2013-08-01

Probability distribution maps for the arbovirus vector Culicoides sonorensis in southern Alberta and Montana under current and future climate scenarios.

Zuliani, Anna

UNIVERSITY OF CALGARY

Probability distribution maps for the arbovirus vector *Culicoides sonorensis* in southern Alberta and Montana under current and future climate scenarios.

by

Anna Zuliani

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

FACULTY OF VETERINARY MEDICINE

CALGARY, ALBERTA

July 2013

ABSTRACT

Culicoides sonorensis transmits Bluetongue and Epizootic Hemorrhagic Disease viruses to ruminants in western North America. This study was conducted to model the current and future distribution of *C. sonorensis* in Southern Alberta (Canada) and Montana (USA). Data on *C. sonorensis*, collected from 2002 to 2011, and environmental and climatic variables, chosen based on the ecology of *Culicoides* spp., were used in the analysis. Under current climatic conditions, the best MaxEnt model selected with the Akaike Information Criterion was built using 50 *C. sonorensis* presence records and four variables (Elevation, Land Cover, mean Precipitation of May and mean Vapour Pressure Deficit of July). Under future climate scenarios, constructed relying on three Representative Concentration Pathways, a northward expansion of *C. sonorensis* was predicted by the 2030s and 2050s. Our models can provide information for the development of future vector surveillance and the assessment of disease transmission risk in the region.

ACKNOWLEDGEMENTS

I would like to thank all of the people who contributed in some way to the work described in this thesis. First, I would like to express my gratitude to my supervisor, Dr. Susan Cork, for her support and guidance. She has faithfully encouraged me to follow my interests and to shape my ideas throughout this Master's project. I would like to thank my co-supervisor, Dr. Alessandro Massolo, for his patience during our long sessions on statistics and modeling techniques. Very special thanks go to my committee members, Dr. Tim Lysyk and Dr. Greg Johnson, without whom this project would have not been possible. Their effort to collect and identify *Culicoides* provided me with a unique entomological database. I hope I have been able to develop and show most of its potential with this work. I would like to emphasize my deep gratitude to Dr. Tim Lysyk whose guidance, advice and insights in research and entomology were invaluable to me. I also wish to thank Dr. Dorothy Geale for her invaluable help on Bluetongue policy and Dr Shawn Marshall, who patiently introduced me to climatology and provided me with kind guidance with climate change scenarios. Many thanks go to former and current members of the Vector Ecology group Dr. Regula Waeckerlin, Allison Cully and Dr. Kathryn Berger. They greatly contributed to my research and cheered me up when the learning curve was too steep. I would also like to acknowledge the Faculty of Veterinary Medicine and the University of Calgary for their economic support and the administrative assistants for their technical help.

Finally, I would like to thank family and friends. They are too many to be named but each of them is truly making a difference in my life. The first law of geography states that everything is related to everything else, but near things are more related than distant things. I am grateful to say that this law was not true for my family and friends in these past years. I promise it will not

become true even in the years to come for the people I lived, worked, sweated, complained, laughed and dreamt with during these years as a Master student.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

LIST OF TERMS

Area under the receiver-operating characteristic (ROC) curve (AUC)	Metric that combines sensitivity and specificity to assess model accuracy
Deviance	Measure of goodness of fit equals to negative 2 times the value of the log-likelihood at its maximum point, $-2 \log L$
Entropy	Measure of uncertainty
Feature	Transformation of original predictors
Information gain in MaxEnt	Measure of goodness of fit, closely related to deviance
Log-likelihood	The natural logarithm of the probability of an observed outcome, $\log L$
Model accuracy	Closeness of the model to the reality that aims to describe or predict
Percentage contribution in MaxEnt	The increased training gain obtained in each iteration of the model is added to the final contribution of the variable to the model
Permutation importance in MaxEnt	The drop of AUC value when the final model is compared to one in which values of the considered variable at presence locations and background points are permutated
Sensitivity	Ability to identify positive cases correctly
Specificity	Ability to identify negative cases correctly

Variance	Average squared deviation of each case from the mean

LIST OF ABBREVIATIONS

A	Aspect
AB -1, 2, 3, 4	Entomological dataset 1, 2, 3, 4 from Alberta
AIC	Akaike Information Criterion
AICc	Akaike Information Criterion with correction for small sample sizes
AUC	Area Under the Curve
aVP	Actual Vapour Pressure
BSE	Bovine Spongiform Encephalopathy
BT	Bluetongue
BTV	Bluetongue Virus
CCCma	Canadian Centre for Climate Modeling and Analysis
CFIA	Canadian Food Inspection Agency
E	Elevation
EFSA	European Food Safety Authority
EHD	Epizootic Hemorrhagic Disease
EHDV	Epizootic Hemorrhagic Disease Virus
GCM	General Circulation Model
GHG	Greenhouse Gas
GLM	General Linear Model
GPS	Global Positioning System
H	Entropy
IGBP	International Geosphere-Biosphere Program

IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
LC	Land Cover
LP DAAC	Land Processes Distributed Active Archive Center
LST	Land Surface Temperature
MIR	Middle Infrared Radiation
MODIS	MODerate Resolution Imaging Spectroradiometer
MT -1, 2, 3, 4	Entomological dataset 1, 2, 3, 4 from Montana
NARR	North America Regional Reanalysis
NCEP	National Centers for Environmental Predictions
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
NOAA	National Oceanic and Atmospheric Administration
OIE	World Organization for Animal Health
P	Precipitation
R	Red
RCP	Representative Concentration Pathway
RH	Relative Humidity
ROC	Receiver-Operating Characteristic
SBD	Schmallenberg Disease
SBV	Schmallenberg Virus
SDM	Species Distribution Model
sVP	Saturated Vapour Pressure
T	Temperature
VPD	Vapour Pressure Deficit
VS	Vesicular Stomatitis

The highest that we can attain to is not Knowledge, but Sympathy with Intelligence.

Henry David Thoreau

INTRODUCTION

Vector borne diseases are complex epidemiological systems in which the interaction of pathogens, vectors, hosts and the environment play a fundamental role in the occurrence and potential spread of the disease (Tabachnick, 2010). Over the past decades, several vector-borne diseases emerged in new geographical areas or affected different host populations. Re-emergence was characterized by disease spread beyond previously known geographical boundaries or by disease occurrence with higher than expected incidence rates (Kilpatrick & Randolph, 2012).

Many pathogens transmitted by arthropod vectors are known to cause severe diseases and economic losses (Magori & Drake, 2013). Due to their increased importance over the last decades, the World Organization for Animal Health, OIE, produced specific guidelines for the surveillance of arthropod vectors (OIE, 2012) in addition to more general disease surveillance requirements (OIE, 2012). The OIE develops international sanitary standards to ensure safe animal trade on the basis of veterinary scientific information collected worldwide.

In the context of the Terrestrial Animal Health Code, risk is defined as the probability of disease occurrence associated with its economic or health consequences. Risk analysis is a process that involves four main steps: hazard identification, risk assessment, risk management and risk communication. The risk assessment framework is used to collect existing scientific information and to identify potential information gaps in order to address disease risk and to improve disease preparedness plans (de Vos, Hoek, Fisher, de Koeijer, & Bremmer, 2010). Within the risk assessment process four steps are identified: entry assessment, exposure assessment, consequence assessment and risk estimation. Exposure assessment is the step in which the probability of disease transmission and spread is evaluated (OIE, 2012). In the case of vector-borne diseases, this step highly depends on the presence and capacity of a competent

arthropod population to transmit a pathogen under certain environmental and climatic conditions. Vectorial capacity estimates the ability of an arthropod population to transmit a pathogen to a vertebrate host population and vector competence estimates the fraction of arthropods physiological capable of sustaining replication and efficiently transmitting the pathogen (Reisen, 1989). Both factors are affected by environmental and climatic conditions in which arthropod vectors live. In order to properly estimate disease transmission risk, a sound knowledge of vector ecology and vector distribution is required. Understanding the relationship between the vector and the environment is an essential pre-requisite for producing a biologically sound risk assessment for vector borne-diseases. This becomes especially important when estimating changing risk profiles in the context of climate change.

This thesis focuses on the development of probability distribution maps for *Culicoides sonorensis*, in southern Alberta and Montana under current and future climatic scenarios. *Culicoides sonorensis* is the main vector of Bluetongue and Epizootic Haemorrhagic Disease in Western North America. It is envisaged that the result of this work will provide information to the risk assessment process as well as help in the development of preparedness plans for emerging *Culicoides*-borne diseases in the region.

1.1 *CULICOIDES* VECTORS

Culicoides biting midges (Diptera: Ceratopogonidae) are small haematophagous insects that occur throughout the world, with the exception of Antarctica and New Zealand (Mellor, Boorman, & Baylis, 2000). Presently 151 species of *Culicoides* belonging to 13 subgenera and six species groups (Grogan & Phillips, 2008) have been described in North America, north of Mexico. Of these, *C. sonorensis* is considered to be the primary vector of Bluetongue virus

(BTV) and Epizootic Hemorrhagic Disease Virus (EHDV) on the basis of its known range, virus isolations and vector competence studies (Tabachnick, 1992).

1.1.1 Taxonomy of *C. sonorensis*

Culicoides sonorensis Wirth and Jones is a member of the family Ceratopogonidae, genus *Culicoides* Latreille, subgenus Monoculicoides Khalaf that also includes *C. gigas* Root and Hoffman, *C. grandensis* Grogan and Phillips, *C. occidentalis* Wirth and Jones and *C. variipennis* in North America.

