if roads directly affect the distribution of species or if the transportation network does not accurately represent the environmental conditions of the region (Loiselle et al. 2008; Boria et al. 2014). As noted earlier, wildlife surveillance data are often not-random and derived from convenience sampling (e.g. road kills, hunter-shot samples) (Nusser et al. 2008) or risk-based sampling methods (Stärk et al. 2006). When the data does exist, the data may be spatially biased by proximity to potential observers of wildlife and may not represent where a species occurs most frequently in nature. Using spatially biased data in distribution models therefore, may be missing important information or giving too much importance to random conditions that may not be indicative of a species actual niche. All these considerations need to be examined before performing any distribution modelling. A method should be chosen based on what is known about the data and how well a specified model can provide additional methods to account for spatial bias. In future chapters, I will examine in detail the spatial bias associated with surveillance data and how "tuning" or not tuning a model can have varying model outputs which should be interpreted carefully. With the increasing threat of EIDs it is important to examine how the types of data available can influence the spatial applications and modelling of this data, as well as the implications of these biased and incomplete data. Since the presence/absence of wild animals is hard to predict but is dependent on environmental and geographical factors (Pfeiffer and Hugh-Jones 2002), examining the habitat requirements of a species may serve as a valuable tool to predict the spatial distribution of a population and identify areas vulnerable to an EID where increased and strategic sampling effort should occur.

Many zoonoses of concern in recent years spill-over from wildlife populations (Mörner et al. 2002a; K. E. Jones et al. 2008). Wildlife are increasingly forced into smaller habitats due to increasing human population and development (Theobald, Miller, and Hobbs 1997; Sala et al. 2000) has likely increased the prevalence of these zoonotic diseases (P Daszak, Cunningham, and Hyatt 2000; P. Daszak, Cunningham, and Hyatt 2001; Cunningham 2005; Hassell et al. 2017). Despite the critical importance of understanding spatial distributions of wildlife; spatial epidemiological analyses of wildlife origin zoonoses are limited by several factors. For wildlife, these investigations are complicated by additional challenges due to sparse, incomplete, and fractured data across several departments and disciplines outside of the health sector (Leighton et al. 1997; Stitt, Mountifield, and Stephen 2007; Stallknecht 2007; Grogan et al. 2014). Section 2.2

will outline the wildlife specific challenges of data and how that can affect the fitness for use of that data. As well, concentrated research, funding and disease management initiatives rarely target wildlife-specific diseases unless they afflict humans, livestock or crops (Voyles et al. 2014; Stephen 2014; Grogan et al. 2014) even though these can have important conservation concerns. For example, in comparison to human diseases, fungal diseases such as chytridiomycosis in amphibians and white-nose syndrome in bats both experienced time-lags in their detection and have resulted in loss of biodiversity and species declines (Skerratt et al. 2007; Blehert et al. 2009; Grogan et al. 2014; Voyles et al. 2014). The data collection for these species or diseases may not occur until well after devastating impacts to the population has occurred. In wildlife studies, often charismatic or flag-ship species are overrepresented in the literature, this is also true in terms of representation in surveillance data.

Wildlife health encompasses complex relationships between space, time, environment, and determinants of health and disease. These interactions can be further understood using wildlife surveillance data in conjunction with GIS and spatial modelling. This thesis will examine how surveillance data can be used in a spatial context for species distribution modelling and investigate spatial data quality elements that can be measured with wildlife surveillance data.

This thesis will focus on answering the following questions:

1) What are the characteristics of traditional wildlife disease surveillance data and to what extent do data quality issues impose limits on the spatial analysis and interpretation of wildlife health data?

2) Is species distribution modelling an appropriate use for wildlife disease surveillance data and how do the characteristics of the data align with assumptions and requirements of SDMs?

2 Background

2.1 Surveillance

Surveillance is defined as: “the ongoing collection, collation, and analysis of information related to animal health and the timely dissemination of information to those who need to know so that action can be taken” (“Access Online: OIE - World Organisation for Animal Health” n.d.). Surveillance methods can vary based on the purpose of the surveillance (e.g. overall health, outbreak detection), targeted species, geographic extent, community involvement, costs and public interest. Wildlife is defined by the World Organisation for Animal Health (OIE) as: “an animal that has a phenotype unaffected by human selection and lives independent of direct human supervision or control” (“Access Online: OIE - World Organisation for Animal Health” n.d.). There are stark contrasts between domestic animals and wildlife. Wildlife cannot be sampled in the same way that domestic animal or human populations are sampled. Domestic animal locations are known and confined, and typically cared for by their farmers and veterinarians (Artois et al. 2009). Human populations and demographics are typically known from census information, they can go to a doctor when they experience symptoms or admit themselves into a hospital. In wildlife we do not have the same amount of knowledge nor does wildlife have access to services that these other two groups have, so choosing the right type of surveillance and area for surveillance are incredibly important to minimize biased sampling. As such, surveillance of domestic animals and wildlife should differ to suit these two specific contexts. According to the OIE, surveillance is aimed at demonstrating the presence, absence or distribution of a disease or infection, or the earliest detection of an exotic or emerging disease. When wildlife are included in a surveillance system, it is often because they can serve as reservoirs for diseases and as indicators of disease risk to humans and domestic animals ("Access Online: OIE - World Organisation for Animal Health" n.d.). Even when wildlife surveillance programs exist, the number of species and sparse geographic areas covered are not at the same scale in which the animals are distributed (Kuiken et al. 2005).

Hoinville et al. (2013) sought to clarify and maintain consistency in the definition of terms commonly used in animal health surveillance to enhance transparency and facilitate the exchange of data. Two common surveillance strategies are passive and active surveillance. Passive surveillance (or opportunistic) relies on the observation of clinically diseased or dead animals and the opportunistic collection of specimens. They are dependent on submissions from observers (the public, hunters, scientists, veterinarians, government, farmers and those in contact with wildlife) to report sick or dead animals and samples can be provided by hunters or collected from road-kill animals (Artois et al. 2009; Ryser-Degiorgis 2013). For a case to be submitted to a diagnostic laboratory, it must persist in the environment, be detected and reported in a timely manner (Stallknecht 2007). Often post-mortem findings are the primary data collected from cases of passive wildlife disease surveillance. The cause of death is generally not available soon enough or at all to assist in early detection since they rely on laboratory analyses that are costly, time consuming or unavailable (Gardner, Hietala, and Boyce 1996; Institute of Medicine (US) Forum on Microbial Threats 2007), or challenges with obtaining the actual specimen and poor specimen quality. Additionally, passive surveillance can be biased towards species and diseases of priority interest (Artois et al. 2009).Whereas active surveillance (or targeted surveillance) collects investigator-initiated (Hoinville et al. 2013) predefined data collection scheme specific to the objectives of the system (Duncan et al. 2008) which requires the capturing or killing of animals for specimen collection with traps, sharpshooters, net guns, drive nets, hazing, chemical immobilization and other methods (Ryser-Degiorgis 2013) or collection of wildlife found dead. A drawback of active surveillance is that it can produce biased results such as when sick animals may be too sick to be trapped or conversely, will be easier to trap through certain methods (Rees et al. 2011) or healthy animals are less accessible for sampling. These methods require a compromise of ideal scenarios that may not be financially feasible or timely to collect like an adequate number of samples for pathogen and antibody detection or provide reliable prevalence estimates and should be representative of the study population (Stallknecht 2007). For some species, alternative sampling schemes may be more cost efficient such as indirect sampling methods like fecal sampling for avian influenza virus (Bevins et al. 2014).

Surveillance of wildlife relies heavily on communities, hunters, conservation officers, rehabilitation centres, zoos, park wardens and others in contact with wildlife to report animals that look sick or are found dead. One of the main issues for wildlife sampling is whether the samples are representative of the population. For example, hunter-killed samples may be restricted in their geography as well as temporally and may not be representative of the age, sex or structure of the population (Stallknecht 2007). Motivation for submitting samples is an important factor that may increase as public awareness and media perceives a disease outbreak in wildlife (Mörner et al. 2002a; Stallknecht 2007). Nielsen et al. (2004) suggests surveillance must address the needs of both the researchers and the hunters (or more generally all submitters) to highlight the mutual beneficial relationship of managing a resource. This can be facilitated by a strong and sustained motivation of each field collaborator by generating training sessions and systemically providing scientific feedback about the fate of each sample to the contributions to the overall picture (Linden et al. 2011). However, despite these potential biases, surveillance data can still be an indicator of events and trends of interest to public health, veterinary and wildlife management purposes and the detection or absence of diseases (Artois et al. 2009).

2.2 Wildlife Data

The quantity of available wildlife distribution and disease data is often limited. Due to the interdisciplinary nature of wildlife and ecology, datasets may be shared and repurposed by other users and often there is limited documentation or metadata associated to aide the user in interpreting the data (Michener et al. 1997). Wildlife occurrence data (location where an animal is found in an environment) can be generated by a variety of sources, most commonly from museum records, scientific studies, conservation groups and citizen science databases (Boakes et al. 2010). These occurrence records, or presences, can then be used in species distribution models to generate a surface of habitat suitability. Similarly, locations of where an animal has been found dead, such as from wildlife surveillance data, can also be used in species distribution modelling, or in risk mapping. Recording cases of morbidity and mortality in each location can provide information on spatial and temporal trends of infection in wildlife (Artois et al. 2009). Data from environmental monitoring schemes and long-term ecological research can provide a baseline for detecting environmental changes which can in turn, indicate the species, ecosystem or habitat health (Pôças et al. 2014). These data provide critical contextual information to evaluate wildlife mortality trends, or other wildlife health issues. Total population estimates are needed to evaluate the health of local populations, estimate the prevalence of a disease, and aid in identifying what populations and locations are at risk or vulnerable (Childs et al. 2007; Stallknecht 2007).

Biased data are a function of either limited resources (e.g., time, accessibility, safety, financial, etc.) or the repurposing of existing data whereby data are sampled non-randomly (Leitão, Moreira, and Osborne 2011). For occurrence data, the spatial biases are typically a result of sampling schemes that are biased towards roads (Kadmon, Farber, and Danin 2004; Reddy and Dávalos 2003), population centers (Barry and Elith 2006; Reddy and Dávalos 2003; Stolar and Nielsen 2015) and parks or protected areas (Leitão, Moreira, and Osborne 2011; Stolar and Nielsen 2015). Additional errors can accumulate through mistakes in transferring field observations to databases, rounding errors, failure to specify the geographic datum used, and geo-referencing of imprecise locality descriptions which can change over time (Wieczorek, Guo, and Hijmans 2004; Graham et al. 2008; Robertson et al. 2016). The quality of data and spatial coverage of occurrence data is also an issue; occurrence data are often incomplete (Araújo and Guisan 2006), the data are presence-only and lack of observations do not necessarily signify true absences (Phillips, Dudík, and Schapire 2004), sample sizes are often small (Barry and Elith 2006) and specific taxonomic groups may be overrepresented or underrepresented (Duncan et al. 2008; Boakes et al. 2010; Varela et al. 2014). Additionally, sick wildlife may not be found before predators and scavengers prey on dying or dead animals making the specimen unavailable to researchers for surveillance purposes (Wobeser 1994; Artois et al. 2009). The quality of wildlife data can affect the outcome of conservation plans where conservation planners typically must use incomplete and available data to make decisions despite the limitations (Grand et al. 2007). Some of these limitations include an over-representation of charismatic or easily detectable species (Grand et al. 2007; Boakes et al. 2010). When data biased by accessibility are used to inform conservation or surveillance plans, we need to be cautious of our interpretation of the spatial and environmental patterns present. Modelled relationships that use biased data (and the spatial distribution maps that follow from them) may be more indicative of the patterns at these accessible areas and represent the sampling effort than the actual niche (Barry and Elith 2006). Ideally, species data would be collected systematically or randomly through surveys where abundance and presence/absence is recorded (Elith et al. 2011). However, this is often expensive, time consuming and skilled personnel, equipment and other resources are not always available to collect the information in inaccessible areas or over larger spatial scales congruent with processes driving pathogen spread and spillover. Knowing the data quality limitations can improve the choice of conservation priorities and guide further efforts to collect additional data (Grand et al. 2007).

Given the variety of information sources for tracking and monitoring wildlife health; data quality evaluation is critical if these data are to be used effectively. For example, repurposing of wildlife health data that is collected for diagnostic testing to monitoring spatial patterns of health and disease may incur biases derived from laboratory locations, taxonomic sampling effort (e.g., mammals vs. amphibians), or spatial accessibility. There is a growing concern with spatial data quality (SDQ) because of the increasing use and amount of data developed by the geospatial community which comes from different sources, different times of acquisition and obtained through different processing methods or techniques (Devillers et al. 2007; Gervais et al. 2009).

The very nature of passive surveillance with many actors submitting observations can lead to uncertainty about the quality of the spatial data being recorded especially if accuracy or uncertainty of spatial information is not quantified within these databases. A representative sample of wildlife can be challenging to match to age, season, disease activity with statistical precision (Pfeiffer and Hugh-Jones 2002). In terms of wildlife health, it is important that software and databases allow easy recording, dissemination, sharing and interpretation (Deem, Karesh, and Weisman 2001) to its users and submitters, especially if these systems are to be integrated with other sources of data and analyzed for large-scale spatial and temporal patterns. Exploratory spatial analysis of wildlife data should minimally include: data description, data summarization, and data quality. These can be simple and straightforward to perform by checking for missing or null values throughout the database, checking for typos and validating spatial locations (e.g. does the label in the database (City, physical descriptor, road intersection, etc.) match the latitude and longitude given?). Summarizing the data can include grouping and counting the data that fits a certain criterion (e.g. # of observations in a season), calculating summary statistics, or classification methods that examine complex relationships within data.

Timeliness is one factor that can be easily determined in surveillance databases. The impact of a disease can be mitigated by rapid detection of first cases (Anonymous 2004) making timeliness a critical characteristic of many surveillance systems. The dissemination of a threat and response based on data interpretation separates surveillance from monitoring activities (Vrbova et al. 2010). Timeliness can be influenced by many factors such as the time to reach a diagnostic facility, the transfer of data, data management, analysis and interpretation which can all depend on the volume and complexity of data and specimens (Sosin 2003). Timeliness can be examined retroactively to determine if there was a time lag in response, which can aide in future efforts to close this gap or investigate methods to improve timeliness.

2.3 CWHC and Wildlife Health Surveillance Sampling in Canada

The lack of a legal mandate and fragmentation of responsibilities for non-zoonotic diseases at local, provincial and federal levels has been cited as one of the main challenges for wildlife surveillance (Vrbova et al. 2010; Childs et al. 2000; Miller, Farnsworth, and Malmberg

2013; Grogan et al. 2014). In Canada, public health surveillance and outbreak detection primarily falls under provincial jurisdiction with some federal agencies tracking zoonotic diseases. For example, research in British Columbia found that one of the principle obstacles to ongoing wildlife disease surveillance was the lack of a legal mandate for agencies to collect pathological or etiological data from wildlife, as well as no centralized authority for contacting and sharing information between agencies and public health officials (Stitt, Mountifield, and Stephen 2007). When data is collected on wildlife, there is no internationally agreed-upon standards or a standardized approach to coding case definitions and recording data (Artois et al. 2009; Childs et al. 2000; Hoinville et al. 2013). The World Organization for Animal Health (OIE) provides guidelines and mandatory reporting for zoonoses that can affect international trade ("Animal Health in the World: OIE - World Organisation for Animal Health" n.d.). For wildlife, the OIE has created the World Animal Health Information System-Wild (WAHIS-WILD) which provides information about non OIE-listed disease in wildlife selected by OIE experts based on their importance for protecting human and animal health based on a voluntary basis for OIE-reporting countries ("Animal Health in the World: OIE - World Organisation for Animal Health" n.d.).

In Canada, the Canadian Wildlife Health Cooperative (CWHC) has established a national network of wildlife disease diagnosticians, researchers and experts in wildlife health and policy advisors ("CWHC - Canadian Wildlife Health Cooperative" n.d.). They maintain a database of mortality events, locations, and associated testing and disease outcome information as well as targeted programs ("CWHC - Canadian Wildlife Health Cooperative" n.d.). The CWHC Wildlife Disease Database is a national repository that eases data sharing and exchange between researchers, wildlife disease managers and other stakeholders and partners (Canadian Wildlife

Health Cooperative n.d.). The database has standardized drop-down lists for specific fields, such as taxonomy and anatomy to keep the data consistent, with some fields that support free text entry and comments from pathologists. Professionals can request access to the database or fill out a data request to the CWHC who will process the needs of the user and export relevant portions of the database according to the research question.

2.4 Spatial Modelling with Wildlife Surveillance Data

The spatial data contained in the CWHC database were used to examine association with respect to auxiliary geographic datasets, such as proximity to various sources (population centers, road networks, provincial and federal parks, or other location types) and to indicate if any locations are geographically biased (i.e., occur more frequently in specific areas).

The details of ecologic parameters associated with occurrences of diseases or of susceptible species may be unclear because of biased reporting or lack of detailed geographic precision in reported wildlife mortalities. However, there are tools to help explain ecological conditions associated with the occurrences of wildlife and in some cases where higher risks of disease are present, even with incomplete data. Species distribution modelling (SDM or ecological niche modelling) encompasses a suite of tools that relate known occurrences of these species to raster GIS layers, which summarize the variation in several environmental dimensions (Peterson 2006).

2.4.1 Maxent

Maxent is a widely used approach for SDM (Phillips, Dudík, and Schapire 2004). The program allows users to input environmental predictors and occurrence locations and then generates a probability distribution of maximum entropy (interpreted as habitat suitability/preference). MaxEnt has minimal assumptions compared to other SDM frameworks such as generalized linear modelling (GLM) and generalized additive models (GAM) and ideal for situations where there are a low number of occurrence locations (e.g. <100), presence only data, limited/incomplete information, and continuous or categorical environmental data (Phillips, Anderson, and Schapire 2006). “Presences” (species occurrences) are locations in the study area where a species occurs. Background points (BP), sometimes referred to as pseudo-absences, are a random sample (n=10,000) of the entire study area *with the assumption that presence locations are also a random sample from the area where the species is present* (Elith et al. 2011). This assumption is often not valid for occurrence data, since the data are non-random, spatially biased and may not encompass the entire study area (Reddy and Dávalos 2003). Values are extracted from the features at each presence and background location. The feature’s response curves can be either 1) highly complex (typical of MaxEnt) and non-linear, which may have high in-sample predictive power but limited biological meaning, or 2) smoother (Merow, Smith, and Silander 2013) which may be easier to interpret and more easily support management decision-making.

Modelling species distributions can be a challenging task as model assumptions may not hold true to the data and parameter decisions can greatly influence model outputs. Recommendations in the literature for best practices, reducing model overfitting, examining data bias and adjusting the model accordingly, may not be followed (See review by: Morales, Fernández, and Baca-González 2017) making it difficult to compare findings.

2.4.2 Spatial Sampling Bias

One weakness of MaxEnt is that it is prone to overfitting. Overfitting in the context of SDMs occurs when a model is estimated that fits the calibration (or test) data too closely. This conflates the signal and noise inherent in the data and reduces model prediction accuracy

(Radosavljevic and Anderson 2014). Over-fit models may have a high AUC (area under the curve) value, but when they are applied to new regions over different time periods the model will lack generalizability and may fail to accurately capture the expected distribution (Robert J. Hijmans, Phillips, and Elith 2017). AUC is interpreted as the probability that a random presence location is ranked higher than a random background location (Merow, Smith, and Silander 2013; Phillips, Anderson, and Schapire 2006). The AUC can be exaggerated when the study area is large (Lobo, Jiménez-Valverde, and Real 2008) or occurrence records exhibit spatial autocorrelation (Veloz 2009). A regularization parameter (β) in Maxent can be adjusted by the user to reduce overfitting and smooth out the response curves of the features. Adjusting the β can penalize the addition of extra parameters (Warren and Seifert 2011) and determines the strength of the penalty (Merow, Smith, and Silander 2013). Sampling bias can also be handled by restricting background points (BP) to the area where samples were collected (restricted area) as opposed to being uniformly distributed over the entire study area (Phillips and Dudík 2008). If the background locations are biased in the same way the occurrence locations are, then the model will already reflect the errors and the predictions will be consistent with the data (Barry and Elith 2006).

Adjusting parameters from their default settings in MaxEnt can reduce the influence of biased data on the resulting model. Accessibility covariates that are retained in models and have a high permutation importance will indicate what covariates are associated with sampling bias for a species. Comparison of models with an adjustment of MaxEnt parameters can indicate which settings are effective at reducing these biases. Examining the resulting models in this context can inform future methodology that can be used to analyse wildlife disease data that are spatially biased.

2.4.3 Methods to Reduce Spatial Bias in Maxent

Subsampling is a common method to reduce spatial clustering by specifying the maximum number of occurrences allowed in a grid-cell and randomly selecting occurrences up to that maximum as a “subsample” (e.g. One occurrence per grid-cell) (Beck et al. 2014; Varela et al. 2014). The downside to subsampling is that it does not correct the lack of data in some areas and there is less data to model (Beck et al. 2014; Varela et al. 2014; Fourcade et al. 2014). The size of the grid cell may alter the distribution of occurrences in environmental space thus the size of the grid cell must be large enough to reduce the bias while also retaining spatial resolution and extent (Fourcade et al. 2014). Boria et al. (2014) found that models that used subsampled data reduced model overfitting and increased AUC scores. Additionally, Fourcade et al. (2014) used subsampling across five separate types of bias scenarios and found that it consistently performed well. In contrast, subsampling in Varela et al. (2014) resulted in occurrence records being discarded that contained relevant climatic information (but were clustered in space) and the remaining records tended to have similar climatic conditions. This suggests that subsampling could have varying impacts on model performance that may be species or sampling specific and could have a negative effect on species living in patchy or spatially heterogeneous environments (Varela et al. 2014).

Background locations in modelling are sometimes referred to as “pseudo-absences”, however these locations do not represent “true absences” thus the term is misleading (Phillips et al. 2009). These locations represent a sample of the set of conditions that are available to a species’ in a region (Phillips et al. 2009). Background locations are typically derived from a random sample of the entire study area; however, this assumption does not account for spatial bias. Therefore, if a model is derived from data biased by accessibility, the model will predict the environmental conditions associated with roads (e.g. Roads in flat, homogeneous terrain) and survey effort rather than the true environmental conditions associated with a species (Phillips et al. 2009). The recommendation in the literature involves restricting background locations so that they reflect the same sample selection bias as the occurrence data so that both data sets contain the same environmental bias (Phillips et al. 2009). The models derived from restricted background locations will ideally highlight the habitats where a species occurs rather than a reflection of the more heavily sampled areas (Phillips et al. 2009). Background samples drawn from too large of an area can lead inflated evaluation measures of model performance like AUC and less informative choice of response variables (VanDerWal et al. 2009). Fourcade et al. (2014) restricted background points based on randomly sampling buffer areas (500 km and 100 km for different species) around occurrence records which failed to improve the model in five separate biased scenarios. Although, it is unclear how they choose the size of these buffers and their ecological importance to the species examined. The results of correcting for sampling bias may not perform equally across various conditions and species depending on the type and intensity of the bias (Fourcade et al. 2014). Therefore, the type of bias in occurrence data should be examined prior to modelling and bias reduction techniques need further case studies to elucidate how these methods may enhance or hinder model performance for various species, in different geographic areas.