Culicoides sonorensis was initially considered one of the five subspecies in the *C. variipennis* complex that also included *C. v. variipennis*, *C. v. albertensis*, *C. v. australis* and *C. v. occidentalis* (Wirth & Jones, 1957). Atchley (1967) suggested that *C. v. australis* was a synonym of *C. v. sonorensis* and Jorgensen (1969) proposed to elevate C. *v. variipennis* and *C. v. occidentalis* to species status. Downes (1978) suggested that *C. variipennis* and *C. occidentalis* be considered species, but that *C. v. sonorensis*, *C. v. albertensis* and *C. v. australis* were subspecies of *C. occidentalis.* Wirth and Morris (1985) advised the use of the term "*Culicoides variipennis* complex" and suggested that *C. v. sonorensis* and *C. v. occidentalis* be considered subspecies of *C. variipennis*. Tabachnick (Tabachnick, 1992) used electrophoretic techniques to define the genetic independence of the variipennis complex members (*C. v. variipennis, C. v. sonorensis and C. v. occidentalis*). The three subspecies were eventually elevated to a species status through further studies whereas *C. australis* and *C. albertensis* were confirmed as synonyms of *C. sonorensis* (Holbrook, Schmidtmann, McKinnon, Bobian, & Grogan, 2000).

1.1.2 *Culicoides sonorensis* life cycle

Adult *C. sonorenis* are grey and 1.5-2.5 mm long. The two wings possess dense hairs that give rise to pigmentation patterns (Figure 1.1). *Culicoides sonorensis* selects breeding sites close to water bodies (e.g. manure lagoons, irrigation lakes and cattle watering sites) with muddy and shallow margins. (McMullen, 1978; Mullens, 1989). Soils surrounding developmental sites for *C. sonorensis* tend to have high saline and alkaline contents (Schmidtmann, Bobian, & Belden, 2000; McMullen, 1978). *Culicoides sonorensis* swarms have been observed over muddy areas, low grass and bushes (Holbrook, Schmidtmann, McKinnon, Bobian, & Grogan, 2000), which might be considered a favourable habitat for the species.

Male adults generally emerge from pupae before the females and tend to form swarms above the breeding sites waiting for the emergence of females to mate (Braverman 1994). Emerging adults usually fly only a few hundred meters from their larval habitats, but a flight range of 4 km has been reported for *C. variipennis* (Lillie, Marquart, & Jones, 1981). However, because of their small size, *Culicoides* can also be carried by wind streams for hundreds of kilometres (Pedgley, 1983). The flying and biting activity of most *C. sonorensis* females is crepuscular and/or nocturnal (Mellor, Boorman, & Baylis, 2000).

Culicoides sonorensis females feed on blood from various species of mammals while males are primarily nectar feeders. Blood meals are required to produce eggs. The oogenesis process (i.e. egg development and oviposition) requires two days at 30˚C and 10 days at 13˚C (Mullens & Holbrook, 1991). *Culicoides sonorensis* can lay from 9 to 254 eggs in laboratory conditions depending on the size of the blood-meal, season and number of ovipositions (Lysyk & Danyk, 2007). Eggs are laid in wet soil or any sort of semi-aquatic habitat (Mullens & Lii, 1987) and hatch after two to nine days in summer temperate conditions (Braverman, 1994). Eggs hatch and

larvae pass through four larval instars. The fully-grown larva is cylindrical, whitish in color and approximately 2-6 mm long. Larvae move in a serpentine fashion and feed on a wide range of micro-organisms and decaying organic material (Mullen, 2009). Third and fourth larval stages serve as overwintering generations in temperate regions (Barnard & Jones, 1980; Lysyk, 2007). Pupation occurs at the end of the fourth larval instar and lasts two to ten days. The adult median longevity of individuals in laboratory conditions varies from 12-19 days at 10˚C to 9-10 days at 30˚C (Lysyk & Danyk, 2007).

Although adult *C. sonorensis* have been collected at temperatures as low as 5°C (Gerry & Mullens, 2000), population peaks usually occur when temperatures exceed 16°C (Lysyk, 2007). In southern Alberta (Lysyk, 2006), the arthropod is present from May to September, with higher abundance in mid-July and late August, and passes through three generations separated from each other by five-six weeks (Lysyk, 2007). Under warmer conditions (> 16°C), a negative correlation is observed between adult survival and temperature (Gerry and Mullens 2000; Wittmann, Mellor and Baylis 2002; Lysyk and Danyk, 2007). Low temperatures (15ºC) and high relative humidity (85%) seem to increase survival compared to the same temperatures and at lower relative humidity rates (40% and 75%). However, at 30ºC, the effect of relative humidity on the insect longevity is reversed (Wittmann, Mellor, & Baylis, 2002). The reason of this finding still is unclear but the authors hypothesized that *C. sonorensis* might be facilitated in eliminating metabolic water in conditions of high temperature and low relative humidities compared to conditions of high temperature and high relative humidity.

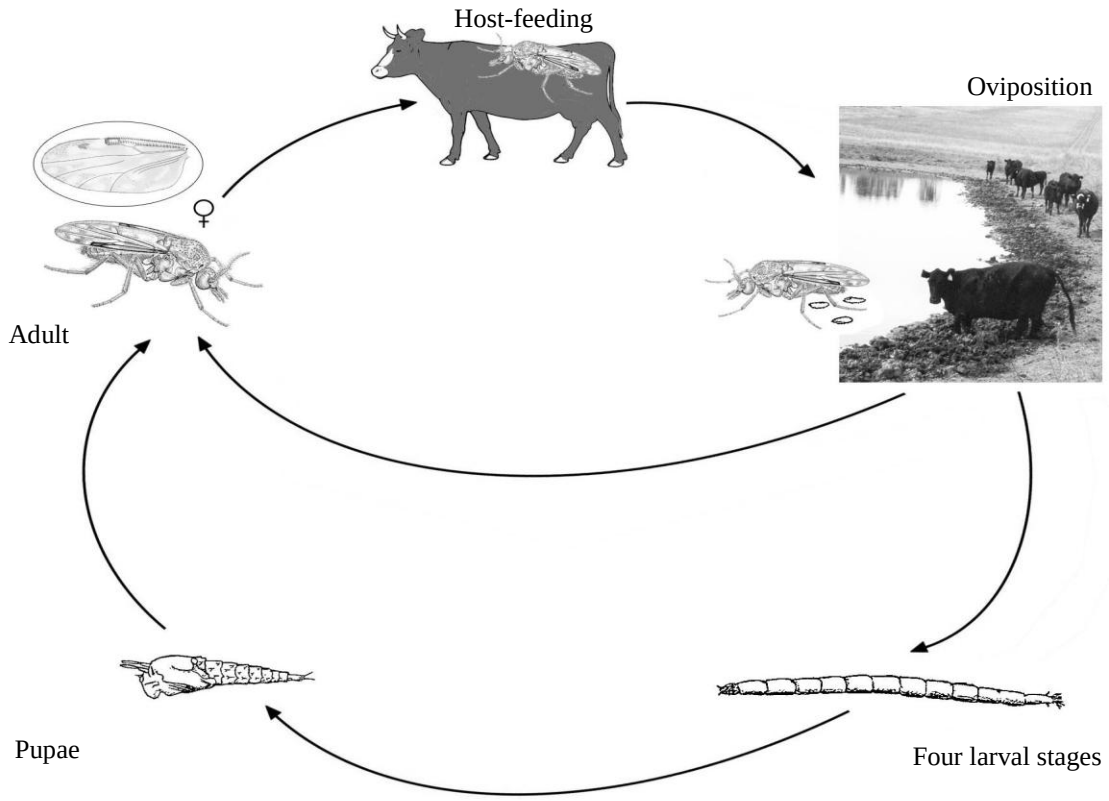


Figure 1.1 *Culicoides sonorensis* life-cycle.

1.1.3 Factors affecting *Culicoides* distribution

Factors influencing the distribution of *C. sonorensis* have not been fully evaluated at a regional scale. However, biotic and abiotic factors that determine Palearctic *Culicoides* spp. occurrence have been studied following BTV incursions into the European continent (Baylis, O`Connell, & Purse, 2004).

Variables (Table 1.1) selected to build distribution maps for Palearctic BTV vectors include topographic variables (Baylis, Mellor, et al. 2001, Wittmann, Mellor and Baylis 2001, Tatem, et al. 2003, Purse, et al. 2004, Acevedo, et al. 2010), temperature (Baylis, Mellor, et al. 2001,Wittmann, Mellor and Baylis 2001, Tatem, et al. 2003, Purse, et al. 2004, Calvete, et al. 2008, Acevedo, et al. 2010), precipitation (Wittmann, Mellor, & Baylis, 2001; Calvete, Estrada, Miranda, Borras, Calvo, & Lucientes, 2008; Acevedo, et al., 2010), vapour pressure deficit

(Baylis, Mellor, Wittmann, & Rogers, 2001; Wittmann, Mellor, & Baylis, 2001), vegetation indices (Baylis, Mellor, et al. 2001, Tatem, et al. 2003, Purse, et al. 2004, Calvete, et al. 2008, Acevedo, et al. 2010), land cover and host density (Acevedo, et al. 2010). The best predictors differ among models (shown in bold in Table 1.1) probably because of the variety of *Culicoides* species and regions that were considered in the studies.

Two studies explored the effects of climate change on *C. imicola* distribution in Mediterranean countries (Wittmann, Mellor, & Baylis, 2001; Acevedo, et al., 2010) considering respectively temperature, and temperature and precipitation as predictors. Wittmann et al. (2001) predicted an extensive range expansion across the Mediterranean Basin while Acevedo et al. (2010) expected an increased abundance with a minimal range expansion. The latter finding is supported by field studies (Capela, et al., 2003; Conte, Gilbert, & Goffredo, 2009) in which no *C. imicola* range expansion has been observed, suggesting that temperature alone is not a good predictor in forecasting *Culicoides* distributions.

Table 1.1 Variables used to model *Culicoides* spp. distribution in Europe. The best predictors are shown in bold.

Variables			**Baylis et al., 2001**	**Wittmann et al., 2001**	**Tatem et al., 2003**	**Purse et al., 2004**	**Calvete et al., 2008**	**Acevedo et al., 2010**
Topography	Elevation		X	X	X	X		X
	Slope							X
Temperature	Land Surface Temp.	Mean, phase and amplitude annual/biannual/ triannual	X		X	**X**		
		Mean/max/min/variance annual+biannual +triannual	X		X	X		
		Variance original value			X	**X**		
	Monthly mean/max/min			**X**				X
	Seasonality (variation coefficient of monthly means)						**X**	X
	Annual mean						X	
Precipitation	Annual total			X			X	
	Seasonality(variation coefficient of monthly means)						**X**	**X**
	Seasonal total							**X**
Vapour Pressure Deficit	Annual daily mean			X				
	Mean, phase and amplitude annual/biannual/triannual		X					
	Mean/max/min/variance annual+biannual+triannual		X					
Vegetation Indices	Normalized Difference Vegetation Index	Monthly Mean					X	X
		Seasonality					X	X
		Mean, phase and amplitude Annual/biannual/ triannual	**X**		X	X		
		Mean/max/min/variance annual+biannual +triannual			X	X		
		Variance original value			X	X		
	Middle Infrared Radiation	Mean, phase and amplitude Annual/biannual/ triannual	X		**X**	**X**		
		Mean/max/min/variance annual+biannual +triannual	X		X	X		
		Variance original value			X	X		
Land Cover								**X**
Host density								X

1.2 PATHOGENS TRANSMITTED BY *CULICOIDES* SPP.

Culicoides spp. can transmit several viruses pathogenic to wild and domestic ungulates (Mellor, Boorman, & Baylis, 2000). The two most important pathogens in North America are the viruses that cause Bluetongue Disease (BT) and Epizootic Hemorrhagic Disease (EHD). These diseases can result in significant morbidity and mortality with economic loss for livestock producers, deer farmers, hunting groups and the wildlife recreation sectors.

1.2.1 Bluetongue (BT)

1.2.1.1 BT virus and disease

Bluetongue virus (BTV) is a double stranded RNA virus that belongs to the genus *Orbivirus* of the family *Reoviridae* (Attoui, Maan, Anthony, & Mertens, 2009). BTV is characterised by the existence of 26 serotypes (Maan, Maan, Belaganahalli, Johnson, Nomikou, & Mertens, 2012). Fifeteen serotypes have been detected in North America, although only five of them (2, 10, 11, 13, 17) are considered endemic to the continent (MacLachlan & Guthrie, 2010).

BT disease affects ruminants causing haemorrhaging and ulceration in the upper gastrointestinal tract as well as laminitis, coronitis, facial and neck oedema, pulmonary oedema, reproductive failure and lameness (Mellor, Baylis, & Mertens, 2009). Mortality as high as 50% and morbidity as high as 70% have been observed in immunologically naïve sheep populations. Cattle typically show mild clinical signs of infection, depending on serotypes (Hourrigan & Klingsporn, 1975). Wild cervids can seroconvert to BTV (Stallknecht, Blue, Rollor, Nettles, Davidson, & Pearson, 1991) but clinical signs have been observed only after experimentally induced infection (Howerth, Greene, & Prestwood, 1988).

1.2.1.2 BTV transmission

Few *Culicoides* species act as competent BTV vectors and those differ from continent to continent. In Africa, Middle East and southern Europe *C. imicola* is the major vector of BTV (Mellor, Boorman, & Baylis, 2000). In northern and eastern Europe the competent BTV *Culicoides* species belong to the *C. obsoletus* and *C. pulicaris* complexes (Purse, Brown, Harrup, Mertens, & Rogers, 2008). In north-western USA and western Canada *C. sonorensis* is the primary BTV vector (Tabachnick, 1992). *Culicoides insignis* is the predominant vector of BTV in South and Central America but it has also been implicated in virus transmission in south-eastern USA (Tabachnick, 2004).

1.2.1.3 BT distribution

The distribution of BTV ranges between 53°N and of 34°S latitudes and the disease occurs on all continents except Antarctica (OIE, Bluetongue, 2011). The disease was first named and described in South Africa in the early 1900s (Spreull, 1905). It was considered an African disease until its first appearance in Cyprus, Turkey, Palestine and Israel in the 1940s (Gambles, 1949). BTV incursions occurred in the Iberian Peninsula in the late 1950`s (Manso-Ribiero, Rosa-Azevedo, Noronha, Branco-Forte-Junior, Grave-Periera, & Vasco-Fernandes, 1957) and in Lesbos and Rhodes in the late 1970s (Dragonas, 1981). The disease reappeared in the Greek islands twenty years later and spread over the Balkans, Italy and France by 2005 (Mellor, Baylis, & Mertens, 2009). In 2006, an outbreak of BTV 8 occurred in the Netherlands, Belgium, Luxemburg, Germany, and northern France through an unknown route. This African serotype was completely new to Europe, was able to overwinter and reappear the following year

spreading further north towards Denmark, UK and Czech Republic (Purse, Brown, Harrup, Mertens, & Rogers, 2008). BT establishment in southern Europe and its unexpected incursions in northern Europe are thought to be influenced by climate changes. These resulted in the expansion of *C. imicola* distribution and enhanced the competence and overwintering ability of *C. pulicaris* and *C. obsoletus* (Purse, Brown, Harrup, Mertens, & Rogers, 2008).

In the USA, a Bluetongue-like disease was first reported in Texas in 1952 (Hardy & Price, 1952). The first virus isolate, BTV 10, was obtained from a sheep in California one year later (McKercher, McGowan, Howarth, & Saito, 1953). Since then four serotypes (10,11,13,17) have become endemic and are widely spread across the USA (Barber, 1979). An additional serotype, BTV 2, was isolated in Florida in 1982 and it is now established in the region (Mecham & Johnson, 2005). Since 2004, seven additional exotic serotypes (1, 3, 5, 6, 9, 12, 14, 19, 22 and 24) have been isolated from blood samples collected in the southeast of USA (Johnson, 2007; Maclachlan, 2010). Most of these serotypes are endemic in Central America, suggesting they originated from the south (Homan, et al., 1990).

BTV outbreaks have sporadically been observed in Canada. BTV 11 was detected in British Columbia's Okanagan valley in imported and native cattle in 1975. Seroreactors were found in 1976 but no virus was isolated from cattle or from *Culicoides* spp. trapped in the region (Thomas, Skinner, & Samagh, 1982). Clinical cases of BT occurred in sheep in 1987 and BTV 11 was isolated (Dulac, et al., 1988). Sentinel cattle seroconvertion was reported in 1988, 1998 and 2004. Animals positive for BTV 17 were detected during the 2007-2008 Bovine Serological Survey. One sample also tested positive for EHDV 2. Two of the positives were sampled in Alberta and traced back to animals from a cattle dealer in the Okanagan Valley (Government of Canada, 2012). Despite several introductions of different serotypes of BTV at different time points, neither active transmission of the virus in British Columbia nor persistence over the

winter have been demonstrated. Moreover, surveillance programs implemented after the outbreaks have all supported the assumption that the virus has disappeared entirely after each incursion (Sterritt & Dulac, 1992).

1.2.1.4 BT policy

Bluetongue was one of the first diseases listed by World Organisation for Animal Health (OIE) in the 1968 International Zoo-Sanitary Code (OIE, 1968). The OIE Terrestrial Animal Health Code (OIE, 2011) sets out detailed requirements in order for a country to be considered BTV-free on the basis of common standards that would ensure safe international trade of animals and animal products. This agreement states that a BT-free country has to demonstrate its status through an ongoing surveillance program to demonstrate no BTV transmission or no *Culicoides* presence. A BTV free country in which potentially competent *Culicoides* spp. are present does not lose its status when importing animals from infected countries if the animals are vaccinated or have tested negative for specific antibodies against BTV 60 days before the shipment (OIE, 2011). The general requirements for an acceptable BT surveillance program include veterinary services and an early warning system which must be in place in order to identify and report suspected cases as well as an active serological and virological surveillance program that is able to determine the infection status of a state or a country. Vector surveillance is used as a tool to identify zones with different levels of risk and has an important role in identifying potential areas of disease transmission.

Canada is adjacent to the USA, a BTV endemic country, (MacLachlan & Guthrie, 2010) and *C. sonorensis* is present in Alberta and British Columbia (Lysyk, 2006, McMullen, 1978). Prior to 2007, Canada`s import policy required serological testing for all ruminants imported from the

USA. Feeder cattle were imported into Canada under the ‘restricted feeder programme’ from selected low-risk states, such as Montana (DeHaven, del Valle Molina, & Evans, 2004). This program allowed the importation of cattle during winter months (October to March) when competent vectors were not likely to be present.