In addition to these bias reduction techniques, model evaluation procedures typically involve “training” and “test” data (Fielding and Bell 1997). Occurrence records are randomly split into groups where a large proportion (70-80%) represent “training” data that are used to fit the model; “test” data (20-30%) are retained for cross-validation and calculating AUC. Unfortunately, if the data collected is biased or sampling effort is spatially autocorrelated the "test" data is not truly independent of training data and may generate an overly optimistic assessment of model accuracy (Veloz 2009).

In summary, the difficulty of measuring wildlife disease distribution over large areas is an ongoing issue, coupled with the requirements of SDMs pose a unique methodological challenge. In this thesis, I will explore elements of spatial data quality of wildlife surveillance data through the intended use case of spatial modelling. In conjunction with the impact of accessibility bias and parameter sensitivity on a widely used species distribution modelling tool, MaxEnt. I compare the performance of different models to quantify the accessibility bias inherent in the occurrence data and couple methods for reducing this bias a priori or tuning modelling methods to reduce this bias.

3 Case Study: White Nose Syndrome in Ontario

3.1 Introduction to White Nose Syndrome

White-nose syndrome (WNS) is an emerging wildlife disease that was first discovered in the winter of 2006/2007 in four caves in New York State. Since its discovery, it is estimated to have killed more than 6 million hibernating bats in North America (U.S. Fish & Wildlife Service 2017). The disease is caused by infection with a cold-loving fungus called *Pseudogymnoascus destructans (Pd*) that grows on the skin and flight membranes of bats (Boyles and Willis 2009; Blehert et al. 2009, Gargas et al. 2009). White nose syndrome in susceptible bat species is associated with population level die-offs that could profoundly affect ecological communities (Ehlman, Cox, and Crowley 2013).

During hibernation bat species susceptible to WNS in Ontario, typically roost in crowded caves. This clustering behaviour increases the likelihood of intra- and interspecies transmission of viruses (Calisher et al. 2006) or in the case of WNS, fungus. The affected bats can exhibit severe weight loss, unusual behaviour, more frequent arousals during hibernation, depletion of fat stores, irritated skin, holes in wings and dehydration (Perry 2013). WNS has expanded across the north-eastern United States and has been confirmed at 31 states as well as in five Canadian provinces, New Brunswick, Nova Scotia, Prince Edward Island, Ontario and Quebec (www.whitenosesyndrome.org, Accessed January 2018). In 2015 WNS was also detected in Washington State.

In Canada, no federal government organization holds bat occurrence data in a centralized database or has the rights to distribute the information to those seeking it. The CWHC database can be a useful proxy when formal species survey data are not available, as it contains spatial locations where the dead bats were found. The CWHC data may then be considered a presence-only data set and we can use SDM to approximate the habitat suitability of bats across our study area. Additionally, a subset of these occurrences that are afflicted with WNS can be used to approximate the habitat suitability of WNS afflicted bats. The accessibility bias of surveillance data can be examined by using covariates that may be driving factors of accessibility (e.g., roads, parks, etc.) to determine which covariates (ecological or accessibility) better predict species occurrences, how the suitability differs between these models and model performance. The accessibility of a habitat could introduce sampling biases, since remote habitats or those with difficult terrain are harder to access. Some accessible geographic areas may be over-represented since transportation is cheaper and lower effort is required compared to areas with more difficult terrain (Anthony et al. 2013; Aguirre, Bröjer, and Mörner 1999). In the case of white nose syndrome, this may account for later detection of WNS spreading across North America, since bat hibernacula are often inaccessible. Ultimately, the WNS case study will shed light on the utility of spatially explicit disease modelling based on SDMs with inherently biased wildlife disease surveillance data.

3.2 Methods

An export of all records between 2010-2015 in the CWHC wildlife disease database was performed (n=4176). A description of the submission process and underlying database and reporting mechanisms can be found on the CWHC website (http://www.cwhc-rcsf.ca/docs/CWHC%20Database%20Overview.pdf accessed Sept 2017). Briefly, the CWHC database is a custom-built application dedicated to storing wildlife disease data where the underlying data model was designed to suit the processes at the CWHC. First an incident (mortality event or sighting of an animal) is reported to the CWHC, then a sample of the animal may be submitted to the CWHC, followed by testing, diagnosis/interpretation and finally reporting. A description of the variable names, attributes of the data and type of data (spatial, numerical, categorical) used for this thesis is described in Table 1. All derived spatial data is also defined in Table 1; spatial data was defined into the NAD 1983 geographic coordinate system from csv files and projected to the Canada Conformal Conic projection.

3.2.1 Data Quality Analysis Approach

Before data are used for modelling, characterization of the data or a spatial data quality assessment should be performed. The International Organization for Standardization (ISO) is a worldwide federation of national standards bodies comprised of international organizations, governmental and non-governmental bodies. The ISO has prepared an international standard for geographic data which will be examined in this chapter. This standard provides principles for describing the quality of geographic data and concepts for handling quality information for geographic data (ISO 2013). Many of the ISO principles may be difficult to evaluate for wildlife data since there is a lack of denominator information, external data sets and the databases used for surveillance may not contain uncertainty measures common in GIS databases.

Since the CWHC data is primarily based on passive surveillance where we lack denominator information and validation from other data sources, there are a limited number of data quality elements that can be assessed. For this chapter, I chose to examine 4 data quality elements: 1) completeness, 2) thematic accuracy 3) temporal consistency and 4) Positional bias derived from the ISO (2013). The derived metrics are summarized in Figure 2 and how they are measured.

Bats may be seen as a less charismatic species due to depictions in popular culture and association with disease, compared to other species contained in the CWHC data and thus may have different data characteristics than other species. Additionally, the fungal pathogen responsible for WNS only afflicts bats, which may contribute to the quality or characteristics displayed in the data compared to zoonotic diseases of concern in other species. To investigate this, the data was further subdivided into bat species (n=442) versus all 185-other species of animals in the database hereafter referred to as “all species” (n=3734) which contained 50 columns which are described in Table 1 based on their type and a brief inferred description from the data.

Completeness is defined as the presence/absence of features, their attributes and relationships. Thematic accuracy is the accuracy of the attributes and classification of features and their relationships. Completeness was assessed by determining the number of null values or values coded as “unknown” in each field. The records in the CWHC database fall into three cases: 1) A diagnosis was unable to be performed or a necropsy was not performed. This case can be because of the quality of the sample may not be suitable or able to be determined or simply because the submitter was not interested in collecting that information. 2) There is one incident code per specimen and null values can be calculated. 3) The values contained for a specimen (or row) are empty but contained in another related row (incident codes). In the first case, the specimens that fall into this case were determined by a keyword search using the following terms:

decomposed, no diagnosis, decomposition, autolysed, maggots, poor body, poor, not suitable, not possible, emaciated, not determined, undetermined, unknown, carcass unsuitable, and no necropsy

If any keywords were matched for a given specimen they were omitted. Since including them would inflate the calculated sum of null values. The second case is the only case that can be directly calculated with confidence since these specimens would not fall into cases 1 and 3. Specimens in case 3 cannot be calculated in any standardized way and is also omitted.

Thematic accuracy was evaluated on classification correctness of species names compared to the National Center for Biotechnology Information (NCBI) database and the Integrated Taxonomic Information system (ITIS). Frequency of observations were aggregated to each represented taxon to identify the most common groups observed in the database and to evaluate species representativeness. The R package “taxize” (Chamberlain et al. 2016) was used to check the scientific names from the CWHC against a reference for the most recent names and to automatically identify further taxonomic groups (taxon).

Temporal quality was evaluated by the temporal consistency (correct order of events e.g. Date Received occurs after Date Found). Temporal consistency was evaluated in conjunction with “timeliness” which is the difference in time between when a sample was found and received by the CWHC. Timeliness would indicate if there was a time lag between when the sample was found versus when the sample was received by the CWHC, which could have implications for early warning systems if there is a significant time delay. If the difference in days is negative this would be an example of an incorrect order of events for temporal quality. The seasonal distribution of when samples were found was also evaluated under “temporal consistency”.

To investigate whether timeliness is impacted by other factors within the CWHC database and extracted spatial components (derived from accessibility), Classification and Regression Trees (CART; Breiman 1984) were generated in R with time delay in days as the dependent variable. Trees explain variation of a single response variable by one or more explanatory variables where the response variable is either categorical (classification trees) or numeric (regression trees) and the explanatory variable can be numeric or categorical (De’ath and Fabricius 2000). The goal is to split the data into groups while keeping the tree small and interpreted easily. Observations that satisfy the condition at each junction are assigned to the left branch and the others are assigned to the right branch (Hastie, Tibshirani, and Friedman 2009). The following variables were examined: Captive Class, Storage Type, Submission Source, Number of specimens, distance to airport, distance to major roads, distance to population centres, distance to provincial/federal parks, distance to waterbody, distance to water source, taxonomic order, Bat species or All species, multiple specimens per incident or one specimen per incident, human population from 2016 census, human population density, and season. The Rpart package in R (Therneau, Atkinson, and Ripley 2015) was used to perform the CART analysis. To prevent the tree from having too many terminal nodes and overfitting, a complexity parameter of 0.01 was used and a minimum bucket of n=20 was chosen, which restricts the terminal nodes to containing no less than 20 observations. After 100 folds, the original tree was pruned to reduce the size of the tree (easier interpretation) and selected the split with minimal cross-validated error.

To further examine the spatial aspects and positional bias of surveillance data, SDMs were used to model relationships between species occurrences and environmental and landscape covariates to output a map of habitat suitability.

3.2.2 Species Distribution Modelling of WNS in Ontario

3.2.2.1 Data

The modelling framework (Figure 3) was adapted from (Franklin and Miller 2010, 10) to include the specifics of modelling with disease surveillance data and MaxEnt. First, we must have ecological theory as to the habitat requirements for these bat species and the environmental conditions associated with white nose syndrome (Appendix Table A3, A4). The next level requires that the data is readily available in a spatial format and/or exists; environmental factors for bats in Ontario were difficult to acquire and were not comprehensive (many hibernacula locations on private land are not recorded). Finally, a model can be created linking the ecological theory with available data and species occurrences (or mortality for WNS). During the modelling process there are many calibration, bias reduction techniques, parameters and outputs that can affect the overall habitat suitability map. Bat occurrence records for species affected by WNS in Ontario (and within 20 km buffer) were obtained from the CWHC's wildlife disease database for 2010-2015 (n=443). The three-bat species: *Eptesicus fuscus* (n=122), *Myotis* *lucifugus* (n=295) and *Myotis septentrionalis* (n=26) were treated as one set of occurrences due to their similar habitat requirements, overlapping distributional ranges and presence in the same hibernaculum (Government of Canada 2015) and swarming locations (Randall and Broders 2014). Bat occurrences that tested positive or suspected for WNS were used for separate models and testing (n=163). Additional bat occurrence data was obtained from Global Biodiversity Information Facility (http://www.gbif.org, accessed October 2016) to compare to CWHC models. The study area was restricted to Ontario based on data availability and access to hibernacula locations that was freely available with sensitivity training. To my knowledge, Ontario was one of the only provinces with a centralized repository for bat hibernacula locations from the National Heritage Information Center. Land Information Ontario contained complete, up-to-date resources including detailed metadata, in various spatial formats and resolution, for road networks and other spatial datasets examined below that did not exist for other provinces with the same attention to detail. The National Heritage Information Center (NHIC) provided historic and current (1934-2015) bat hibernacula locations in Ontario. Normalized Difference Vegetation Indices (NDVI) was obtained from MODIS Satellite MOD13A3 at a monthly 1 km resolution; minimum, maximum and average NDVI values were calculated for 2010-2015. NDVI is used in remote sensing to measure "greenness" or vegetation. Several additional landscape and accessibility covariates were obtained from Land Information Ontario (Appendix Table A1). Environmental covariates were downloaded from the WorldClim database (R.J. Hijmans et al. 2005 http://worldclim.org) at a 1 km resolution. All environmental and accessibility covariates (See: Appendix Table A1 for a list of all covariates used) were transformed to euclidean distance (where necessary) and were re-projected to Canada Lambert Conformal Conic, clipped to the Ontario study area in ArcGIS (10.3.1) (Esri 2015). To minimize missing environmental covariates due to a sharp boundary, the Ontario study area was buffered by 20 km. MaxEnt version 3.3.3k was used within the R package dismo (Robert J. Hijmans, Phillips, and Elith 2017).

3.2.2.2 MaxEnt Parameter and Bias Analysis

Model name conventions follow the format where prefixes denote the occurrence data used (c=CWHC, G=GBIF); covariates were denoted by “eco” for ecologically meaningful covariates, “acc” for accessibility covariates, and “dis” for WNS disease covariates. Models with a suffix indicate a modifier on the settings and parameters where rb=restricted background and ss=subsampled occurrences (Appendix Table A2). For example, the C-eco-rb is a model using CWHC occurrence records with ecological covariates with restricted background points that predicts suitable bat habitat. All possible variable relationships (linear, quadratic, product, threshold and hinge) were used and the output format was logistic. The settings were kept consistent across all models, with exception of where RB were used.

3.2.2.3 Regularization

To examine the effects of changing the regularization parameter (β) value on the complexity of the model, the β value was evaluated at 1,2,4,6,8,10,12, and 16. The default settings of MaxEnt use a regularization (β) value of 1, this value was chosen based on performance across arrange of taxonomic groups (Phillips and Dudík 2008). However, this may not be applicable to all species, occurrence localities, study regions and environmental data (Radosavljevic and Anderson 2014). Each model used cross-validation and 4 replicates. Response curves for environmental covariates were examined for each iteration for an expected ecological response. After an optimal regularization was identified, k-fold partitioning (cross-validation) was used on all occurrences with a k=5, the average response curves, jackknife test and AUC was compared for all folds (Appendix B). K-fold validation was chosen as it is the standard validation approach among most MaxEnt studies. The optimal regularization parameter was then used in all following models below.

3.2.2.4 Ecological and Disease Models

The Ecological Models (denoted by "eco" prefix) use environmental covariates that should reflect bat habitat based on previous studies (See: Appendix Table A3**)**. The entire set of WorldClim variables was retained to determine the effects of different model parameters on variable selection. To simplify the model, response curves, percent contribution, permutation importance and jack knife results were used to eliminate environmental covariates from the final model. In the jackknife procedure, covariates that were used in isolation that increased training or testing gain (how well the covariates fit the presence data) were included in future iterations or covariates that reduced training or testing gain when removed from the model were retained. Final models used 30% of the occurrence points for testing and 70% for training. The CWHC WNS Disease Model (CW-dis) follows a similar procedure, however disease covariates and only a subset of bats that tested positive or suspect for WNS were used.

3.2.2.5 Accessibility Models

To examine sampling bias, we examined how well accessibility covariates (distance to roads, airports, population centers, parks) predict bat occurrences. We would expect that an unbiased sample of bat occurrences would be better predicted by ecological covariates if the sample is indicative of the species niche. However, we know that with most wildlife data, there is often a tendency for occurrence data to be biased. Distance from accessibility covariates was calculated from occurrence point locations, with distance to major indicating that most occurrences occurred within 1 km of a road (Appendix Figure A1**)**. Initial models built with accessibility covariates indicated that distance to major roads had the highest permutation importance compared to all other accessibility covariates. In order to account for this bias, two methods were used: 1) Restricted background locations and 2) Subsampling of occurrence locations.

3.2.2.6 Accounting for Accessibility bias

Restricted Background Sampling

The Ecological Restricted Background model (c-eco+rb), utilized restricted background sampling. Since the survey effort for both CWHC surveillance data and GBIF data were unknown, a sampling bias file cannot be computed to account for spatial sampling bias. With the assumption that our data are biased towards accessible areas, we can use the accessibility model as a surrogate sampling bias file, since this surface will utilize the same bias from MaxEnt. Background points (n=10,000) were sampled from the binary accessibility output (derived from the logistic threshold of maximum test sensitivity plus specificity or true skill statistics (TSS) (Jiménez-Valverde and Lobo (2007)). The TSS can minimize the mean value of the error rate for positive observation values and the error rate for negative observations (Duan et al. 2014).

Subsampling

Occurrences were randomly sampled for one occurrence per 10 km grid cell. The resulting model is referred to as Ecological Subsampled Model (c-eco+ss) using CWHC data and G-eco+ss for GBIF. This systematic sampling approach was used by Fourcade et al. (2014) which lowered spatial clustering bias of occurrence points for most virtual species.

Model Comparison Analysis

Evaluation and comparison of models was done using their AUC values, raster subtraction (Appendix C) to identify the magnitude of change across Ontario and by using the common overlap similarity indices nicheoverlap I and D (Warren, Glor, and Turelli 2008) in *dismo* and the structural similarity index metric (Z. Wang et al. 2004; Robertson et al. 2014)**.** Niche overlap 'D' (Schoener 1970) measures the overall match between two niches over the whole study area and determines whether we can infer the niche characteristics of one species from another (Broennimann et al. 2012). The *D* index was selected to compare the maps of habitat suitability since it is a commonly used metric within MaxEnt studies and has a straightforward interpretation (Warren, Glor, and Turelli 2008). In contrast, the niche overlap 'I' is based on Hellinger distance where *I* treats model outputs as probability distributions and compares how identical the niches are (Warren, Glor, and Turelli 2008). Similar to Veloz (2009), for the purposes of this paper, niche overlap indices can be thought of as the level of agreement between the predictions of two habitat suitability maps. The Structural Similarity Index (SSIM) uses a local moving window to compare two maps (or in this case MaxEnt outputs) by generating independent units relating to local similarities in the mean, variance and spatial correlation (Robertson et al. 2014; Z. Wang et al. 2004, 200). SSIM is well-suited for ecological spatial data since it considers local magnitudes and spatial structure (E. L. Jones et al. 2016). The SSIM function was implemented in R and used to compute the composite SSIM index using a 5 x 5 moving window that moves over the entire MaxEnt output raster.

4 Results

4.1 Data Quality

4.1.1 Completeness

There was a total of 4176 records obtained for this thesis. Each row contains a "Specimen Code" that is a unique ID submitted for a given animal/sample. Attached to each Specimen Code is an Incident ID. Each incident may contain one or more specimens, which is denoted by the Incident ID (Many to one relationship). Each contained varying diagnoses and tests but were obtained at the same time or area (e.g. Submitted together). There were 822 records that fell into case 1; 891 in case 3 and were omitted from further completeness analysis; 2463 records were retained from case 2. Completeness of the data is summarized in the spider diagram in Figure 4. For bats and non-bat species interpretation was the least complete (66% and 54%). These levels of completeness would be lower if not for the omission of other cases, since only one row may contain all of the relevant information for one incident code with the remaining specimens (rows) empty.

4.1.2 Thematic Accuracy

The original data was subset and summarized by incident codes (n=3337). The majority of incidents for non-bat species were submitted by the Public (n=1234), Rehabilitation Centres (n=976) and Provincial Government (n=505). For bats the majority of submissions were from the Public (n=68), the Provincial Government(n=64), Universities (n=21) and Rehabilitation Centres (n=62).

The CWHC database contained 185 unique species names which were then classified using taxize (Chamberlain et al. 2016) and tax_name to get the genus, family, order, class phylum and kingdom. 42 species could not be matched in the database from NCBI and were further examined and 28 species were found in the ITIS database. Some species names in the CWHC database contained more than one name in the field in parentheses which wouldn't allow further classification in taxize without proper formatting. The missing names (n=14) were fuzzy-matched using 'classification' where the species name was extracted from the list. To examine the distribution of taxonomic groups summary statistics and plots were created (Figure 5-7). The top 3 species with the highest number of specimens were *Procyon lotor* (n=860), *Myotis lucifugus* (n=288) and *Corvus brachyrhynchos* (n=222). Species were further aggregated to Class (Figure 5), Family (Figure 6), Order (Figure 7). When comparing by incident code the number of incidents per species is similar: *Procyon lotor* (n=800), *Corvus brachyrhynchos* (n=180), *Lanius ludovicianus* (n=146).

4.1.3 Temporal Consistency

For bats, there were 92 records where timeliness could not be calculated, in these cases Date Received and/or Date Found were missing a value. There were 6 negative values, which indicate there may have been a data entry error. For non-bat species there were 802 NA values and 50 negative values. In both cases, there were large time delays that could be attributed to typos with the dates (e.g., 0201-08-30 instead of 2010-08-30) or the sample was a reference from an earlier time; these were corrected when found (n=6) or not included in further calculations. Timeliness used the same subset of data from thematic accuracy based on incidents. The mean and median values for timeliness were the same across both groups (mean= 34 days and median=13 days).

Timeliness of reports was evaluated using box plots (Figure 8-10) and regression trees (Figure 11). The distribution was non-normal for all species and bat species (Appendix D). A

Wilcoxon Rank Sum Test (Mann-Whitney) determined the median difference in samples was a difference of -1.0 (P=0.2499). Storage Type and Submission source values influenced the time delay for both groups. For both groups, specimens that were "Frozen" had the highest mean time delay (43 days for all species and 39 days for bat species. Fresh specimens had the lowest time delay of 1 day median across both groups (9 day mean for all species and 10 day mean for bats). Submission sources for all species with the highest mean time delay were from Conservation Groups (78 days) followed by Parks Canada (43 days) and Rehabilitation centres (44 days). The submission sources for bats that had the highest mean time delay were different; "Other" (68 days), followed by the Public (55 days) and Environment Canada (45 days).

Species with the highest time delay were *Asio otus, Sistrurus catenatus, Falco peregrinus, Aegolius acadicus, Taxidea taxus s* with > 200 days for mean time delay. Orders of Anura, Strigiformes, Squamata and Rodentia had the highest mean time delay along with the highest standard deviations (>100 days). Classes of Amphibia has the highest standard deviation for time delay (164 days), while unclassified values had a SD of 133 days.

The frequency of submissions by year and month were also computed to indicate if there was a time of year when sampling was more frequent or if the listing of certain species of bats to the Species At-Risk Act inhibited collection. Prior to listing both federally and provincially (2012 and 2013), more observations were recorded yearly in 2010 and 2011 and in winter months. After listing, yearly and winter observations fell which may be in line with recommendations for restricted winter sampling of hibernacula or due to a decrease in population for bats ("Ontario's White-Nose Syndrome Response Plan" 2015; Government of Canada 2015). The seasonality of observations (Figure 10) indicated that the majority of bat observations were found in the spring (42%) followed by summer (23%). During the spring, the CWHC requested samples from its submitters since this is when they were more likely to exhibit signs of disease. For all species summer had the highest number of submissions (32%) followed by fall (20%). The NA values indicate where season could not be determined because Date Found and other associated date fields were missing a value.

4.1.4 Regression Trees

The predictors used in the regression tree were chosen based on consistency and completeness across species. The CART analysis (Figure 11) identified 5 variables as the most important for timeliness: Storage Type (32%), Submission Source (19%), Order (20%), and Population from 2011(22%) and Captive Class (7%). Resulting in 4 splits and 5 terminal nodes. The first split divided observations by Storage Type. Storage Types from Fresh, Fixed and unknown samples ended with one terminal nodes with an average of 11 days. Frozen samples were further split based on submission source, population and order. Conservation groups and NGOs were a terminal node with an average of 103 days. The remaining submission sources were then split on population >= 5555; with one terminal node with an average of 34 days when population is <5555; the final split was based on order, for the grouping including bats (Chiroptera) the average was lower than the other terminal split with 54 days while orders of: Falconiformes, Galliformes, Squamata, Strigiformes had a time delay of 159 days.