Nevertheless, following experts and public consultation, in 2007, the Canadian Government removed the requirement for BT serological testing for all ruminants imported from USA and changed the status of BTV US endemic serotypes from the reportable disease list, for which control and eradication measures are applied, to the immediately notifiable disease list, that includes exotic or rare indigenous diseases for which no control or eradication measures are in place (CFIA, 2007). In order to mitigate potential risk of incursions of BTV from the USA, the Canadian Food Inspection Agency (CFIA) enhanced its BT surveillance activities and supported scientific research to confirm that the risk of BT transmission is low enough to justify the changes in animal import policy and disease reporting (CFIA, 2010a). In 2009, the CFIA modified the BT domestic policy to include new USA serotypes and follow-up epidemiological investigation and testing for any BT serological reactor (CFIA, 2009). Canada`s BT domestic policy was updated in July 2010 in response to key changes in the OIE 2010 Terrestrial Animal Health Code. This directive proposed three different Bluetongue risk zones within Canada considering vector and disease status for each specific zone. These zones are the Okanagan Valley, the prairie provinces and eastern Canada (CFIA, 2010b).

1.2.2 Epizootic Haemorrhagic Disease (EHD)

1.2.2.1 EHD virus and disease

Epizootic Haemorrhagic Disease (EHD) is a double stranded RNA virus, in the genus *Orbivirus*, family *Reoviridae* (Attoui, Maan, Anthony, & Mertens, 2009). There are seven recognised EHD virus (EHDV) serotypes worldwide. Serotype 1 has recently been recognised to be the same as serotype 3. Serotype 2, also called Albertan serotype, is the same as Ibaraki strain. Additional serotypes are 4, 5, 6, 7 and 8 (Anthony, Maan, Maan, Sutton, Attoui, & Mertens, 2009). Three serotypes have been isolated in North America: 1, 2 and 6 (European Food Safety Authority, 2009).

EHD is the most important viral disease of white-tailed deer (*Odocoileus virginianus* Zimmermann) in the USA with an estimated mortality rate of 20% (Savini, et al., 2011). Other ruminants, including mule deer, pronghorns, elk, bison and bighorn sheep usually show subclinical signs (Mullen, 2009). Mortality is low in domestic cattle and morbidity ranges between 1% to 18% (Savini, et al., 2011). However, some exceptions exist. In fact, the Ibaraki strain in Asia and the recent EHD 6 and 7 events in the Mediterranean Basin (European Food Safety Authority, 2009) caused severe clinical disease in cattle with morbidity rates reaching 100%.

1.2.2.2 EHDV transmission

Competent *Culicoides* species are thought to be similar to the ones transmitting BTV, although different levels of competence for individual *Culicoides* species are hypothesized for

EHDV (Savini, et al., 2011). In North America, *C. variipennis* and *C. sonorensis* play a central role in the transmission of EHDV (Foster, Breckon, Luedke, & Jones, 1977).

1.2.2.3 EHD distribution

EHD has predominantly been considered a wildlife issue of North America. Coastal south-east USA regions are considered endemic areas for EHD with the disease being reported every two to three years. (Couvillion, Nettles, Davidson, Pearson, & Gustafson, 1981). In central and mid-western USA, disease epidemics are observed every eight to ten years (Nettles, Hylton, Stallknecht, & Davidson, 1992). EHDV 1 was initially isolated in New Jersey, USA in 1955 (Shope, MacNamara, & Mangold, 1955). Serotype 2 was isolated from white-tailed deer in 1962, in Alberta, Canada. Additional EHD outbreaks occurred in Saskatchewan (1986-1987) and British Columbia (1987 and 1999) (Nettles, Hylton, Stallknecht, & Davidson, 1992; European Food Safety Authority, 2009). EHDV 6 has emerged in the North American in 2006 and caused outbreaks in Indiana and Illinois (Allison, 2010).

Although EHD is a disease known to affect wildlife, the Ibaraki strain (EHDV 2) has been known to cause severe clinical disease in cattle in East Asia since the late 1950s (Omori, Inaba, Morimoto, Tanaka, & Ishitani, 1969) and EHDV 6 and 7 clinically affected cattle in Morocco, Algeria, Turkey and Israel (Yadin, et al., 2008; Temizel, et al., 2009). These latter outbreaks demonstrated a change in host preference and in disease distribution, since EHDV had not previously caused clinical disease in cattle in the Mediterranean Basin. More recently EHDV caused clinical disease in cattle also in Wyoming, USA (ProMED archive number: 20121010.1334173, 2012) which raises awareness for EHDV switches in host preference that

might occur more extensively in North America and could cause important economic losses to the cattle industry.

1.2.2.4 EHD policy

EHD is listed as a multiple species disease in the Terrestrial Animal Health Code of the World Organisation for Animal Health (OIE). It was added to the OIE list of notifiable diseases in May 2008, following the changes in disease epidemiology and distribution in the Mediterranean Basin (OIE, 2009).

In Canada, EHD is a federally immediately notifiable disease under the Health of Animals Regulation (CFIA, 2011). In Alberta, EHD has to be notified to the competent authority when it occurs in wild and domestic cervids (Alberta Agriculture and Rural Development, 2008).

1.2.3 Other *Culicoides*-borne viral diseases

There are several viruses that might use *C. sonorensis* as a disease vector (Mellor, Boorman, & Baylis, 2000). These include Vesicular Stomatitis, which has occurred in North America and Schmallenberg, an emerging disease in European livestock. Both of these diseases have the potential to become emerging viruses in North America.

1.2.3.1 Vesicular Stomatitis (VS)

Vesicular stomatitis (VS) is caused by a vesiculovirus in the Rhabdoviridae family. It is a vesicular disease that affects cattle, horses and pigs, although several other species (humans included) are susceptible in the Americas. The disease is relevant because it is clinically

indistinguishable from foot and mouth disease. Two serotypes, New Jersey and Indiana, have been recognised and both have caused epidemics in the USA on an eight to ten year cycle (Hanson, 1952; Rainwater-Lovett, Pauszek, Kelley, & Rodriguez, 2007).

Most of the outbreaks in the western USA have originated in Mexico and spread northwards along the Rocky Mountains. VS occurred in horses in Montana during summer 2005 (Montana Department of Livestock, 2005). The disease has been reported as far North as Canada in 1937 and 1949 (Hanson, 1952).

Black flies (family Simuliidae) are considered the primary biological vectors for the virus (Schmidtmann, Tabachnick, Hunt, Thompson, & Hurd, 1999). However, colonized *C. sonorensis* have the potential of sustaining VS virus replication and dissemination (Nunamaker, Peréz De León, Campbell, & Lonning, 2000; Drolet, Campbell, M.A., & Wilson, 2005). Virus transmission to cattle has been demonstrated in laboratory settings, after intrathoracic infection of *C. sonorensis* (Perez de Leon & Tabachnick, 2006). Moreover, VSV has been isolated from *C. variipennis* (likely *C. sonorensis*) during a disease outbreak in Colorado and Utah (Kramer, Jones, Holbrook, Walton, & Calisher, 1990)

VS is listed as a notifiable disease by the OIE and it is a federally reportable disease to the Canadian Food Inspection Agency (CFIA, Federally Reportable Diseases for Terrestrial Animals in Canada, 2013).

1.2.3.2 Schmallenberg Disease (SBD)

Schmallenberg virus (SBV) has recently been detected in northern Europe. The virus belongs to the Simbu serogroup in the Bunyaviridae family, together with the better known Akabane disease agent (Goller, Höper, Schirrmeier, Mettenleiter, & Beer, 2012). Clinical signs are

generally mild in adult ruminants, except during pregnancy when foetus malformations and abortions are observed (European Food Safety Authority, 2012).

The virus was initially isolated from dairy cows in Germany in November 2011. By the end of 2012, 15 European countries (Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Poland, Spain, Sweden, Switzerland and United Kingdom) had confirmed SBV transmission in domestic ruminants (European Food Safety Authority, 2012).

The virus has been isolated from pools of Palearctic *Culicoides* species collected during entomological surveillance activities in Europe (De Regge, et al., 2012). Several years ago *C. sonorensis* was proved to be a biological vector of Akabane virus (Jennings & Mellor, 1989). More recently, *C. sonorensis* was tested for vector competence and a full SBV dissemination was demonstrated (Veronesi, et al., 2013). At the present time no policy has been developed in Europe. The OIE recommend the control of potential vectors to decrease the risk of disease transmissions as well as the rescheduling of animal breeding outside the vector season to reduce fetal malformation (OIE, 2013). Schmallenberg has not yet been reported in North America, but the presence of competent vectors suggests the need of a surveillance system in place for both vectors and hosts.

1.3 SPECIES DISTRIBUTION MODELING (SDM)

Species distribution modeling is an empirical process that combines species occurrence data with environmental variables that are deemed to govern, directly or indirectly, species distribution (Franklin, 2009; Elith & Leathwick, 2009; Guisan & Zimmermann, 2000). A well-established framework to properly build a SDM was presented by Austin (2002). This approach consists of three parts: the ecological model; the data model and the statistical model.

1.3.1 The ecological model

The ecological concepts that support an empirical model of a species distribution include the purpose of the study, the characteristics of the species, the climatic and environmental predictors and the possible species response curves to the selected environmental variables. These concepts will determine important choices about data collection and selection, as well as model selection and calibration.

A central concept that underpins SDMs is niche theory. SDMs attempt to describe the ecological niche of a species. Grinnell (1917) defines the niche as a suitable habitat allowing species survival, whereas Hutchinson (1957) considers it as a characteristic of the species more than of the environment that is supporting its presence. Within this broad concept, Hutchinson further identifies the fundamental niche and the realized niche. A fundamental niche is an environmental space in which a species can survive and develop constrained only by its physiological and habitat requirements. As a consequence, a species fundamental niche is a theoretical space that can be potentially occupied when biotic interactions such as species competition are not taken into consideration. A realized niche is defined if biotic interactions and geographical barriers are considered. A realized niche is usually a smaller portion of the fundamental niche, in which the species is actually present (Pulliam, 2000). Awareness of the differences between the potential and realized niche is essential when applying these concepts to SDM and what they actually represent. SDMs are considered to depict a realized niche of a species since they are based on field data rather than theoretical information (Guisan & Zimmermann, 2000; Austin, 2002).

However, this might not be true when considering metapopulation theory (Hanski, 1999; Levins, 1969) and source-sink dynamics (Pulliam, 1988). A metapopulation consists of a regional group of a species that is connected and interact at some level with subpopulations of the same species located in a different region. A source habitat is an area where population growth is positive, whereas sink habitats have populations that cannot maintain themselves and the population is generated by dispersals or migrations from source areas. Assuming that habitat occupancy corresponds to habitat suitability might lead to errors since a species can be found in an unsuitable patch because of migration processes (sink areas) and be absent in suitable environments due to dispersal limitations (Pulliam, 1988). Therefore the relationship between species niche and distribution models is complicated and continues to undergo extensive reviews and discussions (Guisan & Zimmermann, 2000; Austin, 2002; Soberon & Peterson, 2005; Jimenez-Valverde, Lobo, & Hortal, 2008).

The spatial and temporal scale of the data used in SDMs must be identified within the ecological model formulation because at different scales patterns of interaction between species and environment can change (Levin, 1992). Spatial scale is defined by grain and extent. Grain refers to sample resolution and extent refers to the dimension of the study area. (Meyer & Thuiller, 2006). Temporal scale refers to the time-frame in which species and environmental data are sampled and can include climate change scenarios on which to project species data. In practice, the choice of the scale is driven by the study purpose and the understanding of the ecological process but is limited by data availability. The art of modeling consists in the ability to lose some resolution and detail while still capturing the relevant dynamics of the ecological process of interest (Levin, 1992).

1.3.2. The data model

In addition to the ecological assumptions, two other key factors affect the resulting prediction: species data and environmental data. The data model specifies materials and methods for the species and the climate and environmental data to be analysed during the statistical model step (explained in section 1.3.3).

1.3.2.1 Species data

Species data for distribution modeling purposes can be collected using systematic, random or stratified sampling strategies (Franklin, 2009). In a *systematic sampling* approach subjects of the population included in an ordered study frame are selected at regular intervals, once the first case has randomly been chosen. In a *random sampling* method each element has the same probability of being chosen. A *stratified sampling* design involves the identification of different strata with similar characteristics, among which a random sample can be chosen to represent each subpopulation. A well-known approach combining random and stratified sampling methods is "gradsect sampling" (Gillison & Brewer, 1985; Austin & Heyligers, 1989). This approach provides a description of the whole range of biotic characteristics occurring in the study area by sampling only along selected transects having the strongest environmental gradients. However, sometimes the only available species data are existing samples, which are often collected following an opportunistic or purposive method (Franklin, 2009).

Species data obtained in order to build the distribution model can include presence/absence data, presence-only data or even abundance information. When dealing with SDMs, it has being suggested to start looking for correlations between environment and species occurence, before investigating the relationship between response variables and species abundance (Austin & Cunningham, 1981).

Sample size is another important issue which has been proven to be related to SDM performances. As a rule of thumb, a reasonable number of species records is equal to the number of climatic and environmental variables (n) used in the model multiplied by ten (i.e. $n \times 10$) (Anderson, 2008). However, what might be more important than a large sample size is the homogeneous distribution of samples throughout the entire geographical range of the species (Kadmon, Farber, & Danin, 2003). Obtaining environmentally representative species data will help not only to built an accurate model but also to extrapolate the predictions to different climate scenarios (Franklin, 2009).

1.3.2.2 Climate and environmental data

Climate and environment have an impact on the biology and the distribution of disease vectors (Lafferty, 2009). The selection of climatic and environmental data is therefore an important part in species distribution modeling. The increasing availability of candidate predictors, especially from remote sensing sources, facilitate the tendency to include as many variables as possible and leave the selection to an automatic process such as stepwise procedures. However, it is strongly advised to use an *a priori* selection based on ecological relevance and expert opinion in order to obtain a simpler model and a meaningful prediction (Hirtzel & Le Lay, 2008; Anderson, 2008).

Candidate predictors were classified by Austin (2002) into proximal and distal variables. *Proximal variables* are those that cause a stronger responses on species occurrence, whereas *distal variables* are related to proximal predictors and therefore to species distribution, but do not have a strong effects on it. These variables are also called proxies and are easier to obtain and are commonly used in SDM. A further classification by Austin identifies direct, indirect and

resource variables. *Direct predictor*, or conditions, have a direct effect on species physiology and can be differentiated from resource variables because they cannot be consumed by the species. *Indirect factors* are correlated to direct predictors but do not have direct consequences on species distribution. Indirect factors are always proxies. An example of a direct predictor is temperature, whereas examples of resource variables are nutrients and water. A classic indirect, distal predictor is elevation, since it has a direct inverse correlation with temperature.

Generally speaking, the use of proximal and direct variables leads to a stronger prediction of species occurrence. Moreover, models based on proxies will have a limited power when extrapolating the prediction to different scenarios and will be more complicated to explain (Austin, 2002). On the other hand, Guisan and Zimmerman (2000) suggest that a combination of resource and direct factors can be substituted by indirect predictors, reducing the number of variables and therefore model complexity.

The assumptions about species response to climatic and environmental predictors are as important as variable selection in SDM. Response curves graphically describe the relationship between species occurrence (or abundance) and gradients of variables. The general assumption of niche theory is that response functions follow a Gaussian normal distribution (Tilman, 1982). However, skewed, bimodal or multimodal response curves can be expected, especially when dealing with direct and resource predictors at their gradients` extremes (Austin, 2002). The form of a species response curve to indirect gradients depends on its relationship to the underlying direct variable and therefore is unpredictable. It is also important to consider that most variables covary in nature and interact among each other, adding complexity to the interpretation of response curves (Hirtzel & Le Lay, 2008). The selection of biologically important variables and the understanding of their interactions as well as the response of the species to their variability in

current climate is fundamental to be able to model species distribution under climate change conditions.

1.3.2.3 Climate data under future scenarios

The sensitivity of vectors to climate suggests that climate change will affect their distribution in the future (Lafferty, 2009). This hypothesis stimulated the research agenda to focus on the potential consequences of climate change on vectors and vector borne diseases (e.g. Acevedo, et al., 2010).

General Circulation Models (GCMs) are advanced mathematical models that describe physical processes of the Earth system. A GCM can describe a single component, usually the atmosphere, or be coupled to describe the interaction between different processes such as the behaviour of the atmosphere and its relationship with oceans, ice, land and vegetation (Donner, Schubert, & Somerville, 2011). Since the early stages of GCMs development, one of the key research goals was the ability to forecast weather and climate variability at different temporal scales (Pfeiffer, 1960). The proliferation of GCM activities coincided with the great advance in computer technology and increasing concerns about climate change started in the 1980s (Donner, Schubert, & Somerville, 2011). This led to the founding of the Intergovernmental Panel on Climate Change, IPCC in 1988. Since then, IPCC reports regularly assess GCMs reliability, develop climate change scenarios and are an established science-based source for policy implementation (IPCC, 1990; IPCC, 1996; IPCC, 2001; IPCC, 2007).

In the past, the IPCC recommended a set of climate scenarios as a common base for climate change impact experiments (Nakicenovic, Swart, & al, 2000). It depicted four scenario families (A1, A2, B1 and B2) that explore the directions that the world might experience on the basis of different inputs of greenhouse gas (GHG) emission drivers (i.e. population and economic

growth, technological choices). Recently, the scientific community expressed the need of new scenarios that would take into account not only concentrations and emissions of GHG and air pollutants but also land use changes, environmental policy options and adaptation processes (Moss, et al., 2010). Four representative concentration pathways (RCPs) were designed and will be part of the IPCC fifth assessment report which will be finalized in 2014. RCPs are the result of a comprehensive literature review on possible trajectories of the driving forces of climate change and constitute the input information for GCMs in order to produce climate change projections. Each pathway is defined by a radiative forcing value which describes the change in the amount of energy that enters the atmosphere and the quantity that is reflected back, and is expressed in Watts per meter square of surface (W/m^2). The baseline year in which the irradiative balance is set as zero is 1750, while the current value is estimated to be 1.6 W/m^2 (IPCC, 2007). The radiative forcing levels estimated to be reached by the end of the century range from 2.6 W/m^2 to 8.5W/m^2 (van Vuuren, Edmons, Kainuma, Riahi, & et al, 2011). Low radiative forcing levels of 2.6 W/m^2 describe a mitigation scenario and might be compared to the previous B1 scenario (Nakicenovic, Swart, & al, 2000), a medium forcing levels of 4.5 W/m^2 depict a stabilization scenario while high radiative forcing levels (8.5 W/m^2) are comparable to the A1 high emission scenario (Nakicenovic, Swart, & al, 2000). Projected values for specific climatic variables can be obtained for each pathway and may vary depending on the underlying GCM. The resolution of data is generally coarse since circulation models, on which the variables are estimated, attempt to describe the full Earth system and cannot include small-scale physical processes.