4.2 White Nose Syndrome Case Study: Species Distribution Modelling

The default regularization setting produced response curves with irregular shapes that did not represent an ecological response for bats, whereas β values greater than 6 generated more realistic estimates of ecological response (Figure 12). The cross-validated model's average AUC was 0.948. The habitat suitability maps generated with a β of 1 were less generalizable, whereas a β of 6 yielded a smoother surface with higher intensities near Ottawa and in the Greater Toronto Area (GTA) as well as North-Western Ontario (Figure 13). The covariates in the final models (Table 2) indicate that restricting the background points can influence the importance of covariates and to a lesser extent, subsampling occurrences. The AUC scores for all models (Table 3) were >0.75. AUC decreased most when RB was used for both C-eco and G-eco. The habitat suitability maps (Figure 14) range from 0-1 with values closer to 1 being predicted more suitable. The response plots (Figure 15) are similar across most models, with RB smoothing and generalizing these relationships. For models using CWHC occurrences, the permutation importance of distance to hibernacula was consistently >0.62. Restricting the background points decreased permutation importance of hibernacula and increased maximum NDVI to 24%. The models using GBIF occurrences had lower values of permutation importance for hibernacula (<35%). The G-eco Model, bio6 had a dramatic increase to 59.2% and distance to hibernacula was much lower than the C-eco model (13.5%). G-eco-rb had the highest permutation importance for hibernacula among GBIF models (35%), and increased importance of maximum NDVI (44%).The G-eco+ss model increased bio6 to 72% and decreased hibernacula to the lowest permutation importance.

Habitat suitability for bats across all models was highest in southern Ontario, specifically around the GTA, Niagara and Bruce County (Appendix Figure A2). Northern Ontario had localized suitability around hibernacula. The G-eco habitat suitability map is similar to the C-eco Model but follows a more generalized pattern as opposed to the localized areas surrounding hibernacula locations in the C-eco model. A binary representation of the C-acc was calculated at the maximum test sensitivity plus specificity logistic threshold value (0.186) to sample background occurrences from for the c-eco+rb model. The threshold value used for G-eco+rb was 0.202 from the G-acc model.

The C-eco+rb model removed noise from the C-eco Model and areas around the GTA and Central Ontario are less suitable, however the C-eco+ss Model increased habitat suitability in Central Ontario and the GTA and contained less suitable areas in Northern Ontario. It removed the concentrated areas around hibernacula locations that were present in the C-eco Model. Notably, new suitable areas in Northern Ontario that are not present in any other model, emerged through the G-eco+rb Model. This discovery is likely related to the use of the restricted background points used in G-eco+rb.

The WNS models (CW-eco and CW-dis) had the highest AUC (0.978, 0.976 respectively). CW-eco utilized the same ecological covariates as all other eco models while CW-dis used covariates associated with *Pd* growth from other studies (Appendix Table A4). CW-eco had the highest permutation importance of 97.5% for distance to hibernacula. with the rest of the covariates being <=1% (Table 3). The areas on the suitability map (Figure 14) are concentrated around hibernacula locations and have less of a gradient-like effect than the C-eco output. For the CW-dis hibernacula remained the most important with 95.2% while all other covariates were minimal in their contribution. Comparison of the models indicate that G-eco is the least similar to both CW models. This is difficult to compare, since CW models are derived from the same data source, so it is expected that G-eco would be the least similar. One interpretation could be that the models that used GBIF data are more indicative of an environment where WNS is less prevalent or the difference in where citizen science observations take place compared to surveillance.

Distance to major roads was the most important for the accessibility models (Table 3), however G-acc had higher values for distance to airports and population centers. Similarity between the eco and acc models (Table 4) were highest for models where no bias reduction was attempted (C-eco, G-eco), however for C-eco-ss had the highest values *I* and *D*. C-eco+rb and G-eco+rb had the least similarity to acc models. The visual comparison of SSIM (Figure 16) indicated that that most models have low suitability in Northern Ontario. Areas closer to zero (orange) indicate that these models exhibited different patterns. These areas of divergence are common along smaller lakes and areas where there are no roads intersecting them. For the C-eco and G-eco models SSIM shows a large discrepancy on the coast of Lake Huron with ~ -0.4-0.5 change, where GBIF predicted more suitable habitat.

5 Discussion

5.1 Data Quality

The data quality analysis performed here included completeness, thematic accuracy and temporal consistency. The fields that contained the most null values likely have little impact on those using the CWHC data, depending on the purpose of their surveillance activities or special projects from a researcher. If many different tests are being ordered on a sample, it may not be relevant for the submitter to ask for a necropsy and some fields may not be filled out. One issue with the data is the use of two unique ids where one indicates a specimen and another a group of specimens (incident). While grouping by incident can serve a specific purpose for organizing groups of submissions that may be found in a similar or nearby environment or collected for a specific purpose, it is less useful in the context of this thesis where we are already extracting spatial information per specimen. A complication with this data was lack of consistency across fields where data is provided on a per incident basis but not per specimen. For example, a group of 6 specimens may only contain a diagnosis and test results in one row of the data (for one specimen) but this diagnosis is explaining all 6 specimens. This presented challenges when determining WNS status for bats. The WNS status field may have been null if WNS is not suspected or if data is being collected for a different purpose (e.g. If rabies is the main concern of the submitter). Keyword searches on multiple columns and manual identification were used to classify WNS status as a result of the information being inconsistently distributed. Further, during the study period the identification of the fungal pathogen responsible for WNS changed from *Geomyces destructans* to *Pseudogymnoascus destructans* which should be retroactively updated. Unknown values for sex or age may be due to difficulties with desiccated bats where identifying sex or age is impossible. Approximately half of the interpretations were missing an explanation for both bats and all species. In some cases, additional information was provided in other fields, however interpretation may be a valuable free-text field prior to discovery of a disease. For example, commonly database fields for already discovered diseases are added when they are of concern, however prior to discovery there may be characteristics or conditions of the specimen that may be recorded in free-text fields such as interpretation, and history; that when examined after an outbreak may provide further clues to classify the existence of a disease. If these characteristics were identified and recorded under a free-text field such as interpretation, they may have served as an early warning to a spreading disease in bats before catastrophic population declines. Further keyword analysis could be performed on interpretation fields to understand more contextual information about the diagnosis that isn't contained in other fields (Appendix Table E1). For bats there was some missing information for sex and age. For evaluating the population affected or recovering from WNS the sex of the bat can be important for estimating reproduction rates and population size as most bat species affected by WNS have one offspring per year (Barbour and Davis 1969; Thomas, Fenton, and Barclay 1979). As well, age of the bat and timing of pup date of birth can contribute to survival rates of bats (Frick, Reynolds, and Kunz 2010). While the quality of the samples is an ultimate limitation on what can be determined and recorded about a specimen, some of these characteristics may have important implications for determining the surviving bat population and where intervention may be necessary.

Wildlife data can also be biased with respect to seasonality. The seasonality of submissions differed for bats and all other species which can be attributed to life history, seasonality of disease prevalence and host vectors as well as regulations or restrictions for sampling specific species and when surveillance effort was highest. For bats, most samples occur in early spring which may be due to limiting disturbance or restricted sampling of hibernacula in winter months (“Ontario’s White-Nose Syndrome Response Plan” 2015) or spring emergence from hibernacula to maternity roosts (Fenton 1969; Foley et al. 2011). For all other species summer had the most frequent observations which may correspond to seasonal breeding and feeding locations (Mörner et al. 2002b), the seasonal patterns of disease emergence (Bradley and Altizer 2007; van Dijk et al. 2014; Rees 2011) or species were easier to detect or more accessible in summer months; and when people are more likely to be outside (camping, hiking, etc.). The temporal aspects of the database are not able to be fully examined without knowing precise dates and activities the CWHC is calling for submissions or actively targeting a species/disease/area.

The species represented in the database were skewed towards animals with specific disease concern and larger mammals. The largest number of observations was for the Procyonidae family (primarily racoons), which may speak to rabies control efforts and their occurrence in close proximity to humans; as well as targeted research projects for racoons over the study period. The number of observations for a particular species can indicate the importance they may hold for emerging or re-emerging zoonotic diseases, such as Rabies, West Nile Virus or Avian Influenza or the likelihood of being opportunistically discovered dead as part of community hunts or a planned large cull (Schurer et al. 2013). For bat species specifically, there was a large number of observations in the CWHC database which may be representative of testing for rabies as well as mandatory reporting of collisions/mortality associated with wind turbines (Kuvlesky et al. 2007; Kunz et al. 2007; Bird Studies Canada et al. 2017) as well as increasing interest/concern about WNS.

There were slight differences in the timeliness observed between bats and all other species in Ontario specifically when looking at mean with 32 and 31 days respectively. The median values (12 days and 12 days, respectively) were the same**.** Submission source was another field that contributed to timeliness which may speak to the resources and funding available to submission sources and the cost effectiveness of sending samples to a diagnostic lab or mandated diagnosis through provincial and federal government. Special projects may also play a role since there are dedicated and prioritized resources for these projects, however determining special projects from the data was not possible. For all species, conservation groups and rehabilitation centres had the highest median time delay while local, provincial and federal government sources had the shortest delay. For bats, "Other" submission sources had the highest median time delay followed by Environment Canada and rehabilitation centres. Bat submissions from conservation groups, universities, Provincial government and had the shortest time delay. Additionally, the storage type of the samples influenced timeliness as well; where frozen or unknown storage types had a higher median and mean time delay for bats. Fixed and missing values for storage had the highest time delay for all species which be an indicator of overall data quality or the quality of the specimen submitted. Frozen samples may have a longer delay when sent to the lab because of the distance to submission, or cost effectiveness in waiting to send multiple frozen carcasses to the CWHC.

The CART analysis provided further insight into potential predictors for timeliness contained in the database and derived spatial components. The original tree was pruned to reduce complexity and readability, which eliminated some of the spatial components such as: population, distance to population centres and parks. Interestingly, CART did not find importance based on the binary classification of records of “all species” or “bat species” but did group observations by taxonomic order, of which bats had a higher time delay than the other taxonomic order grouping. The original hypothesis was that a large number of specimens associated with increased shipping costs and effort, which would contribute to a delay in timeliness. The number of specimens submitted by incident was surprisingly, not an important determining factor of timeliness from this CART analysis. Storage types that were fresh were received by the CWHC the fastest, however frozen or fixed samples were further divided by submission source and order. Rehabilitation centres had the highest time delay in their node. The CART analysis was mainly exploratory in this case to examine which factors may influence timeliness in the database. One drawback of CART is that it partitions criteria into 2 groups, for data with many categorical variables (like CWHC database) this can be limiting since it aggregates many values in a group together. Fields like “Order” contained 34 possible outcomes and then were forced to be split into two groups. For examining different classifications with many categories, CART is biased in variable selection towards variable that have more splits (Loh and Shih 1997). Overall, CART is an effective way to visualize complex relationships within data, however its utility when certain variables have large groups in one category (e.g. Frozen in Storage), or there are many categories in a variable (e.g. Order) can be limiting.

Further research into the expected timeliness from when a sample reaches a diagnostic facility versus timeliness to detect and communicate a disease outbreak is needed to evaluate "effective timeliness". Experts could also determine the variables to include in CART that may be more meaningful or to address a specific research question, or to categorically define the time variable (eg. <10 days = "Fast", >10 days = "Slow", >30 days =" Unacceptable") to be used in a classification tree, which may improve generalization.

5.2 Ecological and Disease Models

The data quality analysis provided additional insight into the habitat suitability maps and the importance of the submission source. The application of wildlife data was examined by using MaxEnt to model habitat suitability of bats and suitability of WNS afflicted bats. Overall, hibernacula locations were the most important for the Ecological Model. Southern Ontario, specifically around the GTA, Niagara and Bruce county, contain habitat suitable for bats. In the absence of cave or mine locations in Southern Ontario, buildings are common roost sites for more than half of Ontario's bat species (Soper and Fenton 2007). Additionally, dead bats may have been reported at wind farms in Southern Ontario (Government of Canada 2012; Kunz et al. 2007; Bird Studies Canada et al. 2017); which is further supported by the submission source being primary the public, provincial government, rehabilitation centres and universities in Southern Ontario. Further, in highly populated areas, observations were submitted by the public, which maybe simply because large numbers of people are able to observe bats them and interaction due to proximity to humans (bats flying into windows, pest removal and capture by domestic cats). In Northern Ontario, the suitable habitat is more localized than the spread in Southern Ontario and the majority of submissions were by the provincial government. These locations from CWHC surveillance data were often at cave or mine locations where previous bat studies had occurred (e.g. Fenton 1969 and Dubois and Monson 2007). It is worth noting that there may be sampling bias for CWHC and GBIF data associated with these locations, since they are "known" to researchers and cavers, but other undocumented hibernacula in Ontario may exist.

The GBIF models had a greater gradient of suitability and were less concentrated over hibernacula locations. This could be a function of GBIF data being reported from multiple observers across the province, which may not be concentrated around caves (as can be seen by the lower importance of hibernacula in GBIF models). The GBIF models were least similar to the CW models, which could indicate that observed bat occurrences have a different distribution and associated environmental predictors than dead bats (Appendix C). The link between environmental determinants of health for bats is hard to examine with the data issues examined in this paper and requires future research. It's possible that abandoned mines in the Ecological models could contribute to bias the MaxEnt outputs as well, since mine locations may have existing infrastructure such as roads connected to them. An alternate interpretation is that accessibility is causally related to WNS; as a mechanism for WNS spread since these areas may be frequently explored by cavers and researchers alike, where contamination of fungal spores on equipment may spread to uninfected hibernacula locations and disturbance of bats during hibernation may lead to mortality of bats (Foley et al. 2011; Blehert 2012; Government of Canada 2015). The new areas revealed in Northern Ontario through the C-eco+rb model may highlight inaccessible areas suitable for *Pd*. The C-dis model lacked a significant correlation with any of the environmental variables, which may be a symptom of the data, but WNS is also associated with conditions inside bat hibernacula, which would not have been reflected in the WorldClim data. The choice of environmental covariates for MaxEnt can also be subjective to the modeller's interpretation. The environment surrounding hibernating bats was also examined by Puechmaille et al. (2011) where the presence of viable spores of *P. destructans* was found on the surfaces of hibernaculum. This has major implications for understanding the disease transmission and disease modelling, as cave walls could serve as a passive vector or reservoir for spores (Puechmaille et al. 2011; Hallam and Mccracken 2011). Therefore, if the spores remain viable over many seasons, healthy bats could encounter these spores and contract WNS during their hibernation period. Lorch et al. (2013) state that *P. destructans* is capable of surviving in bat hibernacula when bats are either absent or at low densities, which further supports that caves and mines can serve as a reservoir when bats return for hibernation in autumn. Sediment samples from the same study also contained *P. destructans* after the caves had been sealed off without a presence of bats for two years, which indicates that the fungus is very persistent in the environment. In laboratory conditions, Hoyt et al. (2014) showed a long-term persistence of *P. destructans* in the absence of bats, which further suggests that environments contaminated with *P. destructans* may serve as long-term environmental reservoirs. Hoyt et al. (2014) also showed that the fungus can survive outside hibernacula on equipment or clothing if they are stored in cool, dry conditions. While the temperature outside hibernacula will control the proliferation of WNS, especially in cold regions (Flory et al. 2012) it would be beneficial to study microclimates within hibernacula as well. Microclimates used within hibernacula may vary between species and within its geographic range, which could influence the proliferation or inhibit *Pd* growth (Verant et al. 2012).

Using both CWHC and GBIF data sources in this paper illustrate how methods to reduce spatial bias and evaluation procedures can have varying effects on estimated habitat suitability, even when the same species and study area are used. This furthers the need for more rigorous evaluation procedures and methods to reduce spatial bias to be examined, as both data sources did not provide adequate models using the default settings of MaxEnt. The regularization parameter in MaxEnt can affect the shape of the response curves and permutation importance of model covariates. For bats, the default regularization value produced irregular response patterns until a regularization of 6 was reached.

Wildlife mortality data may be used in SDM and may be a viable additional data set to use in these models where occurrence data are sparse or do not exist. While mortality data eliminate the issue of repeated sampling of the same individual from wildlife surveys - there remains the spatial bias common in both wildlife data. It is well known that species occurrence data tend to be biased towards accessible areas (Reddy and Dávalos 2003; Beck et al. 2014; Syfert, Smith, and Coomes 2013; Kramer-Schadt et al. 2013; Dennis and Thomas 2000). Proximity to major roads was the most important bias in this study, in contrast to other studies where occurrence data was biased towards roads, rivers, and population centers (Reddy and Dávalos 2003). This may be due to the nature of surveillance data, but the same can be said of the GBIF citizen science data.

5.3 Model Parameters and Evaluation

A common metric to compare fit, AUC was >0.9 for C-eco and C-acc models, indicating a “good model” fit. Similarly, for GBIF models, the G-eco model had a lower AUC than G-acc model indicating that accessibility covariates *better predict occurrence locations than ecological* *covariates*. The impact of spatial bias, while evident from visualization of spatial distributions is, critically, not detected via AUC. AUC as a model assessment criterion for MaxEnt has been criticized (Jiménez-Valverde and Lobo 2007; Veloz 2009; Merow, Smith, and Silander 2013) due to model overfitting and uncertainty with the value of AUC as a metric. This case study furthers the assertion AUC as an indicator of model performance is flawed, since both models had similar AUC's with completely different inputs and vastly different suitability maps. A high AUC should be interpreted cautiously when data are biased and being used for decision-making as the metric may instill false-confidence in a model that does not accurately reflect habitat suitability.

The global indices used to compare models tended to agree, with *I* having the highest values, followed by *D* and then SSIM. In contrast to Warren (2008), there was a difference between the *I* and *D* metrics, where the values for *I* were seemingly inflated and rarely fluctuated despite differences observed in other evaluations. This may be because of species examined but the use of *I* as a metric was not as useful to discriminate between different models. Although the AUC values for the C-eco and C-acc models are similar, their distance metrics vary for each evaluation measure. The assumption being that since the AUC is high for both models that they would have similar prediction surfaces for the output. SSIM had the lowest value of 0.57 whereas *I* and *D* had higher values of 0.87 and 0.64. Visual comparisons indicate that spatially, these two suitability maps are more different than *I* and *D* suggest. SSIM showed differences between the C-models, particularly around lakes. RB led to a decrease in importance for hibernacula and an increase in climatic and environmental variables, such as bio 6 and NDVI. NDVI would then have a stronger signature for areas on lakes where NDVI is lowest.

The RB models (C-eco+rb-0.47 and G-ec+rb-0.12) both had a lower SSIM than the subsampled models (c-eco+ss-0.574, G-eco+ss -0.462) when compared to the respective acc models. This may indicate that RB is better at reducing the spatial bias and reducing model overfitting than subsampling. Additionally, the G-eco+ss model was more similar to the GBIF Accessibility model, further illustrating subsampling does not reduce bias in this case. In contrast to Fourcade et al. (2014), subsampling did not seem to reduce model overfitting or remove the accessibility bias from the model output. When accounting for spatial bias in species occurrence data, we must be careful that removing records does not alter or remove potentially important information. Filtering of records can weaken prediction - ideally, prior knowledge about the sampling scheme and associated bias is already known and can be used to construct a bias file with 'true absences'. However, this paper demonstrates that RB based on a previous accessibility model can be an effective method when sampling effort and true absences are unknown. This method requires further studies to determine the effectiveness; however, for this paper, it predicted new areas, reduced noise, and was less similar to Accessibility Models than the standard c-eco Model. However, since the C-eco+rb model utilized the binary suitability from the Accessibility model we may expect these models to be the least similar.

6 Conclusions

Utilizing spatial analysis in wildlife health research could aid in decision making and locating at risk areas for diseases or adverse health to wildlife. Focusing sampling and surveillance programs in these areas could prove to be a cost-effective solution for otherwise expensive sampling efforts. The conceptual model guiding these thesis questions (Figure 1) summarizes the levels of complexity and bias that contribute to an understanding the relationship between wildlife & their environment. Going forward in using spatial outputs from surveillance data, the tools used to model species distributions and spatial epidemiology need to place emphasis on these three components. The first component relies on actors (Figure 1) and data users to improving the quality of information available. This encompasses understanding sampling biases, limitations and gaps in current surveillance data (spatially, temporally and taxonomically) and establishing common data standards. The quality of information available can be examined and reported in various output formats such as: by exploratory analysis, graphing relationships (e.g. time of year vs count of observations) and mapping of spatial data or further quantified by using common spatial techniques to evaluate clustering of biased spatial data (e.g. Getis-Ord GI*, Moran's I, and Kernel Density, etc.). Common data standards for wildlife & surveillance is would benefit the international community and any analysis that spans over wide geographic regions. The second component is quantity of data where data such as occurrence records are limited and/or need to be aggregated data from multiple sources and agencies, or the wildlife data does not exist. Lack of data can be collected or supplemented by other data sources such as citizen science, or researchers willing to share their data, however where no data exists, there must be an incentive or need to collect the data, which may not be realized in a timely manner, or by those who would fund such endeavours. The third component is utilizing and understanding the underlying assumptions of statistical approaches and choosing the right approach that can handle large and patchy data. The results of this thesis indicate that although MaxEnt is widely praised and used for SDMs, the adjustments in the default settings of MaxEnt can affect the areas predicted as suitable for a species, as well as the importance of input variables. Approaches to SDMs that feature default settings, sparse or non-reproducible methodology sections are common in the literature although, fortunately, there is also increasing awareness and criticism of these approaches and the associated model evaluation procedures (e.g. Lobo, Jiménez-Valverde, and Real 2008; Morales, Fernández, and Baca-González 2017). These three components need to be examined in whole to further increase confidence in model outputs, data replication and their derived conservation plans and risk disease modelling. These changes may be slow moving and difficult to implement for wildlife health, especially in the near future since the availability of samples and sample distribution are limiting factors; however, utilizing new technologies and a transdisciplinary perspective could increase confidence in wildlife health research and inspire action. An EcoHealth approach to wildlife health and epidemiology could benefit from many different domain experts tackling a complex problem holistically (Stephen et al. 2004; Aguirre 2009). Since wildlife and pathogens can cross borders, ecosystems and species, knowledge spread across many disciplines & researchers must be part of the solution.

The CWHC surveillance database contains a wealth of valuable information about wildlife throughout Canada. The use of this data will vary across actors, users and of course, by discipline. While the database contains spatial information, it is not a spatial database and cannot be easily validated by common spatial database procedures (e.g. is a given latitude or longitude within Canada?). When examining data quality characteristics, the data is commonly looked at by "incident", however in most spatial contexts we are more concerned with the individual specimen's relationship to space and the environment. Mechanisms to translate the format of the CWHC surveillance database to a spatial database would be beneficial when a spatial component is being analysed by a user. This translation would again, require collaboration by the CWHC's data producers, users as well as GIS and IT experts with knowledge of the CWHC's practices. The benefit of a spatial surveillance database would be two-fold in that some spatial validation would be automatically conducted and it would make spatial data formats more available to users who may not be familiar with common spatial procedures.

6.1 Recommendations

6.1.1 Metadata

Metadata is "data about the data" (Orr 1998) which contains information for each field of the database explaining the purpose, type of attribute, values that are contained and what those values mean as well as information about data quality. Ideally metadata comprises all information that is necessary for sufficient long-term secondary use or reuse by original investigators and by other scientists who may not have been directly involved in the data collection or recording process (Michener et al. 1997). Explicit metadata that contains the level of uncertainty in location based measurements should be included along with data products or included within them. As Devillers et al. (2007) notes, positional accuracy contained within a separate metadata document may be too general when one is interested in a specific data point (e.g. A specific building); instead metadata encoded at the feature level is more useful and can be easily examined in a GIS. I recommend that all data collection agencies such as the CWHC should provide easily accessible metadata adhering to a standard (e.g. Dublin Core, ISO-NAP) to ensure that users are understanding the data and the limitations of that data. As well, geographical data collected in wildlife disease surveillance systems should also provide tools to record the level of spatial precision reported for instance, if a GPS was used for precise coordinates and the error associated, or whether a road intersection was given, or a general place marker was estimated; as this information can be incorporated into spatial models that incorporate uncertainty. There were few errors with the ordering of date related fields, however date validation upon user-entry using form-validation tools would also be beneficial (e.g. Date Found has an earlier date than Date Submitted and is validated before submission).

6.1.2 Fitness for Use

"Fitness for Use" is a concept that illustrates the importance of taking in data consumer's viewpoint of data quality (Veregin 1999), since they are the main consumers of said data. The CWHC could follow-up after completing a data request with the data consumers to evaluate perceived data quality, such as how well the CWHC data fit the purposes of the users' study or analysis, what could be improved upon, and any data quality issues found. The results from these surveys could then be used to improve the quality of data, include user training and improvements to the database and metadata documents (R. Y. Wang and Strong 1996).

6.2 Limitations

National, centralized repositories such as the CWHC's wildlife surveillance database are important data sources for wildlife health and provide educational and collaboration opportunities with many researchers and government scientists. The lack of wildlife data is a limiting constraint for wildlife health and zoonotic disease focused studies. Therefore, an obligation exists to evaluate these national surveillance schemes to further our own understanding of these processes and establish standards for collecting, maintaining, and sharing data concerning wildlife at the international level. The analysis here indicates that certain species may be recorded disproportionately depending on public health concern and ease of sampling. Spatial analysis can aid in indicating which spatial factors are underlying the surveillance data.

Using surveillance data in place of occurrence data for species distribution models is challenging since the data is coming from many different actors (Figure 1) who may record spatial precision or accuracy differently or this information may not be available. The information contained within surveillance data may vary by the surveillance objective, specific project, species, disease of interest or number of specimens at a given mortality event. Contextual information regarding when press releases were issued or when targeted surveillance started could also be incorporated to understand an increase in frequency of specimens versus a mortality event that was recorded without much effort. MaxEnt is a valuable tool being used in species distribution modelling and in disease modelling that can handle small sample sizes and limited data, as shown in thesis. However, more work needs to be done in understanding the underlying biases in occurrence data and providing detailed methods used (e.g. ecological basis for model covariates, regularization, cross-validation used, background locations and model assumptions) in published research (Morales, Fernández, and Baca-González 2017). These methods should include an investigation into potential spatial and accessibility biases in occurrence data, an ecological basis for model covariates, listing of the resolution of covariates and any transformations performed on the data, regularization value used (and why), type of cross-validation performed, number of background locations and how they were chosen and knowledge of underlying model assumptions.

There were numerous challenges with acquiring appropriate data for modelling bat species in Ontario. As with Maher et al. (2012), there were difficulties in obtaining Canadian cave data as, to my knowledge, a standardized database does not exist. Several datasets of interest were freely available in the United States but unavailable in Canada, or simply did not exist. There were concerns for privacy of cave locations that were on private property, as well as concerns that if these locations were known to researchers, cavers would find them closed to protect the bats. With the listing of three bat species to Species at Risk Act, only data at certain resolutions were available from the Government of Canada. Aggregation is an unfortunate realization for many spatial and ecological studies such that it can hinder the outcome of the study if these limitations are not realized.

Comparison of this paper to Flory et al. (2012) is difficult since they used WNS-infected hibernacula locations as “occurrence records” rather than actual species occurrence to “map potentially suitable habitat for mortality associated with infection by an organism”. Similarly, this thesis uses mortality data of WNS positive bats to predict the latter, but also uses dead bat locations to map habitat suitability. The results presented in this thesis do not account for differences in disease among species however the species examined are all species afflicted by WNS in Canada. While Flory et al. (2012) used similar climatic data (WorldClim) the variables used in their models and the shape of their response curves were different from the ones used here. Flory et al. (2012) also note the importance of environmental conditions inside of hibernacula (or microclimates) influencing WNS transmission, but due to the urgency of WNS and lack of existing information, they could not incorporate these factors into their models. Alves, Terribile, and Brito (2014) also used MaxEnt and used WorldClim climatic data, however the variables used in their final models were different than observed here. This is also difficult to compare since their models for North America were projected from previous models of WNS presence from European caves, where pathogen and host are assumed to have coevolved (Warnecke et al. 2012; Wibbelt et al. 2010; Alves, Terribile, and Brito 2014).

6.3 Research Contributions

This thesis contributes to the existing body of literature that attempts to find a relationship between environmental predictors and species distributions. In situations where data is limited for wildlife, utilizing surveillance data as “occurrence records” is valuable. This data source may serve as an additional or complimentary dataset to more traditional wildlife surveys or in cases (like this thesis) where no other data is available for a species. More specifically, this thesis outlines the specific biases associated with bat surveillance data in Ontario such as accessibility, and how model parameterization can affect resultant models. Many studies using similar methodologies fail to produce reproducible methods and rely on the “default” settings in software tools. The results from this thesis indicate that the default settings for bats lead to unpredictable response curves and a bias towards accessible areas in model outputs. The methodologies presented to reduce accessibility bias here are novel and lead to habitat suitability that was present in new areas of Ontario.

The knowledge gained from this research will help wildlife managers and the CWHC make evidence-based decisions to allocate resources for the surveillance and monitoring of wildlife, specifically concerning White Nose Syndrome in bats. It is crucial to understand the distribution patterns of bats as well as the distribution that is suitable for development of *Pd.* The tools presented here will contribute to the development of tools that will promote conservation and protection of sensitive ecological areas in Canada. As well as methods to assess the reliability of wildlife health surveillance systems in Canada regarding sampling effort and spatial biases. Future actions and decisions to limit disease spread and protect species should not be made without further validation of models, which may require collaboration with existing bat experts and further field surveys to collect additional species occurrences. While the health of wildlife or, more accurately, an ecosystem is important to each individual's health (including humans), there needs to be a shift in research and funding to not only fund wildlife health research but to recognize the value in adapting existing surveillance and monitoring wildlife strategies to ensure stable ecosystems, prevent or predict epidemics and, of course, protect the integrity of livestock and human health. Another outcome of this work will be to invite commentary on the existing policies in place concerning conservation, species at risk, the environment and disease prevention in Canada. It is my hope that the ideas presented here provide additional motivation for collaboration among researchers and sharing of data so that we can make informed decisions with the best available data to protect our shared environments with wildlife.

Since WNS is a new EID there remains a number of uncertainties and room for future research. There is sparse literature on environmental variables and modelling on predicting the spread of the disease, and denominator bat data. The study of bats in Canada is primarily presence-based and thus would rely on presence-only modelling methods like MaxEnt. The covariates used described conditions outside hibernacula, similar to Flory et al. 2012; Alves, Terribile, and Brito (2014) and indicate that hibernacula are the largest predictor for WNS, which given that WNS spreads readily through hibernating bats is not a surprise. The quality of environmental data also plays a role in the output of a SDM. WorldClim data are interpolated climatic data from weather stations. The issues with interpolation are beyond the scope of this paper, but local weather fluctuations may be more useful in the context of modelling species and disease outbreaks over time (Fernández, Hamilton, and Kueppers 2013), and interpolated values may be more accurate in flat, highly sampled areas than rugged, sparsely sampled areas.

Parameter decisions in SDMs such as regularization and BP can clearly influence model assessment statistics, estimates of variable importance, and thus predicted habitat suitability. Comparisons among models are only valid when the models use the same landscape, background sample, species and test data (Lobo, Jiménez-Valverde, and Real 2008; Elith et al. 2011) which may hinder some of the analyses between using different settings and background locations. The methodologies, settings and parameters of MaxEnt have been investigated by several authors (Warren and Seifert 2011; Yackulic et al. 2013; Merow, Smith, and Silander 2013), however, even after these studies were published, the majority of studies utilizing MaxEnt do not outline the settings and parameters used nor follow recommendations (Morales, Fernández, and Baca-González 2017). The limited methodology section for SDM papers can be a hindrance for replicating work in different geographic extents and more importantly, may overestimate or underestimate areas where a species may be present (Morales, Fernández, and Baca-González 2017). Model parameterization is therefore an important component to incorporate for MaxEnt, especially when these models will be used for conservation and/or surveillance and future decisions relying on these model outputs.

There is a critical need for more research to be done on WNS. As bat species in North America continue to decline, it is paramount that we actively monitor their hibernation patterns, behaviour, symptoms of disease and their spread across the continent. Unfortunately for bats, prior to the discovery of WNS, little is known about their population size and global range (Committee on the Status of Species at Risk in Ontario (COSSARO) 2012) – had surveys and monitoring of bats taken place more rigorously, containment and spread of WNS could have been prevented. Early warning systems are of paramount importance for surveillance systems, not only to prevent and mitigate the spread of disease, but to ensure biodiversity and increase confidence in our models and decisions regarding wildlife. The models presented here provide locations that should be further examined, in combination with continued monitoring within hibernacula. The methods presented here could be further applied to North America (with challenges of data availability across states and provinces) to help forecast at risk populations so that we can implement actions to prevent or reduce exposure and concurrent stressors to WNS. This paper illustrated how biased occurrence data and limited data can be used in future applications concerning bats and WNS in Canada. These models can be used to gather more appropriate data for validation and for future surveillance or conservation schemes.

Figures and Table

Variable Name	**Type**	**Description**
Specimen Code	character	ID assigned to specimen (contains duplicates)
CommonName	character	Common species name
City	character	City where specimen was found
LatinName	character	Latin Name for species
CrossRef	character	Cross reference #, additional IDs for testing done
AgeCat	character	Age Category
Sex	character	Sex of specimen
Weight	character	Weight of specimen
DateFound	date	Date specimen was found
DeathDate	date	Death specimen died
Date_Recei	date	Date received by CWHC
Date_Submi	date	Date submitted to CWHC
MortState	character	Mortality state
Captive_Ca	character	Class of specimen (Wild, Rehab, Research, Farm/Zoo)
Storage	character	How specimen was stored Fixed, Frozen, Fresh or combination)
Found	character	How the specimen was found (e.g. hunted, found alive, died, etc.)
AgeT	character	Age text comment field for description of age
Diagnostic	character	Type of sample diagnosis was performed on (Carcass, blood, feces, etc)
CarcassCla	character	Carcass class and condition specimen is in
FinderAgen	character	The name of the agency who found the animal
SubmitAgen	character	The submitting agency/organization
SubmitPers	character	The name of the person who submitted the specimen
FinderPers	character	The name of the person who found the animal
History	character	Description and history of circumstances surrounding the submission
Interpreta	character	Interpretation of findings
SubSource	character	Submission Source or category of the type of organization that submitted the carcass
Latitude	numeric	Spatial coordinates
Longitude	numeric	Spatial coordinates
UTMEast	numeric	Projected coordinates
UTMNorth	numeric	Projected coordinates
DiagnosisT	character	Diagnosis type (Preliminary, final, open)
Diagnosi_1	character	Diagnosis text describing the results of the necropsy
Diagnosing	character	Lab that ran the diagnostic tests or where the necropsy was performed
DiagnosisB	character	Name of the pathologist that performed examination

Category	character	Category of mortality or diagnosis category (e.g. poisoning, trauma, emaciation, etc.)
Accuracy	character	Accuracy of category of mortality
Num_Spec	integer	Number of specimens submitted as part of the same incident
Necropsy	character	Results from Necropsy - descriptive text of necropsy results
Histology	character	Results from Histology – descriptive text of histo results
WNS_Status	character	Summary of WNS test results (positive, suspect, negative, etc…)
Bacteriolo	character	Results from all Bacteriology testing – free text
PCR	character	Results from all PCR tests (these could be for anything)
MatrixPCR	character	Results from Matrix PCR (this is the AIV screening PCR – birds only)
Virology	character	Results from all Virology testing – free text
Toxicology	character	Results from all Toxicology testing – free text
Parasitolo	character	Results from all Parasitology testing – free text
Immunohist	character	Results from all Immunohistochemistry testing – free text
Testing_Me	character	Testing methods used
West_Nile	character	WNV test result – summary of all tests (positive, negative, pending, etc…). e
AIV_H5_PCR	character	Results from H5 PCR (AIV specific – birds only)
AIV_H7_PCR	character	Results from H7 PCR (AIV specific – birds only)
Calculated/Derived fields for spatial analysis		
airport	numeric	distance of specimen to airports
majrods	numeric	distance of specimen to major roads
popltnc	numeric	distance of specimen to population centers
prvp_f_	numeric	distance of specimen to provincial and federal parks
watrbdy	numeric	distance of specimen to waterbodies
watrcrs	numeric	distance of specimen to water and rivers
species	numeric	binary classification of "bats" or "all other species" in Ontario
time	numeric	Timeliness (days) from Date Found – Date Received
month	date	Month specimen received
season	character	Season specimen received (spring, summer, fall, winter)
kingdom	character	Taxon of specimen - kingdom
phylum	character	Taxon of specimen - phylum
class	character	Taxon of specimen - class
order	character	Taxon of specimen - order
family	character	Taxon of specimen - family
genus	character	Taxon of specimen - genus
speciesname	character	Corrected Latin Name from Taxize database

Table 1: Description of CWHC database field/variable names with type and description (derived from inference) and new fields calculated for further spatial analysis and data characterization.

a. Ecological Models

Covariates	c-eco	c-eco+rb	c-eco+SS	G-eco	G-eco+rb	G-eco+SS	CW-eco
Distance to Hibernacula	83.7	61.9	71.6	13.5	34.5	3.7	97.5
Bio 6 – Min. T of Coldest Month	4.3	6.1	10.4	59.2	5	72.3	0.1
Modis Max – Maximum NDVI	8.3	23.9	9.8	9.6	44.3	7.5	1
Bio 15 – Precipitation Seasonality (Coefficient of Variation)	2.7	4	6.3	3	8.2	3.5	0.7
Distance to Abandoned Mines	1.1	4.1	2	14.7	8	12.9	0.6

b. Accessibility Models

Covariates	C-acc	G-acc
Distance to Major Roads	92.8	62.3
Distance to Airports	7.1	28.9
Distance to Population Centers	0.2	8.8

c. WNS Disease Model

Covariates	CW-dis
Distance to Hibernacula	95.2
Modis Max – Maximum NDVI	1.3
Bio 15 – Precipitation Seasonality (Coefficient of Variation)	0.9
Bio 9 - Mean Temperature of Driest Quarter	0.1
Distance to Watercourse	0.4
Distance to Abandoned mine	2.1

Table 2: Summary of permutation importance (%) for final model covariates across all models. Table 1a shows the final model covariates for all eco models, 6b the acc models and 6c the dis model.

Model	AUC Train	AUC Test
C-eco	0.946	0.941
C-acc	0.928	0.915
C-eco+RB	0.866	0.846
C-eco+SS	0.943	0.929
G-eco	0.876	0.878
G-acc	0.912	0.920
G-eco+RB	0.773	0.768
G-eco+SS	0.887	0.909
CW-dis	0.975	0.978
CW-eco	0.974	0.976

Table 3: Comparison of AUC training and Test Values across all models

	Model	I	D	SSIM
1	C-acc vs G-acc	0.97	0.85	0.80
2	C-acc vs C-eco	0.87	0.64	0.57
3	C-acc vs C-eco+RB	0.85	0.61	0.47
4	C-acc vs C-eco+ss	0.88	0.66	0.57
5	C-eco vs C-eco+RB	0.97	0.87	0.83
6	C-eco vs C-eco+ss	0.99	0.90	0.96
7	C-eco VS CW-eco	0.87	0.62	0.69
8	C-eco vs CW-dis	0.86	0.60	0.69
9	CW-eco vs CW-dis	0.99	0.92	0.98
10	G-acc vs G-eco	0.90	0.67	0.51
11	C-eco vs G-eco	0.95	0.80	0.65
12	G-acc vs G-eco+rb	0.67	0.38	0.12
13	G-acc vs G-eco+ss	0.90	0.66	0.46
14	G-eco vs CW-eco	0.76	0.49	0.43
15	G-eco vs CW-dis	0.75	0.48	0.43
16	G-eco vs G-eco+rb	0.71	0.45	0.45
17	G-eco vs G-eco+ss	0.99	0.89	0.86

Table 4: Comparison of Models based on global metrics niche I, D and SSIM.

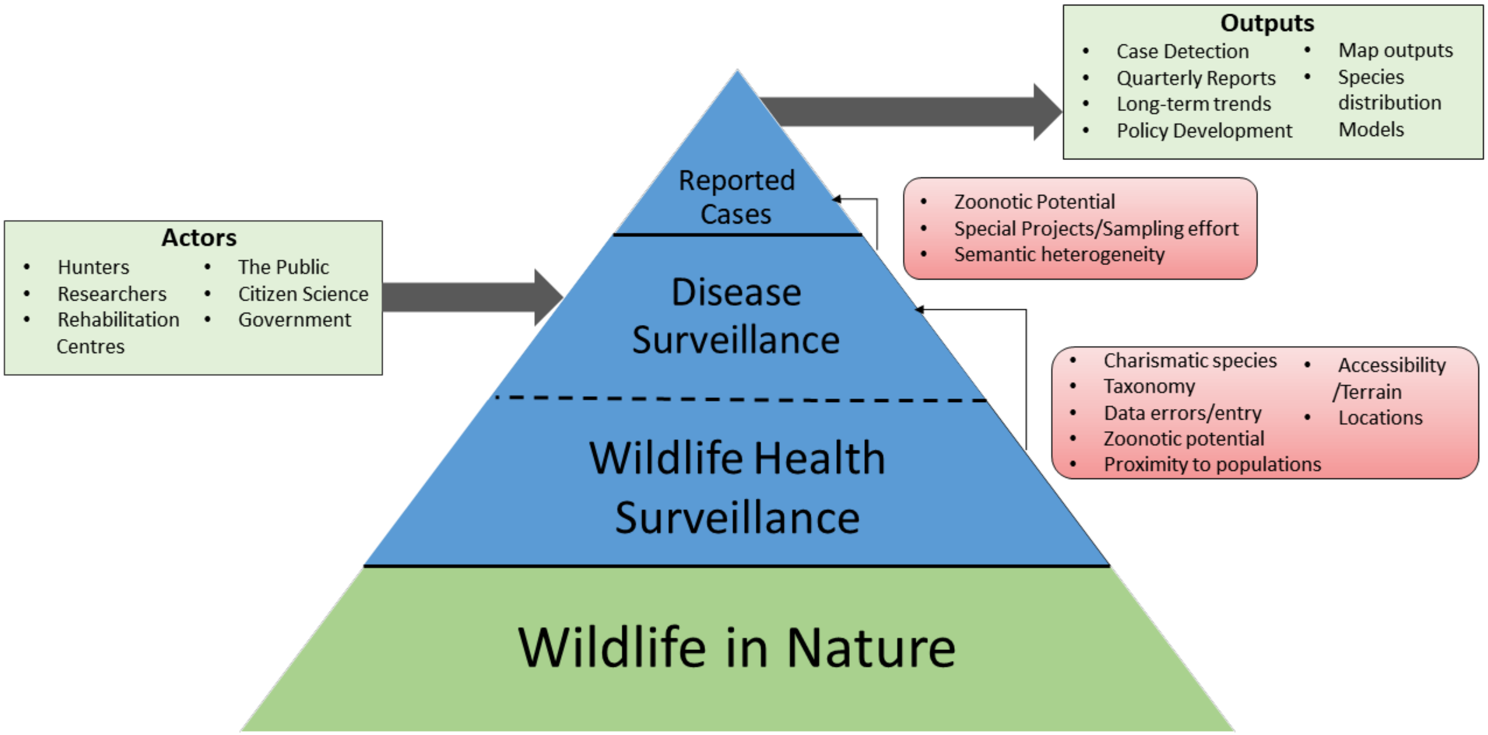


Figure 1: Conceptual Model of Surveillance of wildlife in Nature. At each step moving up the pyramid biases are introduced (in red) which contribute to the complexity of surveillance data and the methods used for the outputs. Surveillance is only capturing a fraction of what occurs in nature and depends upon many actors. The output we are exploring in detail is map outputs and species distribution models.

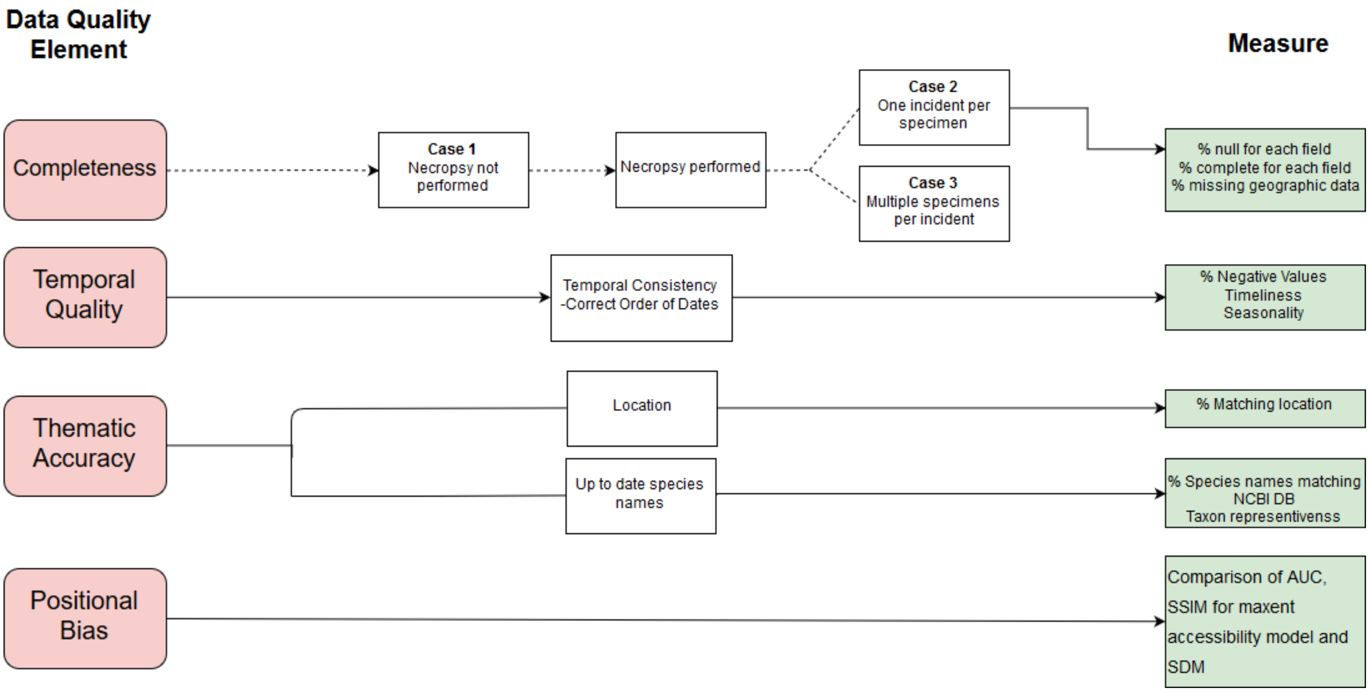


Figure 2: Data Quality Framework. The data quality elements are listed on the left while the steps between them indicate the set of conditions that were then measured.