1.3.3. The statistical model

Austins` statistical model phase (2002) follows three main steps: model formulation, model calibration and model evaluation which were formulated in detail by Guisan & Zimmermann (2000). In this context, these steps will be generally reviewed to justify the choice of MaxEnt as the statistical approach to model *C. sonorensis* distribution in this study.

1.3.3.1 Statistical model formulation

Statistical model formulation involves the choice of a specific statistical approach which varies according to the available data (e.g. presence-absence, presence-only) and its statistical distribution. Several statistical approaches are used in SDM. Linear regression models predict a response (or dependent) variable from a set of predictor (or independent) variables. Since linear models make several assumptions about the structure of the data and these assumptions are often violated in ecology, other statistical approaches are commonly used to model species distribution (Franklin, 2009). Generalized linear models, GLMs, are generalizations of linear models that are able to deal with normal as well as non-normal distributions of the response variable. The generalization is made through a link function that combines predictor variables to the response variable (Guisan, Edwards, & Hastie, 2002). Logistic regression is an example of GLM in which a binomial distribution is used to describe the distribution of the response variable and the logit link function is used to relate the response and predictor variables. Logistic regression is frequently applied in SDMs (Rushton, Ormerod, & Kerby, 2004) and many examples exist for *Culicoides* spp. (e.g. Baylis, Meiswinkel, & Venter, 1999).

A similar approach is discriminant analysis that assumes a multivariate normal distribution for the independent as well as for the dependent variable, which is treated as categorical (i.e.

presence-absence) as in a logistic regression method. This technique creates a multidimensional space in which each point in the study area is assigned a probability of belonging to a group on the basis of independent variables (Rogers, 2006). The number of dimensions depends on the different ways to combine indipendent variables in order to be able to discriminate among groups of dependent variables (Tabachnick & Fidell, 2013). A modified version of discriminant analysis has been used to predict *Culicoides* spp. distribution in Europe (e.g. Baylis, Mellor, Wittmann, & Rogers, 2001).

Machine learning methods belong to a different group of SDM where, through a given algorithm, rules are developed by the machine in order to correctly classify new cases on the basis of the existing relationship between observations and environmental data. (Franklin, 2009). Two common machine learning methods used for SDM are decision tree-based methods and maximum entropy.

Decision tree-based models are processes that sort dependent categorical variables in similar groups, based on the range of values of independent categorical and continuous variables. Each decision process is made of three steps: tree building, tree stopping and tree pruning. In the *tree building* part, the database is partitioned in subgroups with similar characteristics at decision nodes. The splitting process continues until the *tree stopping* phase is reached (terminal node or "leaf"). The last step, *tree pruning*, aims to simplify the decision tree, reducing the number of nodes (Olden, Lawler, & Poff, 2008). One example of decision tree is Random Forests. Random Forests consists of a large number of decision trees that are built independently and then averaged (Breiman, 2001).This approach has been used for *Culicoides* spp. modelling in the Iberian Peninsula (Peters, et al., 2011) and performs well in comparison to linear discriminant analysis and logistic regression.

Maximum entropy is another machine learning approach increasingly applied in SDM, although it has not yet been used for modelling *Culicoides* spp. distribution. A dedicated software, MaxEnt (Phillips, Anderson, & Schapire, 2006), has been developed for species presence-only data. This choice, compared to presence/absence approaches, avoids the uncertainty around negative findings while not making any assumptions (i.e. leaving maximum freedom or entropy) about unsampled locations. More details on this specific method will be given in section 1.4.

1.3.3.2 Statistical model calibration

Model calibration, or model fitting, adjusts the parameters of the model so that the overall prediction becomes closer to the empirical data (Rykiel, 1996). Parameter estimation is done through measurements of variance or deviance reduction in regression models. Generally, the model with lower deviance is considered the best, providing a simpler and clearer interpretation. Deviance reduction is obtained by maximizing the log-likelihood by adjusting model parameters (Franklin, 2009). In discriminant analysis, parameters are calibrated through Wilk`s λ statistics, which is a multivariate analysis of variance (Everitt & Dunn, 1991). The smaller the lambda for an independent variable, the more that variable contributes to the prediction. In tree based techniques pruning, together with cross-validation, can be considered a method for complexity reduction. Decision-tree based techniques do not address the problem of variable and model selection, since the final decison tree is not necessarily the best (Olden, Lawler, & Poff, 2008). Random Forests partially solves this issue since terminal nodes are averaged and the most recurring ones are selected(Breiman, 2001).

MaxEnt calibrates its model output by calculating the gain, which is a measure of goodness-of-fit similar to deviance reduction in GLMs. The gain is defined by the improvement in an average log-likelihood, minus a constant that describes a uniform distribution corresponding to a null model with zero gain (Elith, Phillips, Hastie, Dudik, Chee, & Yates, 2011).

Guisan & Zimmermann (2000) extended model fitting to the variable selection process. Variable selection, as well as model selection, aims to identify the best set of predictors, based on their fit to the response variables. In regression analysis, variable selection is performed by stepwise elimination of redundant variables, retaining the best performing subsample of variables (Franklin, 2009). The Akaike Information Criterion, AIC (Akaike, 1973) is another well established approach that selects models based on parsimony (i.e. few explanatory variables) and performance (i.e. more deviance explained). The AIC approach has its basis in information theory, where the main goal is to quantify the amount of information lost when a model is used to describe reality. In practice, a model is selected among a group when it yields the smallest AIC value (or AIC_c, an AIC corrected for small sample sizes) that corresponds to the minimum amount of information lost when estimating reality with the selected model (Anderson, 2008). The outcome identifies the best model in a group of similar models and is not related to any absolute measure of goodness of fit. This topic will be examined in the section below.

1.3.3.3 Statistical model evaluation

Model evaluation or validation, is the final step of model building that defines the degree of prediction accuracy. Evaluation also refers to the assessment of model performance based on the specific aim of the study and its applicability (Rykiel, 1996).

There are two main options when undertaking the validation process. The first method consists in using a single database whilst the second involves two separate databases.

A single dataset is commonly used for calibration as well as validation when the number of observations is too small to allow further splits of the sample. When the sample size is large enough, a single dataset can be divided into two sets (Guisan & Zimmermann, 2000); one is used as a training dataset for model calibration and the other one is used as a testing dataset for model validation. In both cases the most common and appropriate techniques applied to validate model outputs are cross-validation, bootstrap and jackknife. These methods can be used to assess the stability of the prediction before evaluating its accuracy through threshold-dependent and independent measures of accuracy. Sensitivity and specificity are examples of threshold-dependent measures of accuracy. These methods are used when dealing with categorical response variables (presence/absence) that can be arranged in a contingency table and examined for false positive and false negative rates. The area under the curve (AUC) of the receiver-operating characteristic (ROC) plot (Hanley & McNeil, 1982) is a threshold-independent method that shows the probability of a case to be correctly classified by the model compared to a random prediction. It ranges from 0 (no fit) to 1 (perfect fit) and in presence-only approaches, such as MaxEnt, a value above 0.75 is suggesting that the model can correctly classify presence locations against random background points (Fielding & Bell, 1997). The AUC value is independent of the prevalence of species occurrence and is therefore a useful evaluation method when using presence-only data, in which any prevalence information is excluded.

1.4 MAXIMUM ENTROPY APPROACH AND SDM

In 1948, Shannon took the expression of entropy from the field of thermodynamics and made it the core of information theory (Shannon, 1948). More than 50 years later, Phillips et al. (2006) implemented a machine learning method, called MaxEnt, based on the maximum entropy principle and specifically designed for SDMs. In this section, I review the basic principles in the entropy concept, its application in MaxEnt as well as its properties, extensions and limitations.

1.4.1 Information theory, entropy and maximum entropy

Information theory was developed by C.E. Shannon (1948) to quantify the amount of information that can be obtained while studying a process. Entropy is the probability of occurrence of a particular event within a process and is defined by:

$$H = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant and p_i is the probability of occurrence of event i.

The main properties of H are the following: a) it is continuous in p_i, b) it is a monotonic increasing function when all p_i are equal. This would be the most uncertain situation, since more choice is involved c) it is 0 when all but one p_i is 0 because there is no uncertainty around one possible event.

Information theory and the entropy expression provided the basis for key developments in inference statistics. A reasonable statement that can be made about a partially known process is that it is a description of the probability of occurrence of certain events, with their own probability distribution. The best way to be consistent with the available information about the process is to make inferences using the events probability distributions with maximized entropy (Jaynes, 1957).

The maximum entropy distribution concept has been used to describe stochastic processes in the machine learning community (Berger, Della Pietra, & Della Pietra, 1996). Stochastic modellers aim to describe a random process on the basis of some empirical measurements. In other words, machine learning researchers look for the conditional probability, among the set of all conditional probabilities, that better explains an outcome *y*, given a value *x*, *f(x,y)*. The maximum entropy principle provides the criterion to select the most appropriate distribution. The best conditional probability distribution is the one with maximum entropy that is the closest to uniform or equal across the whole study area, respecting the constraints set by the empirical observations. A constraint is a numerical relationship between the expected value of *f(x,y)* in the model and its empirical value. The function *f(x,y)* is commonly called *feature* in the maximum entropy environment.