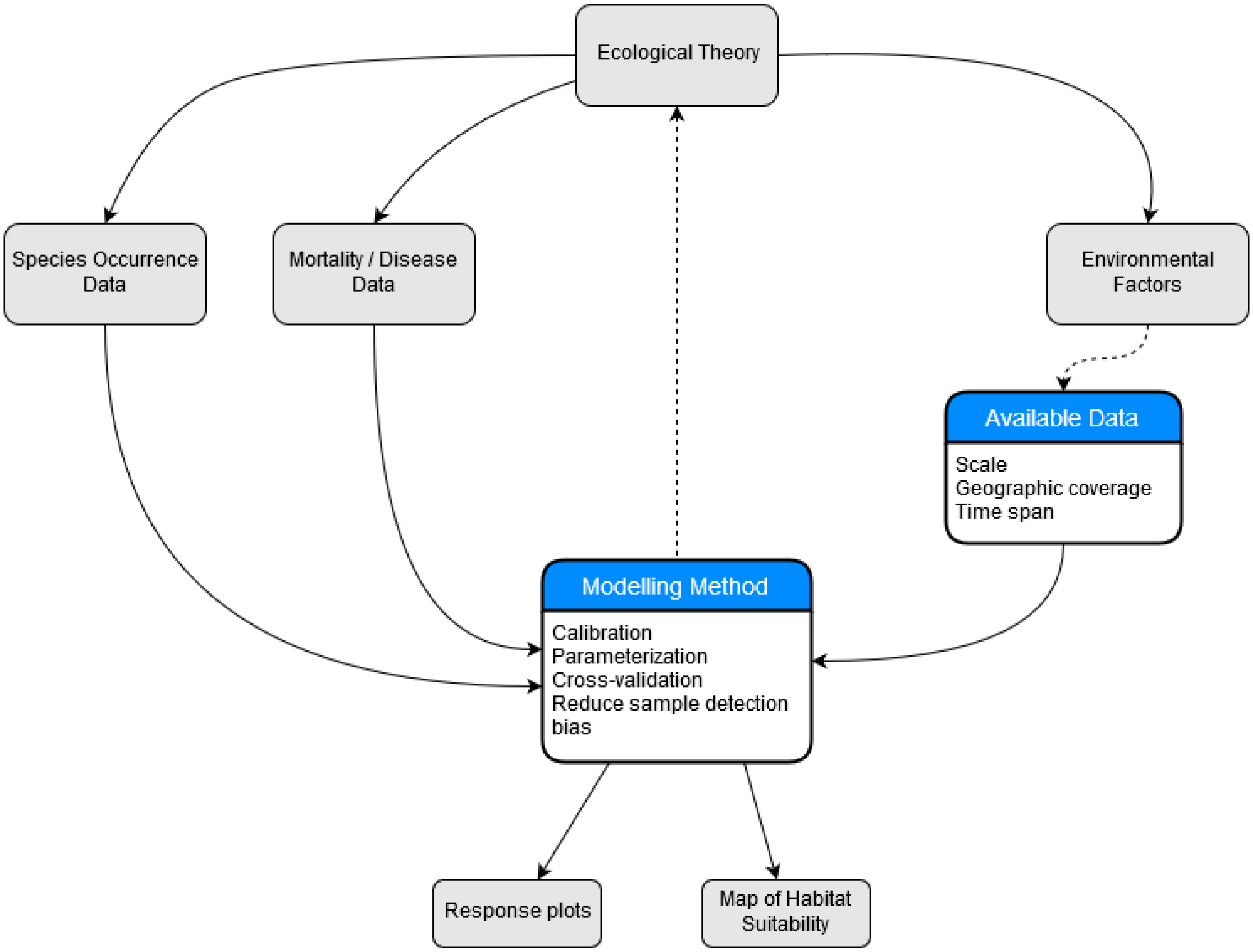


Figure 3: Framework for species distribution/disease modelling. The robustness of a model is dependent on ecological theory and previous studies. Environmental factors that may reflect ecological theory are limited by available data, which is complicated by spatial scale, geographic coverage and the time span that environmental factors are available. The data used for modelling may be the "best available" data that may not be a perfect fit for our occurrence and mortality data. The modelling method also has conditions and specifications within it that will affect the outputs of response plots and final maps of habitat suitability.

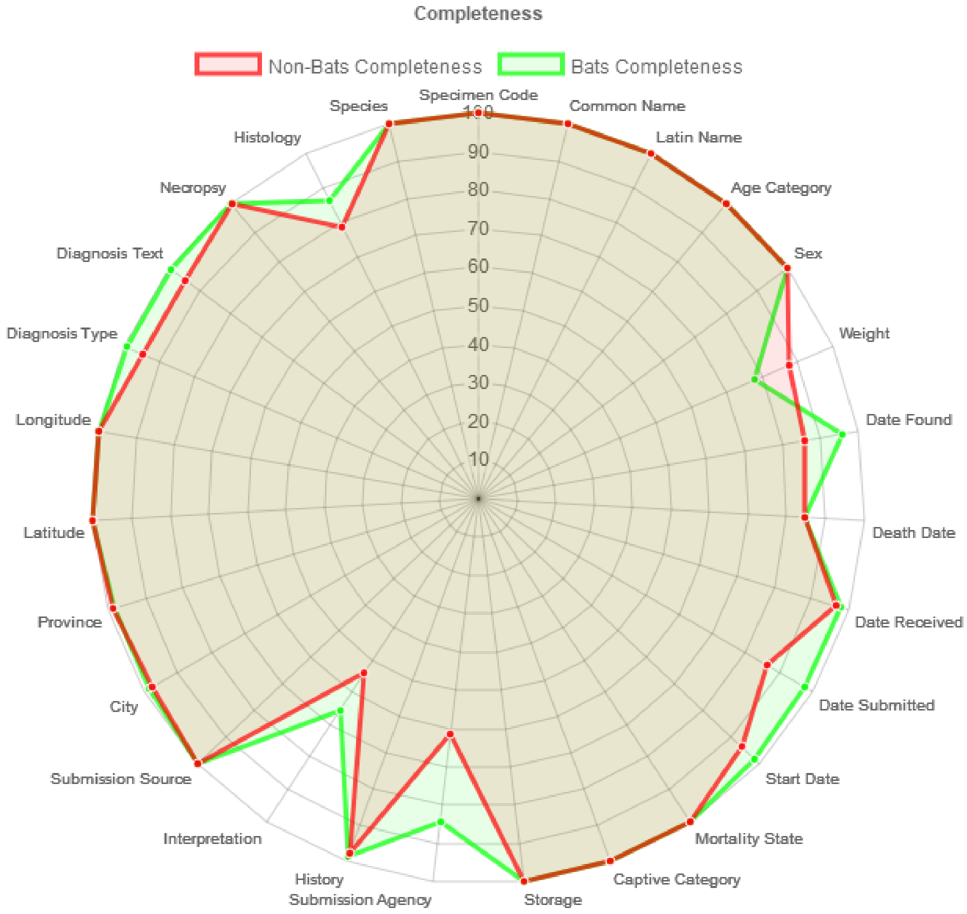


Figure 4:Completeness of fields where the outer rings indicate % 100 complete (no missing values) , red = all species in Ontario, green = bats in Ontario;.

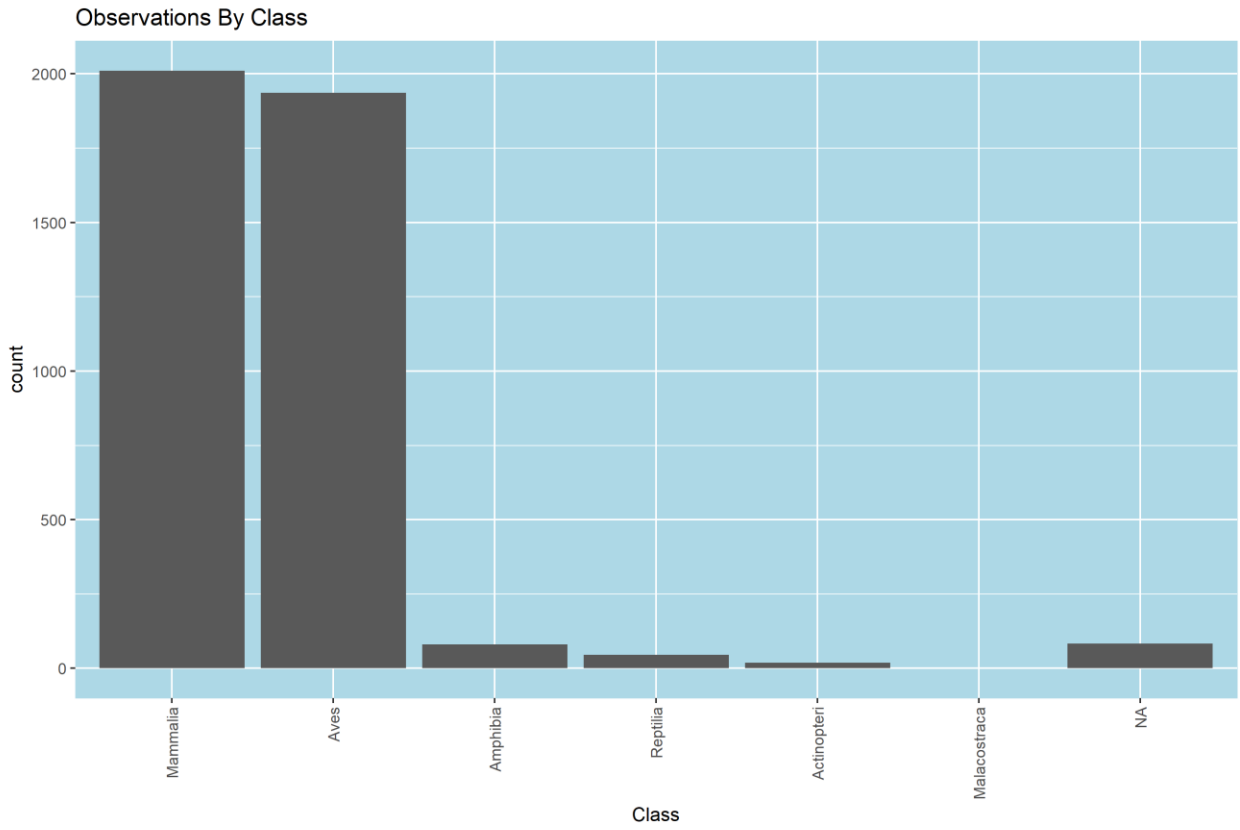


Figure 5:Histogram representing count of observations by Class.

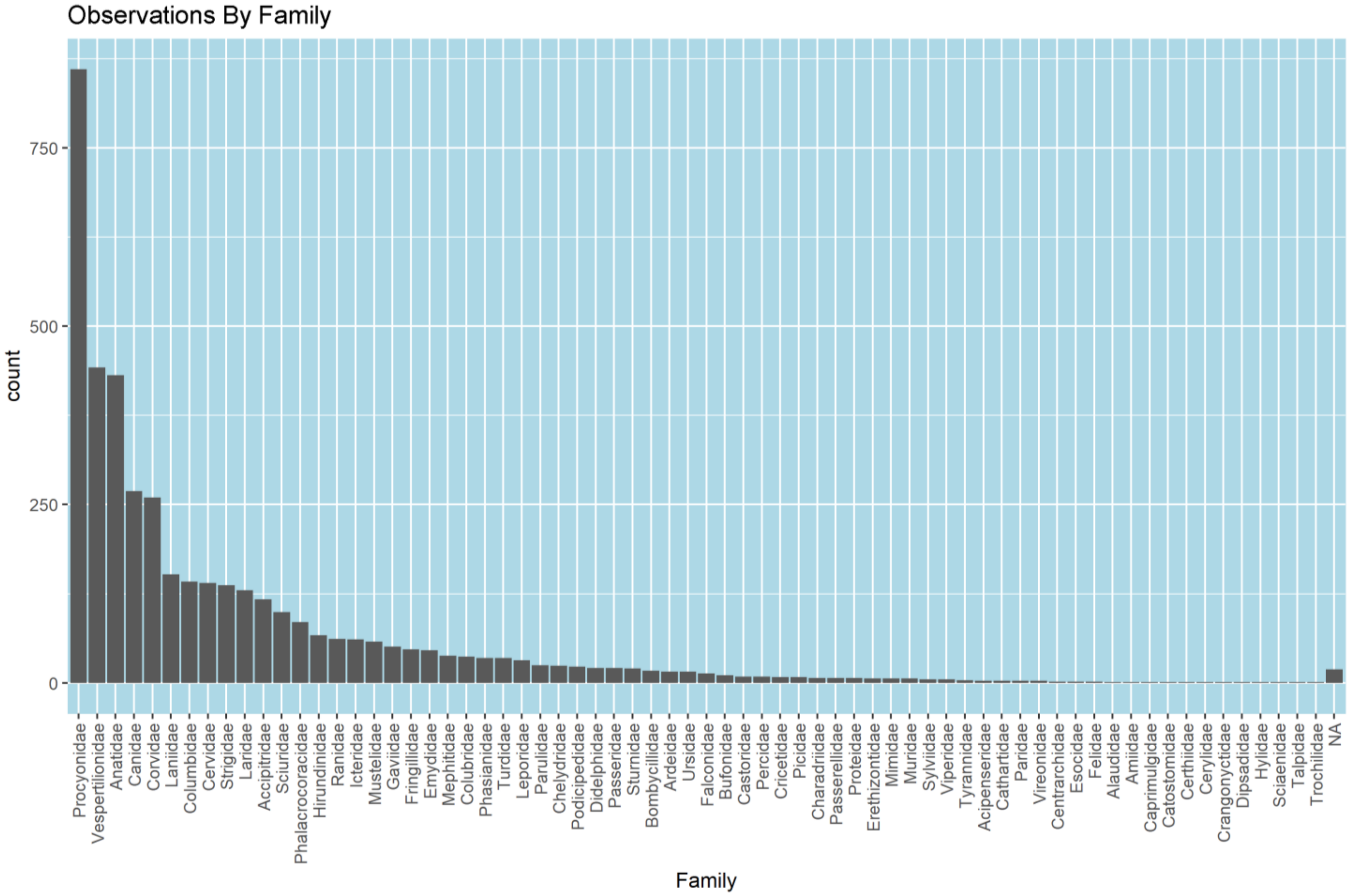


Figure 6:Histogram representing count of observations by Family.

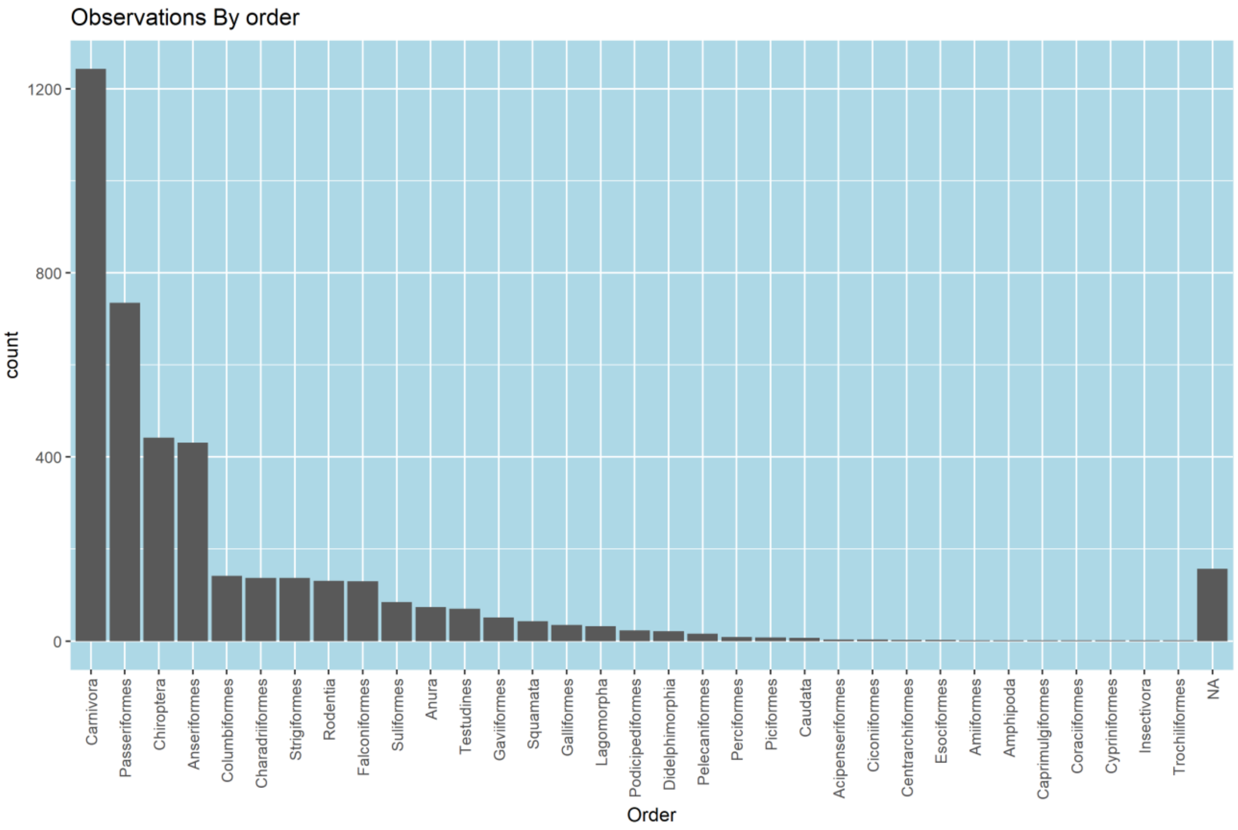


Figure 7: Histogram representing count of observations by Order.

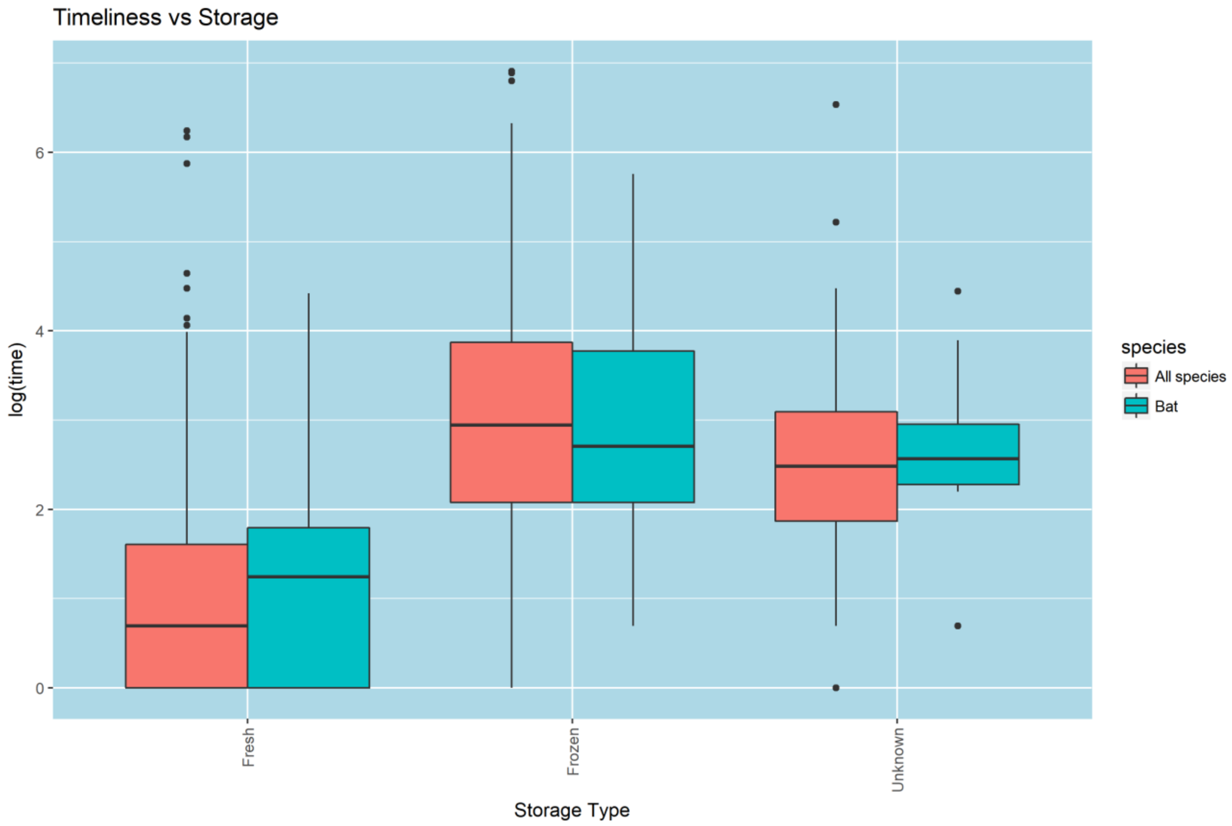


Figure 8: Log timeliness versus storage box plot. All species are in pink and bat species are in blue. The points represent outliers and horizonal lines are the median values.