1.4.2 MaxEnt

MaxEnt (Phillips, Anderson, & Schapire, 2006) is a machine learning method that was developed at Princeton University, USA and aims to apply the maximum entropy concept to species distribution modelling. MaxEnt produces a model of the species probability distribution with maximum entropy subject to environmental constrains, called features, which ensure a solution as close as possible to reality. The study area is the space where the MaxEnt probability distributions are defined. The species occurrence records are the sample points and the environmental variables are the features. These constraints require that the expected mean of each feature has to be similar to its empirical mean over the presence sites. MaxEnt software, version package 3.3.3e, provides six feature classes: *linear* which is equal to the continuous environmental variable, *quadratic* which equals the square of the variable, *product* which equals

the product of a pair of variables whereas *threshold* and *hinge* features are step functions that describe different responses below or above a threshold or knot. In addition, a *category* indicator is implemented in order to allow the use of nominal variables in the model (Phillips & Dudik, 2008).

As a consequence, the main focus of MaxEnt is to maximize the entropy of the distribution, minimizing the relative entropy or information gain, a measure of goodness of fit which is obtained for each iteration of the model. The maximum entropy distribution is defined as the ratio between the probability density distribution of environmental variables where the species is present $f_1(x)$ over the probability density distribution of environmental variables across the study area $f(x)$ (Elith, Phillips, Hastie, Dudik, Chee, & Yates, 2011).

The probability distribution resulting from such a ratio is a Gibbs distribution (Jaynes, 1957; Berger, Della Pietra, & Della Pietra, 1996):

$$f_1(x)=f(x)e^{\alpha+\beta*h(x)}$$

where α is a normalizing constant, β is the contribution weight for each feature and $h(z)$ is the vector of features. A regularization parameter lambda (λ) has been introduced to prevent MaxEnt of matching the empirical feature means too closely to the real feature means of the study area providing complex models not appropriate for generalization. *Lambda* sets the range of values, called error bounds and conceptually similar to a confidence interval, for each feature class over presence sites *m* (Elith, Phillips, Hastie, Dudik, Chee, & Yates, 2011). Default λ parameters in MaxEnt were tuned over presence points of an international dataset covering six geographic regions (Elith, Graham et al., 2006; Phillips & Dudik, 2008).

The MaxEnt distribution is defined by the following equation:

$$\frac{1}{m}\sum_{i=1}^{m} \ln\left(f(\mathrm{x_i})e^{\alpha+\beta*h(x)}\right) - \sum_{j=1}^{n} \lambda_\mathrm{j}\,|\beta_\mathrm{j}|$$

where x is the feature vector for presence point i of sites m and for j = 1...n features.

The first term of the equation is a logarithm of the likelihood representing how well the model identifies presence sites compared to background points. The second term is ruled by the regularization parameter λ and becomes larger when the weights β are bigger and therefore the model tends to overfit.

The best model would be the one that maximizes the difference between the log likelihood and the regularization, or in other words, the one that minimizes the relative entropy subject to the error bound constraints, balancing model fit and complexity (Elith, Graham, & et al, 2006).

MaxEnt delivers three model output formats. The core one is a "raw" output that assigns a relative probability of presence to each site (i.e. pixel) during the model training. Since the assigned values must sum to 1 over the whole study area, they might end up being very small, especially when dealing with large numbers of background points. To overcome this problem, a cumulative format was introduced. It provides an output describing the omission rate (i.e. the fraction of positive localities that fall into sites that were classified as negative) predicted by the model. However this output delivers values that are not proportional to the probability of presence, the reason that the logistic output was implemented (Phillips & Dudik, 2008). The logistic output provides the probability that the species is present, given environmental information on the study area and is defined by the following equation:

$$\Pr(y=1 \mid x) = \tau e^{\alpha+\beta*h(x)-r} / (1-\tau + \tau e^{\alpha+\beta*h(x)-r})$$

where τ is the probability of presence in sites estimated to host the species and is set at 0.5 in MaxEnt. The value r is the estimated relative entropy of $f_1(x)$ from $f(x)$.

MaxEnt software is a powerful tool that provides a species probability distribution that converges to the one with maximum entropy. It has several advantages, including the need of few presence points even in large scale studies and the ability to use it with both continuous and categorical variables (Phillips, Anderson, & Schapire, 2006). Recent software extensions allow

MaxEnt to fit complex dependencies between presence data and environmental variables and avoid over-fitting through fine-tuning of regularization parameters (Phillips & Dudik, 2008). It has also shown to perform better than other highly performing presence-only methods (Elith, et al., 2006). Despite these advantages, MaxEnt predictions can be highly affected by background environmental data and therefore need to be used with caution when trying to extrapolate them to different areas or climate conditions (Phillips & Dudik, 2008).

Elith et al. (2010) suggest that more reliable predictions can be obtained with MaxEnt when using few biologically meaningful variables to build predictions because it reduces model complexity and the contingency of species-variable relationships. Additional care has to be taken when selecting climate change scenarios. Prediction reliability decreases when projecting to extreme climate change scenarios because of the potential novelty in variable interactions and therefore species response to them. The coarse resolution of climate projections and the impossibility of validating species distribution models under future climate scenarios suggests further caution when interpreting the results and when drawing conclusions on potential distribution changes.

On the basis of the above information on *Culicoides*-borne diseases and *C. sonorensis* ecology, this thesis presents the process undertaken to develop *C. sonorensis* distribution maps under current and future climate conditions, using a MaxEnt approach, as the first step towards the identification of areas at greater risk for *Culicoides*-borne disease emergence in southern Alberta (Canada) and Montana (USA).

RESEARCH RATIONALE AND OBJECTIVES

Alberta is Canada's largest producer and exporter of live cattle. Between 2000 and 2010 (excluding 2004 in which cattle export was banned due to bovine spongiform encephalopathy, BSE), Alberta exported 4,243,533 head of live cattle to the USA, with an average of 424,353 head per year. The economic value of such a trade has been estimated to exceed 500 million Canadian dollars (Statistics Canada, CATSNET Analytics) per year, with the exception of 2004. In addition, deer farming is an expanding sector in Canada, with almost 2,000 licensed deer farms in the whole country and more than 1000 of them located in Alberta and Saskatchewan (Deer and Elk Farmers' Information Network, 2003). Most of the farming facilities are located in southern Alberta, an area that corresponds to the northern edge of *C. sonorensis* range (Lysyk, 2006). Montana, which lies on the southern border of Alberta, recently reported *Culicoides*-borne diseases like BT, EHD (Rolston & Johnson, 2012) and VS, raising concerns about vector-borne disease risk for susceptible Alberta livestock in the near future.

Culicoides-borne diseases like BT, EHD and VS are listed diseases according to the OIE because of their economic consequences in the case of outbreaks. The OIE has also recently introduced a section in its Terrestrial Animal Health Code to outline surveillance methods for vectors of arthropod-borne listed and emerging animal diseases (OIE, 2012). This requires knowledge of the potential vector distribution. Modeling can be an essential component of the risk assessment process and requires reliable information on the vector and its distribution to be able to later address specific disease transmission and distribution questions. Probability distribution maps of vectors are therefore needed to start the risk assessment process on vector-borne diseases. They can be used by risk management experts to help determine the extent of vector surveillance

required, to build early warning systems in the case of disease introduction, as well as provide solid platforms for investigations on potential disease transmission.

The key objectives of this research were:

- to investigate the current distribution of *C. sonorensis* on a large scale using historical and current entomological occurrence data from southern Alberta and Montana. A MaxEnt approach was used to predict the probability distribution of the disease vector on the basis of its relationship with selected climate and environmental variables. The results will be used to highlight areas at greater risk of *C. sonorensis* occurrence.
- to assess potential changes in vector distribution in southern Alberta and Montana under future climatic scenarios. The distribution model under current climate conditions was used to develop short term (i.e. 2030) and medium term (i.e. 2050) probability distribution maps of *C. sonorensis* under the IPCC most recent scenarios on future climate. The results will forecast potential changes in vector distribution under future climate scenarios, providing risk managers with useful tools for disease preparedness.

The materials and methods used to select the predictor variables and develop the probability maps under current and future climatic conditions are presented in Chapter 3. Chapter 4 outlines the results of the present research. Chapter 5 provides the overall discussion and implications of research findings while suggesting potential future directions for *Culicoides*-borne disease research in the region.

MATERIALS AND METHODS

3.1 STUDY AREA

The present study is based on data collected in the southern part of Alberta, Canada and the state of Montana, USA (Figure 3.1). The area spans from the 44th to the 54th parallel North and from the 104th to 120th meridian West, measuring 618,337 km^2. The Rocky Mountains characterize south-western Alberta and western Montana. South-eastern Alberta, central and eastern Montana are defined by prairie landscape. Elevation (mean per Km^2) ranges from 545 m to 3699 m (DEM, ©DMTIspatial via University of Calgary). The main rivers in the region are the South Saskatchewan River (i.e. Bow River in the Rockies) in southern Alberta and the Missouri and Yellowstone rivers in Montana. The climate is considered to be dry continental, with higher precipitation rates along the Rocky Mountains (Peel, Finlayson, & McMahon, 2007).

3.2 ENTOMOLOGICAL DATA

Data on the presence of *C. sonorensis* midges in southern Alberta and Montana were supplied respectively by Dr. Tim Lysyk (Agriculture and Agri-Food Canada, Lethbridge, Alberta) and Greg Johnson (Department of Animal and Range Science, College of Agriculture, Montana State University, Bozeman, Montana). Data were collected over five (i.e. 2002, 2003, 2009, 2010, 2011) and four (i.e. 2002, 2003, 2008, 2011) years respectively within a ten-year period (2002-2011).