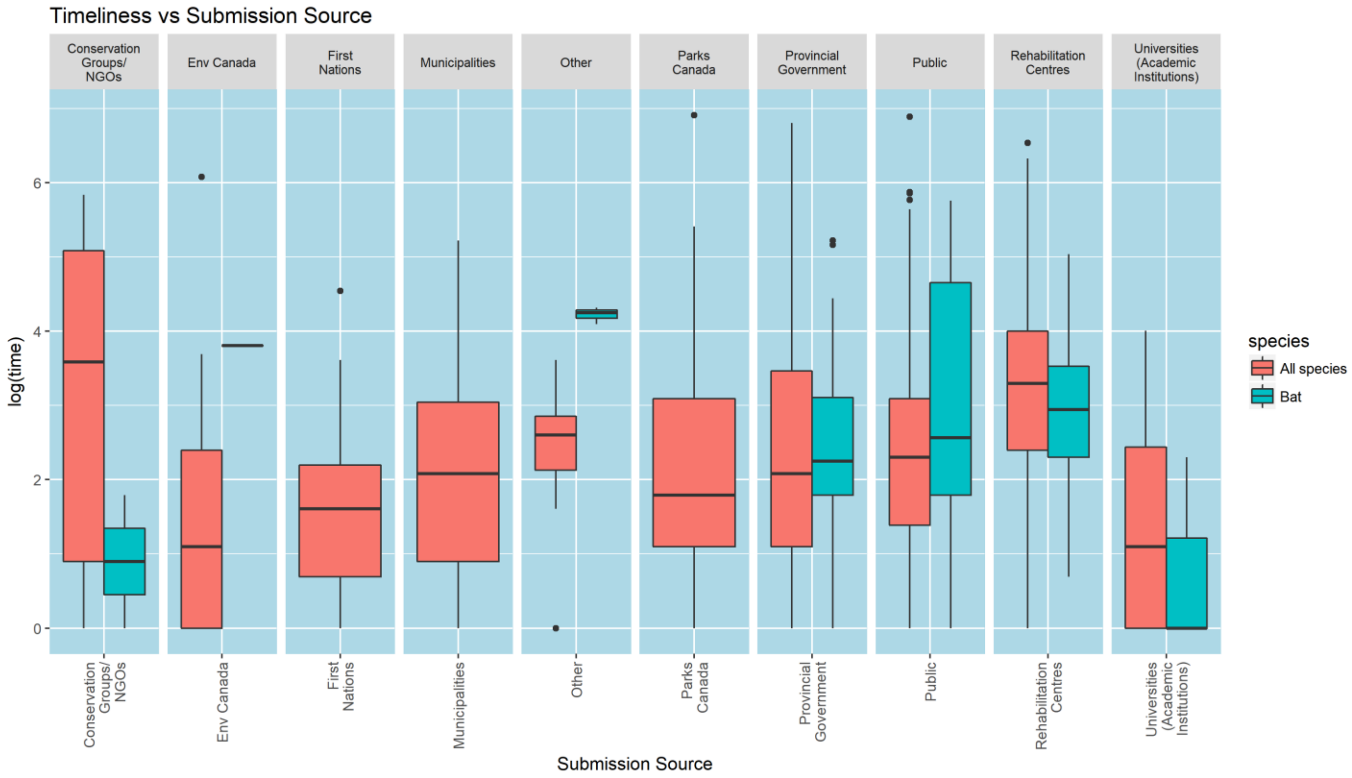


Figure 9: Log timeliness versus submission source where All species are in pink and bat species are in blue. The points represent outliers and horizontal lines are the median values.

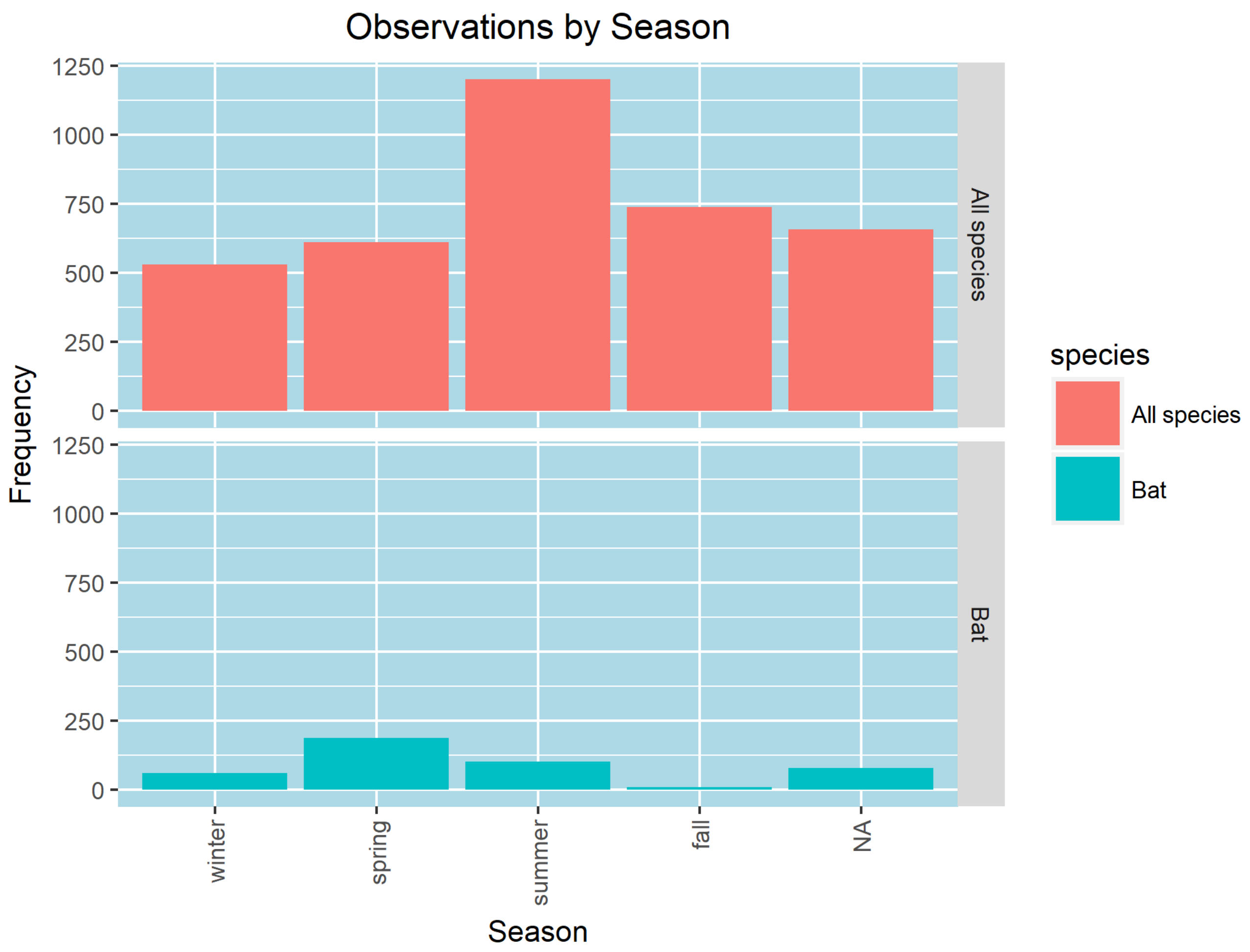
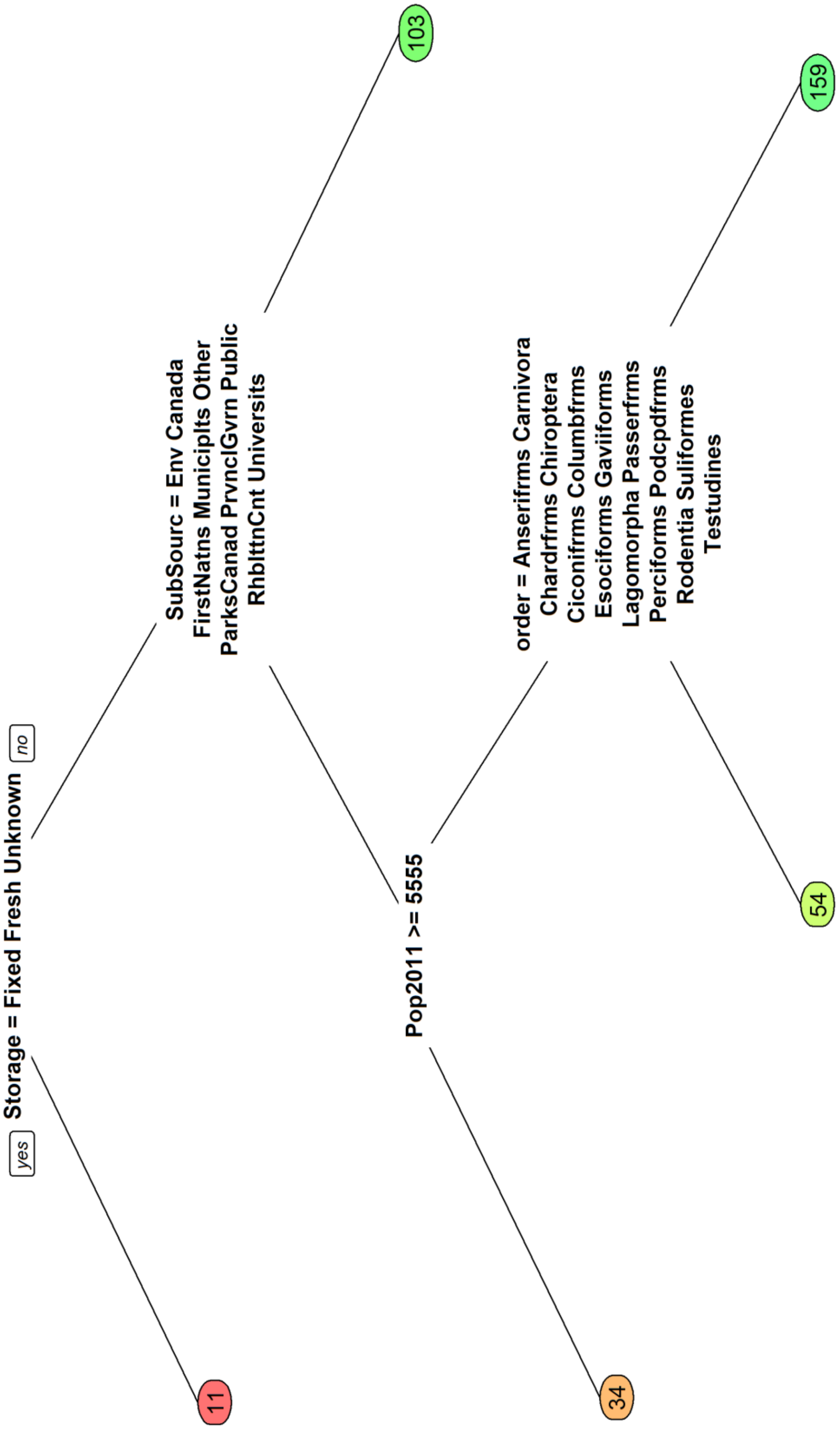


Figure 10: Frequency of observations by seasons of all species (pink) and bats (blue)



.

Figure 11: Regression tree with dependent variable timeliness

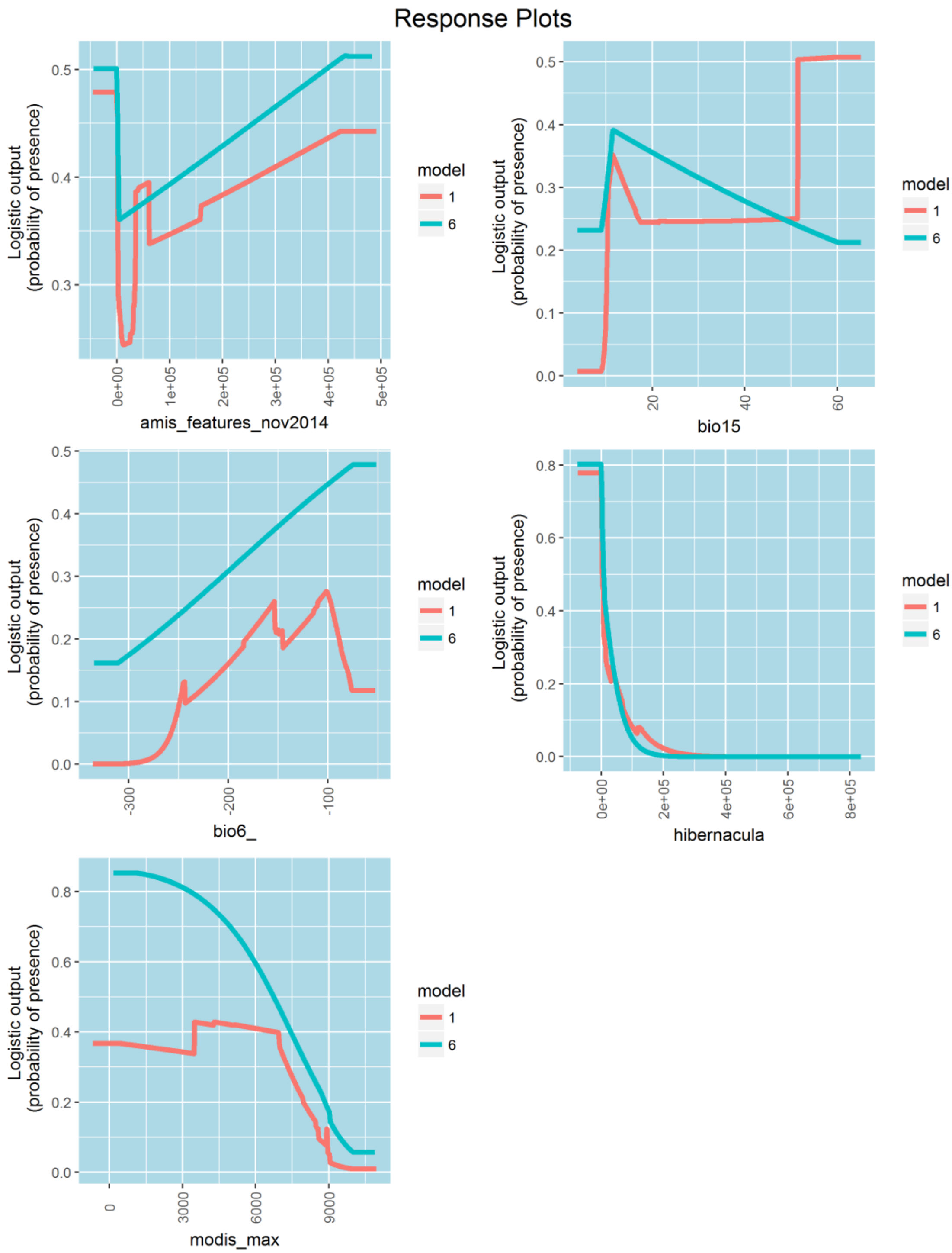


Figure 12: Response plots for differing regularization where the default regularization (pink) is irregular compared to regularization of 6 (blue) which has smoother, realistic estimates of bat responses.

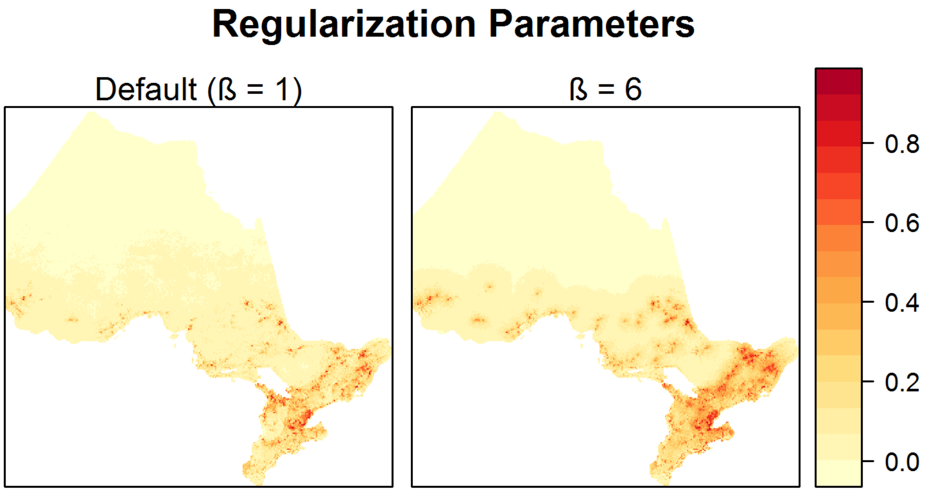


Figure 13: MaxEnt Output with default regularization (left) and regularization of 6 (right).

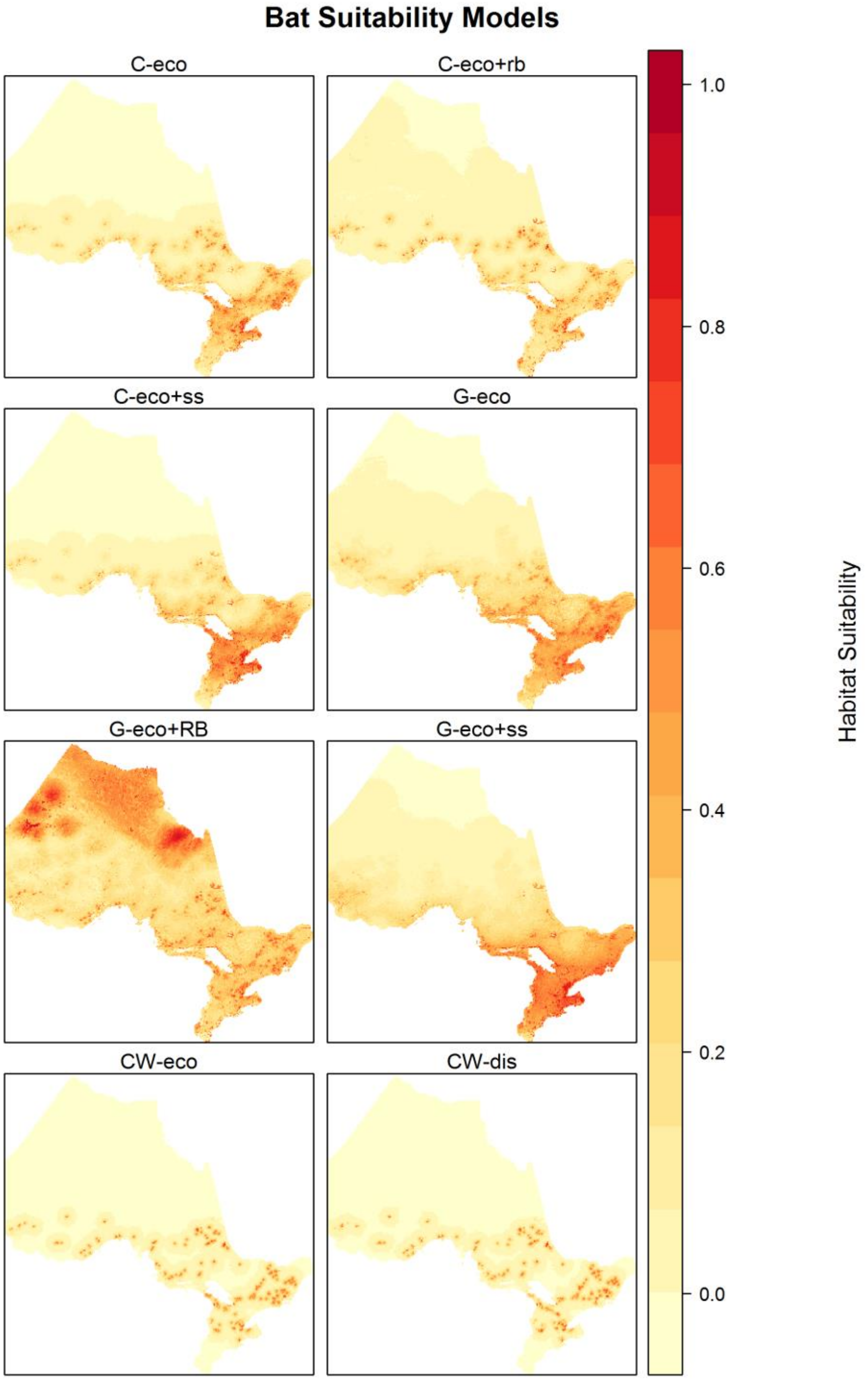


Figure 14: Final habitat suitability models where red=more suitable habitat.

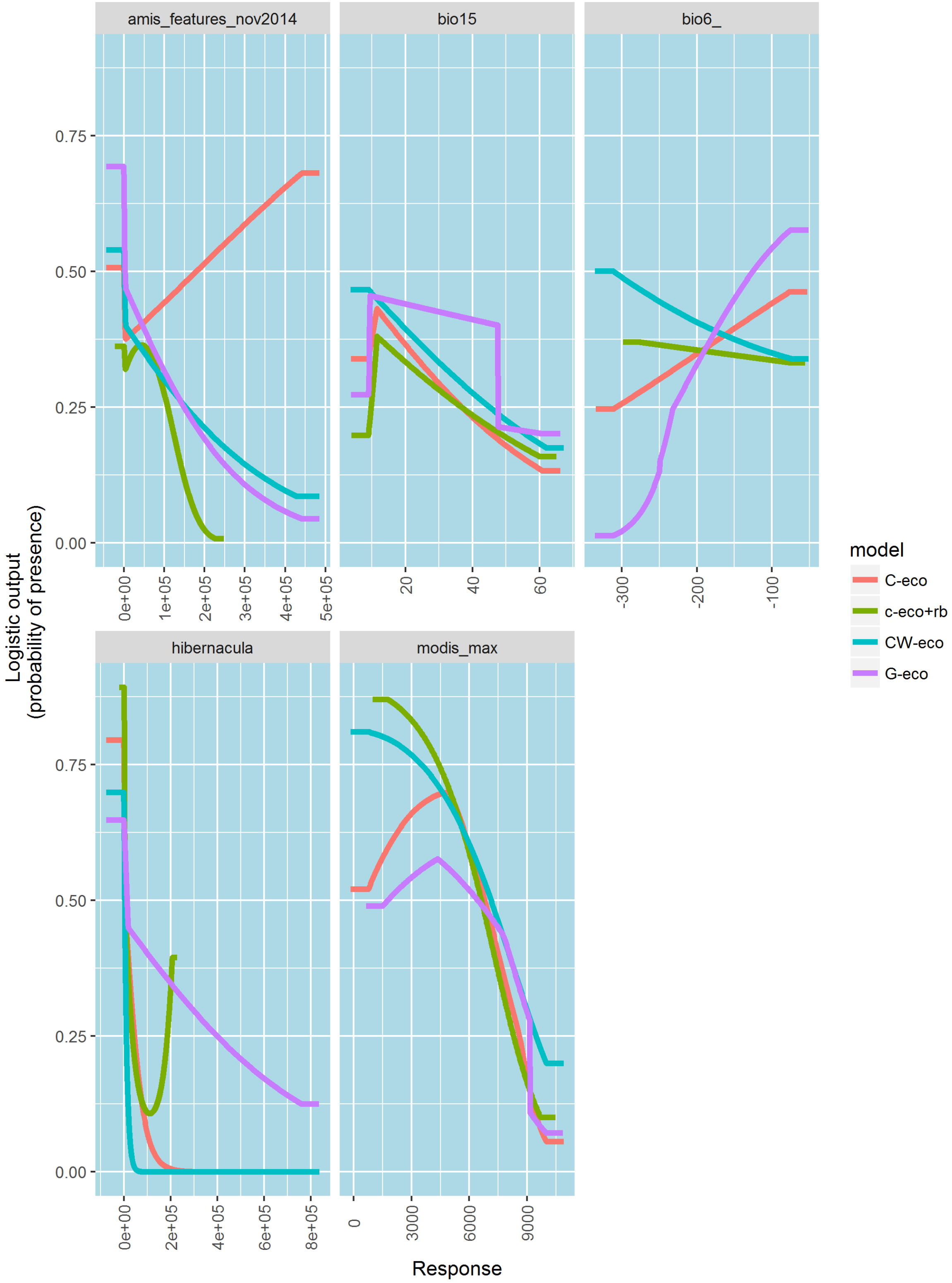


Figure 15: Response plots for C-eco (pink), C-eco+RB (green), CW-eco (blue), G-eco (purple).

Where y=logistic output and x=Response.

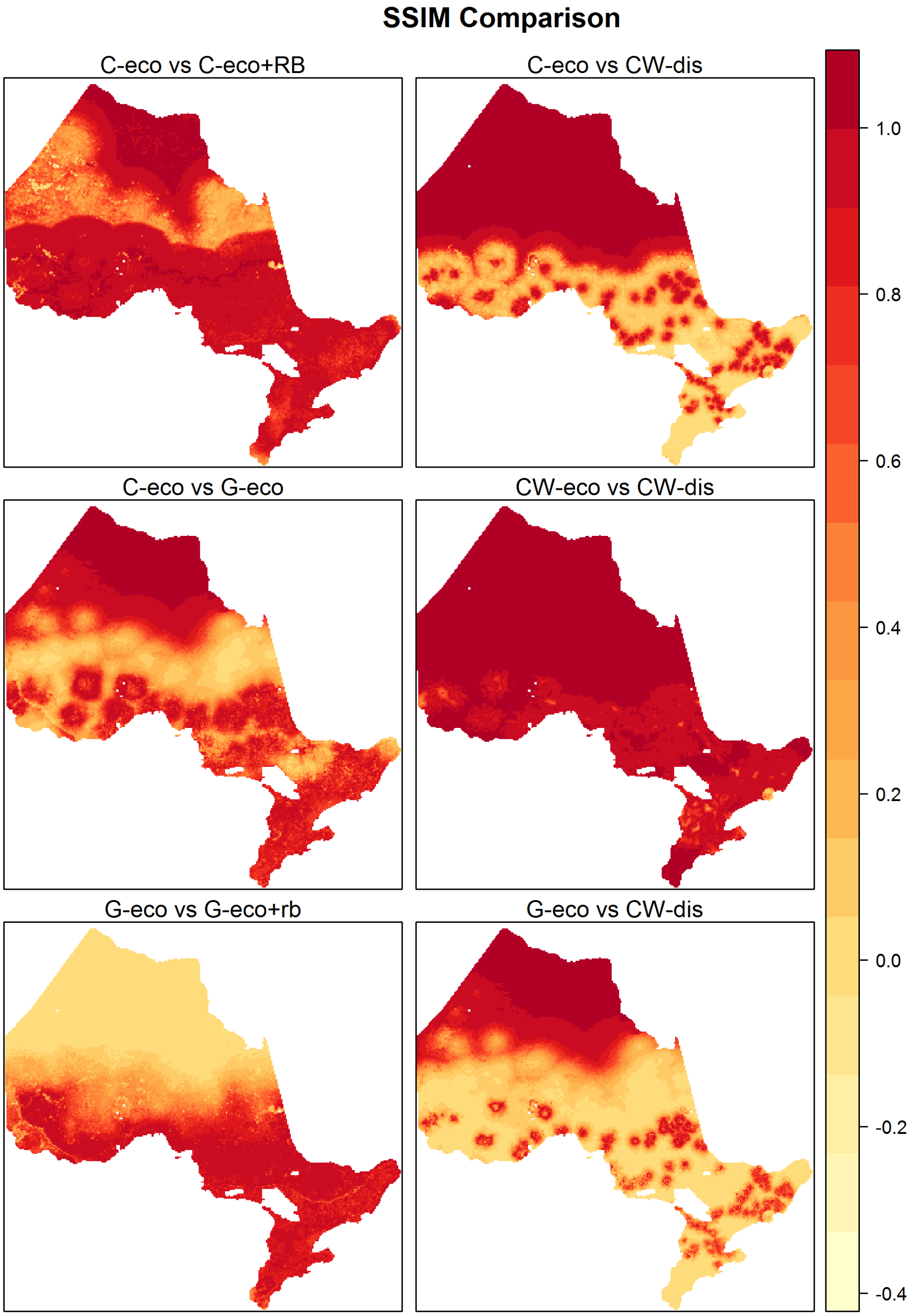


Figure 16: Structural Similarity Index comparison (-1 to 1) where values closer to 1 indicate similar underlying structural similarity and values <0 indicate diverging habitat suitability.

7 References

"Access Online: OIE - World Organisation for Animal Health." n.d. Accessed July 15, 2015. http://www.oie.int/index.php?id=169&L=0&htmfile=chapitre_surveillance_general.htm.

Aguirre, A. 2009. "Essential Veterinary Education in Zoological and Wildlife Medicine: A Global Perspective." *Revue Scientifique et Technique - Office International Des Épizooties* 28 (2): 605–10.

Alves, Davi M. C. C., Levi C. Terribile, and Daniel Brito. 2014. "The Potential Impact of White-Nose Syndrome on the Conservation Status of North American Bats." Edited by R. Mark Brigham. *PLoS ONE* 9 (9): e107395. https://doi.org/10.1371/journal.pone.0107395.

"Animal Health in the World: OIE - World Organisation for Animal Health." n.d. Accessed October 15, 2017. http://www.oie.int/en/animal-health-in-the-world/oie-listed-diseases-2014/.

Anonymous. 2004. "Report of the WHO/FAO/OIE Joint Consultation on Emerging Zoonotic Diseases. Food and Agriculture Organization of the United Nations (FAO). World Health Orgnization (WHO), and Word Organisation for Animal Health(OIE), Geneva, Switzerland." http://apps.who.int/iris/bitstream/10665/68899/1/WHO_CDS_CPE_ZFK_2004.9.pdf.

Araújo, Miguel B., and Antoine Guisan. 2006. "Five (or so) Challenges for Species Distribution Modelling." *Journal of Biogeography* 33 (10): 1677–88. https://doi.org/10.1111/j.1365-2699.2006.01584.x.

Artois, Marc, Roy Bengis, Richard J. Delahay, Marie-José Duchêne, J. Paul Duff, Ezio Ferroglio, Christian Gortazar, et al. 2009. "Wildlife Disease Surveillance and Monitoring." In *Management of Disease in Wild Mammals*, edited by Richard J. Delahay, Graham C. Smith, and Michael R. Hutchings, 187–213. Springer Japan. http://link.springer.com/chapter/10.1007/978-4-431-77134-0_10.

Barbour, Roger W Davis, and Wayne H Davis. 1969. *Bats of America*. 599.40973 B3. Lexington: University Press of Kentucky.

Barry, Simon, and Jane Elith. 2006. "Error and Uncertainty in Habitat Models." *Journal of Applied Ecology* 43 (3): 413–23.

Beck, Jan, Marianne Böller, Andreas Erhardt, and Wolfgang Schwanghart. 2014. "Spatial Bias in the GBIF Database and Its Effect on Modeling Species' Geographic Distributions." *Ecological Informatics* 19 (January): 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002.

Bevins, Sarah N., Kerri Pedersen, Mark W. Lutman, John A. Baroch, Brandon S. Schmit, Dennis Kohler, Thomas Gidlewski, Dale L. Nolte, Seth R. Swafford, and Thomas J. DeLiberto. 2014. "Large-Scale Avian Influenza Surveillance in Wild Birds throughout the United States." *PLoS ONE* 9 (8). https://doi.org/10.1371/journal.pone.0104360.

Bird Studies Canada, Canadian Wind Energy Association, Environment Canada, and Ontario Ministry of Natural Resources. 2017. "Wind Energy and Bat Monitoring Database Summary of the Findings from Post-Construction Monitoring Reports." https://www.bsc-eoc.org/resources/wind/Jul2017_Wind_Database_Summary.pdf.

Blehert, David S., Alan C. Hicks, Melissa Behr, Carol U. Meteyer, Brenda M. Berlowski-Zier, Elizabeth L. Buckles, Jeremy T. H. Coleman, et al. 2009. "Bat White-Nose Syndrome: An Emerging Fungal Pathogen?" *Science* 323 (5911): 227–227. https://doi.org/10.1126/science.1163874.

Boakes, Elizabeth H., Philip J. K. McGowan, Richard A. Fuller, Ding Chang-qing, Natalie E. Clark, Kim O'Connor, and Georgina M. Mace. 2010. "Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data." *PLOS Biology* 8 (6): e1000385. https://doi.org/10.1371/journal.pbio.1000385.

Boots, Barry. 2000. "Using GIS to Promote Spatial Analysis." *Journal of Geographical Systems* 2 (1): 17–21. https://doi.org/10.1007/s101090050024.

Boria, Robert A., Link E. Olson, Steven M. Goodman, and Robert P. Anderson. 2014. "Spatial Filtering to Reduce Sampling Bias Can Improve the Performance of Ecological Niche Models." *Ecological Modelling* 275 (Supplement C): 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012.

Bradley, Catherine A., and Sonia Altizer. 2007. "Urbanization and the Ecology of Wildlife Diseases." *Trends in Ecology & Evolution* 22 (2): 95–102. https://doi.org/10.1016/j.tree.2006.11.001.

Breiman, Leo. 1984. *Classification and Regression Trees*. Wadsworth International Group.

Broennimann, Olivier, Matthew C. Fitzpatrick, Peter B. Pearman, Blaise Petitpierre, Loïc Pellissier, Nigel G. Yoccoz, Wilfried Thuiller, et al. 2012. "Measuring Ecological Niche Overlap from Occurrence and Spatial Environmental Data." *Global Ecology and Biogeography* 21 (4): 481–97. https://doi.org/10.1111/j.1466-8238.2011.00698.x.

Brownstein, John S., Hilary Rosen, Dianne Purdy, James R. Miller, Mario Merlino, Farzad Mostashari, and Durland Fish. 2004. "Spatial Analysis of West Nile Virus: Rapid Risk Assessment of an Introduced Vector-Borne Zoonosis." Research-article. Http://Www.liebertpub.com/Vbz. July 9, 2004. https://doi.org/10.1089/15303660260613729.

Canadian Wildlife Health Cooperative. n.d. "Wildlife Disease DataBase Overview." Accessed October 1, 2017. http://www.cwhc-rcsf.ca/docs/CWHC%20Database%20Overview.pdf.

Chamberlain, Scott, Eduard Szocs, Carl Boettiger, Karthik Ram, Ignasi Bartomeus, John Baumgartner, Zachary Foster, and James O'Donnell. 2016. *Taxize: Taxonomic Information from around the Web*. https://github.com/ropensci/taxize.

Childs, James E., Aaron T. Curns, Meghan E. Dey, Leslie A. Real, Lisa Feinstein, Ottar N. Bjørnstad, and John W. Krebs. 2000. "Predicting the Local Dynamics of Epizootic Rabies among Raccoons in the United States." *Proceedings of the National Academy of Sciences* 97 (25): 13666–71. https://doi.org/10.1073/pnas.240326697.

Childs, James E., J. W. Krebs, L. A. Real, and E. R. Gordon. 2007. "Animal-Based National Surveillance for Zoonotic Disease: Quality, Limitations, and Implications of a Model System for Monitoring Rabies." *Preventive Veterinary Medicine* 78 (3): 246–61. https://doi.org/10.1016/j.prevetmed.2006.10.014.

Committee on the Status of Species at Risk in Ontario (COSSARO). 2012. "COSSARO Candidate Species at Risk Evaluation for Little Brown Myotis (Myotis Lucifugus)." http://files.ontario.ca/environment-and-energy/species-at-risk/stdprod_101774.pdf.

Cunningham, Andrew A. 2005. "A Walk on the Wild Side—emerging Wildlife Diseases." *BMJ : British Medical Journal* 331 (7527): 1214–15.

"CWHC - Canadian Wildlife Health Cooperative." n.d. Accessed May 25, 2017a. http://www.cwhc-rcsf.ca/who_we_are.php.

———. "CWHC - Canadian Wildlife Health Cooperative." n.d. Accessed February 21, 2017b. http://www.cwhc-rcsf.ca/wildlife_disease_database.php.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2001. “Anthropogenic Environmental Change and the Emergence of Infectious Diseases in Wildlife.” *Acta Tropica* 78 (2): 103–16. https://doi.org/10.1016/S0001-706X(00)00179-0.

Daszak, P, Andrew A. Cunningham, and Alex D. Hyatt. 2000. “Emerging Infectious Diseases of Wildlife-- Threats to Biodiversity and Human Health.” *Science* 287 (5452): 443–49. https://doi.org/10.1126/science.287.5452.443.

Daszak, Peter, Carlos Zambrana-Torrelio, Tiffany L. Bogich, Miguel Fernandez, Jonathan H. Epstein, Kris A. Murray, and Healy Hamilton. 2013. “Interdisciplinary Approaches to Understanding Disease Emergence: The Past, Present, and Future Drivers of Nipah Virus Emergence.” *Proceedings of the National Academy of Sciences* 110 (Supplement 1): 3681–88. https://doi.org/10.1073/pnas.1201243109.

De’ath, Glenn, and Katharina E. Fabricius. 2000. “Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis.” *Ecology* 81 (11): 3178–92. https://doi.org/10.2307/177409.

Deem, Sharon L., William B. Karesh, and Wendy Weisman. 2001. “Putting Theory into Practice: Wildlife Health in Conservation.” *Conservation Biology* 15 (5): 1224–33. https://doi.org/10.1111/j.1523-1739.2001.00336.x.

Dennis, R. L. H., and C. D. Thomas. 2000. “Bias in Butterfly Distribution Maps: The Influence of Hot Spots and Recorder’s Home Range.” *Journal of Insect Conservation* 4 (2): 73–77. https://doi.org/10.1023/A:1009690919835.

Devillers, R., Y. Bédard, R. Jeansoulin, and B. Moulin. 2007. “Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data.” *International Journal of Geographical Information Science* 21 (3): 261–82. https://doi.org/10.1080/13658810600911879.

Dijk, Jacintha G. B. van, Bethany J. Hoye, Josanne H. Verhagen, Bart A. Nolet, Ron A. M. Fouchier, and Marcel Klaassen. 2014. “Juveniles and Migrants as Drivers for Seasonal Epizootics of Avian Influenza Virus.” *Journal of Animal Ecology* 83 (1): 266–75. https://doi.org/10.1111/1365-2656.12131.

Duan, Ren-Yan, Xiao-Quan Kong, Min-Yi Huang, Wei-Yi Fan, and Zhi-Gao Wang. 2014. “The Predictive Performance and Stability of Six Species Distribution Models.” *PLoS ONE* 9 (11). https://doi.org/10.1371/journal.pone.0112764.

Dubois, John E., and Kimberly M. Monson. 2007. “Recent Distribution Records of the Little Brown Bat, Myotis Lucifugus , in Manitoba and Northwestern Ontario.” *The Canadian Field-Naturalist* 121 (1): 57–61.

Duncan, C., L. Backus, T. Lynn, B. Powers, and M. Salman. 2008. “Passive, Opportunistic Wildlife Disease Surveillance in the Rocky Mountain Region, USA.” *Transboundary and Emerging Diseases* 55 (7): 308–14. https://doi.org/10.1111/j.1865-1682.2008.01039.x.

Ehlman, Sean M., John J. Cox, and Philip H. Crowley. 2013. “Evaporative Water Loss, Spatial Distributions, and Survival in White-Nose-Syndrome–affected Little Brown Myotis: A Model.” *Journal of Mammalogy* 94 (3): 572–83. https://doi.org/10.1644/12-MAMM-A-111.1.

Eisen, Lars, and Rebecca J. Eisen. 2010. “Using Geographic Information Systems and Decision Support Systems for the Prediction, Prevention, and Control of Vector-Borne Diseases.” *Annual Review of Entomology* 56 (1): 41–61. https://doi.org/10.1146/annurev-ento-120709-144847.

Elith, Jane, Catherine H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, et al. 2006. “Novel Methods Improve Prediction of Species’ Distributions from Occurrence Data.” *Ecography* 29 (2): 129–51.

Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. “A Statistical Explanation of MaxEnt for Ecologists.” *Diversity and Distributions* 17 (1): 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x.

Elliott, Paul, and Daniel Wartenberg. 2004. “Spatial Epidemiology: Current Approaches and Future Challenges.” *Environmental Health Perspectives* 112 (9): 998–1006. https://doi.org/10.1289/ehp.6735.

Esri. 2015. *ArcGIS Desktop: Release 10.3.1.* Redlands, CA: Environmental Systems Research Institute.

Fenton, M. Brock. 1969. “Summer Activity of Myotis Lucifugus (Chiroptera:Vespertilionidae) at Hibernacula in Ontario and Quebec.” *Canadian Journal of Zoology* 47 (4): 597–602. https://doi.org/10.1139/z69-103.

Fernández, M., H. Hamilton, and L. M. Kueppers. 2013. “Characterizing Uncertainty in Species Distribution Models Derived from Interpolated Weather Station Data.” *Ecosphere* 4 (5): 1–17. https://doi.org/10.1890/ES13-00049.1.

Fielding, Alan H., and John F. Bell. 1997. “A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models.” *Environmental Conservation* 24 (1): 38–49.

Flory, Abigail R., Sunil Kumar, Thomas J. Stohlgren, and Paul M. Cryan. 2012. “Environmental Conditions Associated with Bat White-Nose Syndrome Mortality in the North-Eastern United States.” *Journal of Applied Ecology* 49 (3): 680–89. https://doi.org/10.1111/j.1365-2664.2012.02129.x.

Foley, Janet, Deana Clifford, Kevin Castle, Paul Cryan, and Richard S. Ostfeld. 2011. “Investigating and Managing the Rapid Emergence of White-Nose Syndrome, a Novel, Fatal, Infectious Disease of Hibernating Bats: White-Nose Syndrome in Bats.” *Conservation Biology*, January, no-no. https://doi.org/10.1111/j.1523-1739.2010.01638.x.

Fourcade, Yoan, Jan O. Engler, Dennis Rödder, and Jean Secondi. 2014. “Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias.” *PLOS ONE* 9 (5): e97122. https://doi.org/10.1371/journal.pone.0097122.

Franklin, Janet, and Jennifer A. Miller. 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810602.

Frick, Winifred F., D. Scott Reynolds, and Thomas H. Kunz. 2010. “Influence of Climate and Reproductive Timing on Demography of Little Brown Myotis Myotis Lucifugus.” *Journal of Animal Ecology* 79 (1): 128–36. https://doi.org/10.1111/j.1365-2656.2009.01615.x.

Gardner, I.A., S. Hietala, and W.M. Boyce. 1996. “Validity of Using Serological Tests for Diagnosis of Diseases in Wild Animals.” *Revue Scientifique et Technique (International Office of Epizootics)* 15 (1): 323–35.

Gervais, Marc, Yvan Bedard, M. Levesque, Eveline Bernier, and Rodolphe Devillers. 2009. “Data Quality Issues and Geographic Knowledge Discovery.” *Geographic Data Mining and Knowledge Discovery*, 99–115.

Goodchild, Michael F. 2010. “Twenty Years of Progress: GIScience in 2010.” *Journal of Spatial Information Science* 2010 (1): 3–20. https://doi.org/10.5311/JOSIS.2010.1.2.

Government of Canada, Environment Canada. 2012. “Species at Risk Public Registry - Technical Summary and Supporting Information for an Emergency Assessment of the Little Brown Myotis Myotis Lucifugus – 2012.” February 23, 2012. http://www.registrelep-sararegistry.gc.ca/default.asp?lang=En&n=2A256834-1.

———. 2015. “Species at Risk Public Registry - Recovery Strategy for Little Brown Myotis (Myotis Lucifugus), Northern Myotis (Myotis Septentrionalis), and Tri-Colored Bat (Perimyotis Subflavus) in Canada - 2015 [Proposed].” November 12, 2015. http://www.registrelep-sararegistry.gc.ca/default.asp?lang=En&n=2A04680B-1#_03_2.

Graham, Catherine H, Jane Elith, Robert J Hijmans, Antoine Guisan, A Townsend Peterson, Bette A Loiselle, and The Nceas Predicting Species Distributions Working Group. 2008. “The Influence of Spatial Errors in Species Occurrence Data Used in Distribution Models.” *Journal of Applied Ecology* 45 (1): 239–47. https://doi.org/10.1111/j.1365-2664.2007.01408.x.

Grand, Joanna, Michael P. Cummings, Tony G. Rebelo, Taylor H. Ricketts, and Maile C. Neel. 2007. “Biased Data Reduce Efficiency and Effectiveness of Conservation Reserve Networks.” *Ecology Letters* 10 (5): 364–74. https://doi.org/10.1111/j.1461-0248.2007.01025.x.

Grogan, Laura F., Lee Berger, Karrie Rose, Victoria Grillo, Scott D. Cashins, and Lee F. Skerratt. 2014. “Surveillance for Emerging Biodiversity Diseases of Wildlife.” *PLOS Pathogens* 10 (5): e1004015. https://doi.org/10.1371/journal.ppat.1004015.

Hassell, James M., Michael Begon, Melissa J. Ward, and Eric M. Fèvre. 2017. “Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface.” *Trends in Ecology & Evolution* 32 (1): 55–67. https://doi.org/10.1016/j.tree.2016.09.012.

Hastie, Trevor, Robert Tibshirani, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY: Springer.

Hijmans, R.J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. “Very High Resolution Interpolated Climate Surfaces for Global Land Areas.” *International Journal of Climatology* 25: 1965–78.

Hijmans, Robert J., Steven Phillips, and John Leathwick and Jane Elith. 2017. *Dismo: Species Distribution Modeling* (version 1.1-4). https://cran.r-project.org/web/packages/dismo/index.html.

Hoinville, L. J., L. Alban, J. A. Drewe, J. C. Gibbens, L. Gustafson, B. Häsler, C. Saegerman, M. Salman, and K. D. C. Stärk. 2013. “Proposed Terms and Concepts for Describing and Evaluating Animal-Health Surveillance Systems.” *Preventive Veterinary Medicine* 112 (1): 1–12. https://doi.org/10.1016/j.prevetmed.2013.06.006.

Institute of Medicine (US) Forum on Microbial Threats. 2007. *Global Infectious Disease Surveillance and Detection: Assessing the Challenges—Finding Solutions, Workshop Summary*. The National Academies Collection: Reports Funded by National Institutes of Health. Washington (DC): National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK52867/.

ISO. 2013. “ISO 19157:2013(E) - Geographic Information - Data Quality.” ISO. 2013. http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=32575.

Jiménez-Valverde, Alberto, and Jorge M. Lobo. 2007. “Threshold Criteria for Conversion of Probability of Species Presence to Either–or Presence–absence.” *Acta Oecologica* 31 (3): 361–69. https://doi.org/10.1016/j.actao.2007.02.001.

Jones, Esther L., Luke Rendell, Enrico Pirotta, and Jed A. Long. 2016. “Novel Application of a Quantitative Spatial Comparison Tool to Species Distribution Data.” *Ecological Indicators*, Navigating Urban Complexity: Advancing Understanding of Urban Social – Ecological Systems for Transformation and Resilience, 70 (November): 67–76. https://doi.org/10.1016/j.ecolind.2016.05.051.

Jones, Kate E., Nikkita G. Patel, Marc A. Levy, Adam Storeygard, Deborah Balk, John L. Gittleman, and Peter Daszak. 2008. “Global Trends in Emerging Infectious Diseases.” *Nature* 451 (7181): 990–93. https://doi.org/10.1038/nature06536.

Judson, Seth D., Robert Fischer, Andrew Judson, and Vincent J. Munster. 2016. “Ecological Contexts of Index Cases and Spillover Events of Different Ebolaviruses.” *PLOS Pathogens* 12 (8): e1005780. https://doi.org/10.1371/journal.ppat.1005780.

Kadmon, Ronen, Oren Farber, and Avinoam Danin. 2004. “Effect of Roadside Bias on the Accuracy of Predictive Maps Produced by Bioclimatic Models.” *Ecological Applications* 14 (2): 401–13.

Kaiser, Reinhard, Paul B. Spiegel, Alden K. Henderson, and Michael L. Gerber. 2003. “The Application of Geographic Information Systems and Global Positioning Systems in Humanitarian Emergencies: Lessons Learned, Programme Implications and Future Research.” *Disasters* 27 (2): 127–40. https://doi.org/10.1111/1467-7717.00224.

Kramer-Schadt, Stephanie, Jürgen Niedballa, John D. Pilgrim, Boris Schröder, Jana Lindenborn, Vanessa Reinfelder, Milena Stillfried, et al. 2013. “The Importance of Correcting for Sampling Bias in MaxEnt Species Distribution Models.” *Diversity and Distributions* 19 (11): 1366–79. https://doi.org/10.1111/ddi.12096.

Kuiken, T., F. A. Leighton, R. A. M. Fouchier, J. W. LeDuc, J. S. M. Peiris, A. Schudel, K. Stöhr, and A. D. M. E. Osterhaus. 2005. “Pathogen Surveillance in Animals.” *Science*, New Series, 309 (5741): 1680–81.

Kulkarni, Manisha A., Lea Berrang-Ford, Peter A. Buck, Michael A. Drebot, L. Robbin Lindsay, and Nicholas H. Ogden. 2015. “Major Emerging Vector-Borne Zoonotic Diseases of Public Health Importance in Canada.” *Emerging Microbes & Infections* 4 (6): e33. https://doi.org/10.1038/emi.2015.33.

Kunz, Thomas H., Edward B. Arnett, Brian M. Cooper, Wallace P. Erickson, Ronald P. Larkin, Todd Mabee, Michael L. Morrison, M. Dale Strickland, and Joseph M. Szewczak. 2007. “Assessing Impacts of Wind-Energy Development on Nocturnally Active Birds and Bats: A Guidance Document.” *The Journal of Wildlife Management* 71 (8): 2449–86. https://doi.org/10.2307/4496367.

Kuvlesky, William P., Leonard A. Brennan, Michael L. Morrison, Kathy K. Boydston, Bart M. Ballard, and Fred C. Bryant. 2007. “Wind Energy Development and Wildlife Conservation: Challenges and Opportunities.” *The Journal of Wildlife Management* 71 (8): 2487–98.

Lai, P.C., C.M. Wong, A.J. Hedley, S.V. Lo, P.Y. Leung, J. Kong, and G.M. Leung. 2004. “Understanding the Spatial Clustering of Severe Acute Respiratory Syndrome (SARS) in Hong Kong.” *Environmental Health Perspectives* 112 (15): 1550–56. https://doi.org/10.1289/ehp.7117.

Leighton, F A, G A Wobeser, I K Barker, P Y Daoust, and D Martineau. 1997. “The Canadian Cooperative Wildlife Health Centre and Surveillance of Wild Animal Diseases in Canada.” *The Canadian Veterinary Journal* 38 (5): 279–84.

Leitão, Pedro J., Francisco Moreira, and Patrick E. Osborne. 2011. “Effects of Geographical Data Sampling Bias on Habitat Models of Species Distributions: A Case Study with Steppe Birds in Southern Portugal.” *International Journal of Geographical Information Science* 25 (3): 439–54. https://doi.org/10.1080/13658816.2010.531020.

Linard, Catherine, and Andrew J. Tatem. 2012. “Large-Scale Spatial Population Databases in Infectious Disease Research.” *International Journal of Health Geographics* 11 (March): 7. https://doi.org/10.1186/1476-072X-11-7.

Linden, Annick, Marc Wirtgen, S. Volpe, A. Nahayo, Jessica Pirson, Julien Paternostre, Fabien Grégoire, and others. 2011. “Surveillance of Wildlife Diseases in Belgium.” *Épidémiologie et Santé Animale*, no. 59/60: 213–215.

Lobo, Jorge M., Alberto Jiménez-Valverde, and Raimundo Real. 2008. “AUC: A Misleading Measure of the Performance of Predictive Distribution Models.” *Global Ecology and Biogeography* 17 (2): 145–51. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Loh, Wei-Yin, and Yu-Shan Shih. 1997. *Split Selection Methods for Classification Trees*.

Loiselle, Bette A., Peter M. Jørgensen, Trisha Consiglio, Iván Jiménez, John G. Blake, Lúcia G. Lohmann, and Olga Martha Montiel. 2008. “Predicting Species Distributions from Herbarium Collections: Does Climate Bias in Collection Sampling Influence Model Outcomes?” *Journal of Biogeography* 35 (1): 105–16. https://doi.org/10.1111/j.1365-2699.2007.01779.x.

Maher, Sean P., Andrew M. Kramer, J. Tomlin Pulliam, Marcus A. Zokan, Sarah E. Bowden, Heather D. Barton, Krisztian Magori, and John M. Drake. 2012. “Spread of White-Nose Syndrome on a Network Regulated by Geography and Climate.” *Nature Communications* 3 (December): 1306. https://doi.org/10.1038/ncomms2301.

Merow, Cory, Matthew J. Smith, and John A. Silander. 2013. “A Practical Guide to MaxEnt for Modeling Species’ Distributions: What It Does, and Why Inputs and Settings Matter.” *Ecography* 36 (10): 1058–69. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Michener, William K., James W. Brunt, John J. Helly, Thomas B. Kirchner, and Susan G. Stafford. 1997. “Nongeospatial Metadata for the Ecological Sciences.” *Ecological Applications* 7 (1): 330–42. https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2.

Miller, Ryan S., Matthew L. Farnsworth, and Jennifer L. Malmberg. 2013. “Diseases at the Livestock–wildlife Interface: Status, Challenges, and Opportunities in the United States.” *Preventive Veterinary Medicine* 110 (2): 119–32. https://doi.org/10.1016/j.prevetmed.2012.11.021.

Morales, Narkis S., Ignacio C. Fernández, and Victoria Baca-González. 2017. “MaxEnt’s Parameter Configuration and Small Samples: Are We Paying Attention to Recommendations? A Systematic Review.” *PeerJ* 5 (March). https://doi.org/10.7717/peerj.3093.

Mörner, T., D. L. Obendorf, M. Artois, and M. H. Woodford. 2002a. “Surveillance and Monitoring of Wildlife Diseases.” *Revue Scientifique Et Technique (International Office of Epizootics)* 21 (1): 67–76.

———. 2002b. “Surveillance and Monitoring of Wildlife Diseases.” *Revue Scientifique Et Technique (International Office of Epizootics)* 21 (1): 67–76.

Mostashari, Farzad, Martin Kulldorff, Jessica J. Hartman, James R. Miller, and Varuni Kulasekera. 2003. “Dead Bird Clusters as an Early Warning System for West Nile Virus Activity.” *Emerging Infectious Diseases* 9 (6): 641–46. https://doi.org/10.3201/eid0906.020794.

Nusser, Sarah M., William R. Clark, David L. Otis, and Ling Huang. 2008. “Sampling Considerations for Disease Surveillance in Wildlife Populations.” *Journal of Wildlife Management* 72 (1): 52–60. https://doi.org/10.2193/2007-317.

“Ontario’s White-Nose Syndrome Response Plan.” 2015. Ontario.ca. June 2, 2015. https://www.ontario.ca/document/ontarios-white-nose-syndrome-response-plan.

Openshaw, Stan. 1990. “Spatial Analysis and Geographical Information Systems: A Review of Progress and Possibilities.” In *Geographical Information Systems for Urban and Regional Planning*, 153–63. The GeoJournal Library. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1677-2_14.

Orr, Ken. 1998. “Data Quality and Systems Theory.” *Communications of the ACM* 41 (2): 66–71. https://doi.org/10.1145/269012.269023.

Ostfeld, Richard S., Gregory E. Glass, and Felicia Keesing. 2005. “Spatial Epidemiology: An Emerging (or Re-Emerging) Discipline.” *Trends in Ecology & Evolution* 20 (6): 328–36. https://doi.org/10.1016/j.tree.2005.03.009.

Peterson, A. Townsend. 2006. “Ecologic Niche Modeling and Spatial Patterns of Disease Transmission.” *Emerging Infectious Diseases* 12 (12): 1822–26. https://doi.org/10.3201/eid1212.060373.

Pfeiffer, D. U., and M. Hugh-Jones. 2002. “Geographical Information Systems as a Tool in Epidemiological Assessment and Wildlife Disease Management.” *Revue Scientifique Et Technique (International Office of Epizootics)* 21 (1): 91–102.

Phillips, Steven J, Robert P Anderson, and Robert E Schapire. 2006. “Maximum Entropy Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–259.

Phillips, Steven J., and Miroslav Dudík. 2008. “Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation.” *Ecography* 31 (2): 161–75. https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Phillips, Steven J., Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. 2009. “Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data.” *Ecological Applications* 19 (1): 181–97. https://doi.org/10.1890/07-2153.1.

Phillips, Steven J., Miroslav Dudík, and Robert E. Schapire. 2004. “A Maximum Entropy Approach to Species Distribution Modeling.” In *Proceedings of the Twenty-First International Conference on Machine Learning*, 83–. ICML ’04. New York, NY, USA: ACM. https://doi.org/10.1145/1015330.1015412.

Pierkot, Christelle, Esteban Zimányi, Yuan Lin, and Thérèse Libourel. 2011. “Advocacy for External Quality in GIS.” In *GeoSpatial Semantics*, 151–65. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20630-6_10.

Pôças, Isabel, João Gonçalves, Bruno Marcos, Joaquim Alonso, Pedro Castro, and João P. Honrado. 2014. “Evaluating the Fitness for Use of Spatial Data Sets to Promote Quality in Ecological Assessment and Monitoring.” *International Journal of Geographical Information Science* 28 (11): 2356–71. https://doi.org/10.1080/13658816.2014.924627.

Radosavljevic, Aleksandar, and Robert P. Anderson. 2014. “Making Better Maxent Models of Species Distributions: Complexity, Overfitting and Evaluation.” *Journal of Biogeography* 41 (4): 629–43. https://doi.org/10.1111/jbi.12227.

Randall, Jennifer, and Hugh G. Broders. 2014. “Identification and Characterization of Swarming Sites Used by Bats in Nova Scotia, Canada.” *Acta Chiropterologica* 16 (1): 109–16. https://doi.org/10.3161/150811014X683327.

Reddy, Sushma, and Liliana M. Dávalos. 2003. “Geographical Sampling Bias and Its Implications for Conservation Priorities in Africa.” *Journal of Biogeography* 30 (11): 1719–27. https://doi.org/10.1046/j.1365-2699.2003.00946.x.

Rees, Erin E. 2011. “Targeted Surveillance of Raccoon Rabies in Quéébec, Canada” 75 (6): 1406–16.

Rees, Erin E., Denise Bélanger, Frédérick Lelièvre, Nathalie Coté, and Louise Lambert. 2011. “Targeted Surveillance of Raccoon Rabies in Québec, Canada.” *The Journal of Wildlife Management* 75 (6): 1406–16. https://doi.org/10.1002/jwmg.178.

Robertson, Colin, Jed A. Long, Farouk S. Nathoo, Trisalyn A. Nelson, and Cameron C. F. Plouffe. 2014. “Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks.” *Geographical Analysis* 46 (1): 53–74. https://doi.org/10.1111/gean.12028.

Robertson, Colin, Lauren Yee, Julia Metelka, and Craig Stephen. 2016. “Spatial Data Issues in Geographical Zoonoses Research.” *The Canadian Geographer / Le Géographe Canadien* 60 (3): 300–319. https://doi.org/10.1111/cag.12280.

Ryser-Degiorgis, Marie-Pierre. 2013. “Wildlife Health Investigations: Needs, Challenges and Recommendations.” *BMC Veterinary Research* 9 (1): 223.

Rytkönen, Mika J. 2004. “Not All Maps Are Equal: GIS and Spatial Analysis in Epidemiology.” *International Journal of Circumpolar Health* 63 (1): 9–24.

Sala, Osvaldo E., F. Stuart Chapin, Iii, Juan J. Armesto, Eric Berlow, Janine Bloomfield, Rodolfo Dirzo, et al. 2000. “Global Biodiversity Scenarios for the Year 2100.” *Science* 287 (5459): 1770–74. https://doi.org/10.1126/science.287.5459.1770.

Schoener, Thomas W. 1970. “Nonsynchronous Spatial Overlap of Lizards in Patchy Habitats.” *Ecology* 51 (3): 408–18. https://doi.org/10.2307/1935376.

Schurer, Janna, Todd Shury, Frederick Leighton, and Emily Jenkins. 2013. “Surveillance for Echinococcus Canadensis Genotypes in Canadian Ungulates.” *International Journal for Parasitology: Parasites and Wildlife* 2 (Supplement C): 97–101. https://doi.org/10.1016/j.ijppaw.2013.02.004.

Skerratt, Lee Francis, Lee Berger, Richard Speare, Scott Cashins, Keith Raymond McDonald, Andrea Dawn Phillott, Harry Bryan Hines, and Nicole Kenyon. 2007. “Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs.” *Ecohealth* 4 (2): 125–34. https://doi.org/10.1007/s10393-007-0093-5.

Soper, Karen D., and M. Brock Fenton. 2007. “Availability of Building Roosts for Bats in Four Towns in Southwestern Ontario, Canada.” *Acta Chiropterologica* 9 (2): 542–46. https://doi.org/10.3161/1733-5329(2007)9[542:AOBRFB]2.0.CO;2.

Sosin, Daniel M. 2003. “Draft Framework for Evaluating Syndromic Surveillance Systems.” *Journal of Urban Health : Bulletin of the New York Academy of Medicine* 80 (Suppl 1): i8–13. https://doi.org/10.1007/PL00022309.

Stallknecht, D. E. 2007. “Impediments to Wildlife Disease Surveillance, Research, and Diagnostics.” In *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances* *and Consequences of Cross-Species Transmission*, edited by Senior Research Scientist James E. Childs, Professor John S. Mackenzie, and Veterinary Medical Officer Jürgen A. Richt, 445–61. Current Topics in Microbiology and Immunology 315. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70962-6_17.

Stärk, Katharina DC, Gertraud Regula, Jorge Hernandez, Lea Knopf, Klemens Fuchs, Roger S. Morris, and Peter Davies. 2006. "Concepts for Risk-Based Surveillance in the Field of Veterinary Medicine and Veterinary Public Health: Review of Current Approaches." *BMC Health Services Research* 6 (February): 20. https://doi.org/10.1186/1472-6963-6-20.

Stephen, Craig. 2014. "Toward a Modernized Definition of Wildlife Health." *Journal of Wildlife Diseases* 50 (3): 427–30. https://doi.org/10.7589/2013-11-305.

———. 2015. "The Canadian Wildlife Health Cooperative: Addressing Wildlife Health Challenges in the 21st Century." *The Canadian Veterinary Journal* 56 (9): 925.

Stephen, Craig, Harvey Artsob, William R Bowie, Michael Drebot, Erin Fraser, Ted Leighton, Muhammad Morshed, Corinne Ong, and David Patrick. 2004. "Perspectives on Emerging Zoonotic Disease Research and Capacity Building in Canada." *The Canadian Journal of Infectious Diseases & Medical Microbiology* 15 (6): 339–44.

Stitt, Tyler, Julie Mountifield, and Craig Stephen. 2007. "Opportunities and Obstacles to Collecting Wildlife Disease Data for Public Health Purposes: Results of a Pilot Study on Vancouver Island, British Columbia." *The Canadian Veterinary Journal* 48 (1): 83–90.

Stolar, Jessica, and Scott E. Nielsen. 2015. "Accounting for Spatially Biased Sampling Effort in Presence-Only Species Distribution Modelling." *Diversity and Distributions* 21 (5): 595–608. https://doi.org/10.1111/ddi.12279.

Syfert, Mindy M., Matthew J. Smith, and David A. Coomes. 2013. "The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models." *PLOS ONE* 8 (2): e55158. https://doi.org/10.1371/journal.pone.0055158.

Theobald, David M., James R. Miller, and N. Thompson Hobbs. 1997. "Estimating the Cumulative Effects of Development on Wildlife Habitat." *Landscape and Urban Planning* 39 (1): 25–36. https://doi.org/10.1016/S0169-2046(97)00041-8.

Therneau, Terry, Beth Atkinson, and Brian Ripley. 2015. *Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1–10*.

Thomas, Donald W., M. Brock Fenton, and Robert M. R. Barclay. 1979. "Social Behavior of the Little Brown Bat, Myotis Lucifugus: I. Mating Behavior." *Behavioral Ecology and Sociobiology* 6 (2): 129–36.

Unwin, David J. 2005. "Fiddling on a Different Planet?" *Geoforum* 36 (6): 681–84. https://doi.org/10.1016/j.geoforum.2005.04.003.

U.S. Fish & Wildlife Service. 2017. "White-Nose Syndrome: The Devastating Disease of Hibernating Bats in North America." https://www.whitenosesyndrome.org/sites/default/files/resource/white-nose_fact_sheet_10-2017_1.pdf.

Van Oort, P. a. J., and A. K. Bregt. 2005. "Do Users Ignore Spatial Data Quality? A Decision-Theoretic Perspective." *Risk Analysis* 25 (6): 1599–1610. https://doi.org/10.1111/j.1539-6924.2005.00678.x.

VanDerWal, Jeremy, Luke P. Shoo, Catherine Graham, and Stephen E. Williams. 2009. "Selecting Pseudo-Absence Data for Presence-Only Distribution Modeling: How Far

Should You Stray from What You Know?” *Ecological Modelling* 220 (4): 589–94. https://doi.org/10.1016/j.ecolmodel.2008.11.010.

Varela, Sara, Robert P. Anderson, Raúl García-Valdés, and Federico Fernández-González. 2014. “Environmental Filters Reduce the Effects of Sampling Bias and Improve Predictions of Ecological Niche Models.” *Ecography* 37 (11): 1084–91. https://doi.org/10.1111/j.1600-0587.2013.00441.x.

Veloz, Samuel D. 2009. “Spatially Autocorrelated Sampling Falsely Inflates Measures of Accuracy for Presence-Only Niche Models.” *Journal of Biogeography* 36 (12): 2290–99. https://doi.org/10.1111/j.1365-2699.2009.02174.x.

Verant, Michelle L., Justin G. Boyles, William Waldrep, Gudrun Wibbelt, and David S. Blehert. 2012. “Temperature-Dependent Growth of Geomyces Destructans, the Fungus That Causes Bat White-Nose Syndrome.” Edited by Matthew Charles Fisher. *PLoS ONE* 7 (9): e46280. https://doi.org/10.1371/journal.pone.0046280.

Veregin, Howard. 1999. “Data Quality Parameters.” *Geographical Information Systems* 1: 177–189.

Voyles, Jamie, A. Marm Kilpatrick, James P. Collins, Matthew C. Fisher, Winifred F. Frick, Hamish McCallum, Craig K. R. Willis, et al. 2014. “Moving Beyond Too Little, Too Late: Managing Emerging Infectious Diseases in Wild Populations Requires International Policy and Partnerships.” *EcoHealth* 12 (3): 404–7. https://doi.org/10.1007/s10393-014-0980-5.

Vrbova, L., C. Stephen, N. Kasman, R. Boehnke, M. Doyle-Waters, A. Chablitt-Clark, B. Gibson, M. FitzGerald, and D. M. Patrick. 2010. “Systematic Review of Surveillance Systems for Emerging Zoonoses.” *Transboundary and Emerging Diseases* 57 (3): 154–61. https://doi.org/10.1111/j.1865-1682.2010.01100.x.

Walter, SD. 2000. “Disease Mapping: A Historical Perspective.” In *Spatial Epidemiology: Methods and Applications,* 223–52. Oxford: Oxford University Press. http://www.oxfordscholarship.com.libproxy.wlu.ca/view/10.1093/acprof:oso/9780198515326.001.0001/acprof-9780198515326-chapter-12.

Wang, Richard Y., and Diane M. Strong. 1996. “Beyond Accuracy: What Data Quality Means to Data Consumers.” *Journal of Management Information Systems* 12 (4): 5–33. https://doi.org/10.1080/07421222.1996.11518099.

Wang, Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. “Image Quality Assessment: From Error Visibility to Structural Similarity.” *IEEE Transactions on Image Processing* 13 (4): 600–612. https://doi.org/10.1109/TIP.2003.819861.

Warnecke, Lisa, James M. Turner, Trent K. Bollinger, Jeffrey M. Lorch, Vikram Misra, Paul M. Cryan, Gudrun Wibbelt, David S. Blehert, and Craig K. R. Willis. 2012. “Inoculation of Bats with European Geomyces Destructans Supports the Novel Pathogen Hypothesis for the Origin of White-Nose Syndrome.” *Proceedings of the National Academy of Sciences* 109 (18): 6999–7003. https://doi.org/10.1073/pnas.1200374109.

Warns-Petit, Eva, Eric Morignat, Marc Artois, and Didier Calavas. 2010. “Unsupervised Clustering of Wildlife Necropsy Data for Syndromic Surveillance.” *BMC Veterinary Research* 6 (1): 56. https://doi.org/10.1186/1746-6148-6-56.

Warren, Dan L., Richard E. Glor, and Michael Turelli. 2008. “ENVIRONMENTAL NICHE EQUIVALENCY VERSUS CONSERVATISM: QUANTITATIVE APPROACHES TO NICHE EVOLUTION.” *Evolution* 62 (11): 2868–83. https://doi.org/10.1111/j.1558-5646.2008.00482.x.

Warren, Dan L., and Stephanie N. Seifert. 2011. “Ecological Niche Modeling in Maxent: The Importance of Model Complexity and the Performance of Model Selection Criteria.” *Ecological Applications* 21 (2): 335–42. https://doi.org/10.1890/10-1171.1.

WHO. 2006. “Interregional Meeting on Preventing and Controlling Plague: The Old Calamity Still Has a Future.” http://www.who.int/csr/resources/publications/WHO_HSE_EPR_2008_3w.pdf.

Wibbelt, Gudrun, Andreas Kurth, David Hellmann, Manfred Weishaar, Alex Barlow, Michael Veith, Julia Prüger, et al. 2010. “White-Nose Syndrome Fungus (Geomyces Destructans) in Bats, Europe.” *Emerging Infectious Diseases* 16 (8): 1237–43. https://doi.org/10.3201/eid1608.100002.

Wieczorek, John, Qinghua Guo, and Robert Hijmans. 2004. “The Point-Radius Method for Georeferencing Locality Descriptions and Calculating Associated Uncertainty.” *International Journal of Geographical Information Science* 18 (8): 745–67. https://doi.org/10.1080/13658810412331280211.

Wobeser, G. A. 1994. “Investigation and Management of Disease in Wild Animals.,” vi + 265 pp.

Yackulic, Charles B., Richard Chandler, Elise F. Zipkin, J. Andrew Royle, James D. Nichols, Evan H. Campbell Grant, and Sophie Veran. 2013. “Presence-Only Modelling Using MAXENT: When Can We Trust the Inferences?” *Methods in Ecology and Evolution* 4 (3): 236–43. https://doi.org/10.1111/2041-210x.12004.