In Alberta, midge trapping was conducted using blacklight traps (J.W. Hock Inc., Gainesville, FL) as it captures males that are required to distinguish *C. sonorensis* from *C. occidentalis.*

Females are morphologically indistinguishable while males of *C. sonorensis* have spicules on the aedeagus that differentiate them from *C. occidentalis* (Holbrook, Schmidtmann, McKinnon, Bobian, & Grogan, 2000).

Collections in southern Alberta were designed to study abundance patterns and species composition at cattle feedlot locations and resulted in a comprehensive entomological survey (Lysyk, 2006). In Montana, traps were run as part of a BTV surveillance pilot project that aimed at confirming *C. sonorensis* presence in the state, with a special focus on high risk areas for BTV transmission. The latitude and longitude coordinates of every trap site were recorded with a global positioning system (GPS) device and imported into a geographic information system, ArcMap 10 (ESRI®Inc, 2011).

Four datasets were supplied for Alberta. Set AB - 1 consisted of weekly *Culicoides* spp. catch data for 8 sites sampled in 2002 and 2003 from May to October for a total number of samples ranging from 18 to 22 per site/year (Lysyk, 2006). Set AB - 2 consisted of weekly collections at seven sites in 2009 with 20 to 21 collections per site. Set AB - 3 was collected weekly at ten sites in 2010 and 2011 with 17 to 20 collections per site/year. Set AB - 4 was collected at nine sites in 2010 and 2011 for a total of two weeks at the end of July/beginning of August. Four datasets were supplied for Montana. Set MT - 1 included nightly collections at 31 sites in 2002 for one to nine nights per location. Set MT - 2 consisted of collections made at 17 sites sampled nightly for five to ten nights in 2007. Set MT - 3 consisted of collections made at eight sites for four to eight nights during the summer of 2008. Set MT - 4 consisted of nightly collections for a total of six nights (one site was sampled once) for 12 sites during 2011.

Because of resampling at some locations, 78 trap sites were identified for the study period 2002-2011 (Figure 3.1). The northernmost trap site was located at 53°N-111°W in Alberta and the southernmost trap site was run in Montana at 45°N-105°W. A site was considered a presence

point for *C. sonorensis* if the species was captured at least once for a minimum of one year throughout the sampling period (2002-2011).

3.3 CLIMATE AND ENVIRONMENTAL DATA

3.3.1 Climate and environmental data under current conditions

Candidate environmental and climatic variables (Table 3.1) were chosen based on *Culicoides* spp. distribution modelling (Baylis, Mellor, Wittmann, & Rogers, 2001; Purse, et al., 2004; Acevedo, et al., 2010; De Liberato, et al., 2010) and results of previous studies on *Culicoides* spp. ecology (Schmidtmann, Mullens, Schwager, & Spear, 1983; Mullens, 1989; Lysyk & Danyk, 2007). Each variable was calculated at 1 x 1 km resolution, which is the finest achievable resolution considering the available data and the dimension of the study area (618,337 km^2).

Primary terrain attributes (i.e. elevation and aspect) were obtained from a 30 m resolution Digital Elevation Model (DEM, ©DMTIspatial via University of Calgary). Elevation (mean, minimum, maximum and standard deviation) and aspect (4 classes: North-East, South-East, South-West and North-West) were calculated directly from digital elevation data and used to provide information on the geomorphology of the landscape.

Land Cover, NDVI (Normalized Difference Vegetation Index) and MIR (Middle Infrared Radiation) values at 1 km^2 resolution were derived from MODerate Resolution Imaging Spectroradiometer (MODIS) instrument operating on Terra spacecraft and distributed by the Land Processes Distributed Active Archive Center (NASA Land Processes Distributed Active Archive Center, 2001).

NDVI was used as an indicator of plant growth, vegetation cover and biomass production and is often correlated with soil moisture and rainfall (Hay, 2000). NDVI is calculated by contrasting the red-light region (R) of the electromagnetic spectrum where plants' chlorophyll absorbs incoming sunlight and the near-infrared region (NIR) of the spectrum, where plants' leaf mesophyll determines reflectance, using the following formula:

$$(NIR-R)/(NIR+R)$$

Green vegetation with abundant chlorophyll, has high NDVI values due to low red-light absorption and high NIR reflectance properties. NDVI values range from -1 to 1.

MIR is a sub region of the infrared (IR) wave length and was used as an indicator of water content, surface temperature, and tree canopy density and structure (Hay, 2000).

Cumulative NDVI and MIR monthly means and standard deviations for each month of our sampling season (May-October) were calculated over a 10 year period (2002-2011), which was the longest available using MODIS data.

Land Cover Type 1 dataset is a yearly product in which 17 land cover classes (Appendix 1) are identified by the International Geosphere Biosphere Programme (IGBP). This classification identifies 11 natural vegetation classes, 3 developed and mosaicked land classes, and 3 non-vegetated land classes. We selected the 2004 product since it was the most recent product available when the analyses were performed. We grouped the 17 classes into 7 classes considering the role that they play in *Culicoides* spp. habitat suitability. The new categorical variable included water bodies, evergreen needle leaf forests, open shrubland, grassland, cropland, barren and others.

Monthly mean temperature (T), monthly mean relative humidity (RH) and monthly mean precipitation (P) for each month from May to October were obtained at 32 km resolution from NOAA's (National Oceanic and Atmospheric Administration) National Centers for

Environmental Prediction (NCEP) within the NARR (North America Regional Reanalysis) project (http://www.esrl.noaa.gov/psd/data/gridded/data.narr.html). Vapour pressure deficit (VPD) is a measure of the drying power of the air and it is an important proxy for insect survival (Johnson, 1942). VPD is calculated as the difference between the saturation vapour pressure, *sVP* estimated using mean temperature values, T and the actual vapour pressure deficit, *aVP* estimated using mean relative humidity values, RH for a given month, using the following formulas:

$$sVP = 0.6108 exp^{[17.27T/(T+237.3)]}$$

$$aVP = (RH/100)sVP$$

$$VPD = sVP - aVP$$ (FAO, 1998).

Cumulative means for temperature, relative humidity, vapour pressure deficit and precipitation were calculated by pooling data for a twenty year period (1991-2010) to better describe the climate of the study area. Layers for each variable were computed at 1 km^2 resolution using an ordinary kriging interpolation technique in ArcGis.

3.3.2 Climate data under future scenarios

We considered three Representative Concentration Pathways (RCP) which describe a possible future energy state range of the Earth on the basis of different trends in climate change drivers (Moss, et al., 2010). The RCPs (i.e. RCP 2.6, RCP 4.5 and RCP 8.5) are expressed in W/m^2 and are named according to the radiative forcing levels that might be reached by the end of the century (van Vuuren, Edmons, Kainuma, Riahi, & et al, 2011). Projected values for relevant climatic variables (i.e. Precipitation of May and Vapour Pressure Deficit of July) were obtained from the Fourth Generation Canadian Coupled General Circulation Model (CanESM2)

developed by the Canadian Centre for Climate Modeling and Analysis (CCCma) following the reliability criteria defined by the IPPC. Climate change data for each variable and each RCP scenario were extracted in MATLAB 8 (The MathWorks Inc., 2012) for two time points, the 2030's (2021-2040) and the 2050's (2041-2060).

Since CanESM2 outputs, like all GCM, were at too coarse resolution (i.e. 250 km^2) to represent current weather, historical data (1991-2010) were obtained from the same GCM and for the same predictors to build climate layers following the IPCC guidelines (IPCC, Climate Scenario Data, 2011). Thus, more reliable layers were built by computing the difference (or the ratio in the case of Vapour Pressure Deficit of July) between the CanESM2 projected variables and their CanESM2 baseline values. A further computation was made to account for the variability between the baseline NARR climate model from which Temperature, Relative Humidity and Precipitation data were retrieved (NASA Land Processes Distributed Active Archive Center, 2001) and the CanESM2 GCM. Precipitation in May was calculated for each scenario:

$$P\ May = \left(\frac{CanESM2\ P\ May\ projected}{CanESM2\ P\ May\ baseline}\right) x\ NARR\ P\ May$$

Vapour Pressure Deficit of July was calculated as:

$$VPD\ Jul = (CanESM2\ VPD\ Jul\ projected - CanESM2\ VPD\ Jul\ baseline) - NARR\ VPD\ Jul$$

Predictor surfaces were created for each time point and each scenario in ArcGis 10 (ESRI, 2011). Elevation and land cover were assumed constant for all projections.

3.4 SPATIAL MODELING AND STATISTICAL ANALYSIS

3.4.1 *Culicoides sonorensis* distribution under current conditions

The maximum entropy algorithm, implemented through MaxEnt software version 3.3 (Elith, Phillips, Hastie, Dudik, Chee, & Yates, 2011) was used to model species distribution. It was selected because of the lack in homogeneity in temporal and spatial sampling effort and the paucity of absence points (twenty-eight) compared to presence points (fifty).

As an initial step, we ran the algorithm with five replicates, 500 iterations and 10000 random background points including all the climatic and environmental variables transformed into hinge feature that combines linear and step functions. This feature class alone was chosen since it improves model performance when there are at least 15 presence points (Phillips & Dudik, 2008). Variables that had less than 1% contribution to the increasing training gain or less than 1% permutation importance were considered unimportant and excluded from the subsequent analyses. The 1% threshold was chosen following the example of Flory et al. (2012), whereas the exclusion of variables that were not rated above the selected threshold by both methods was based on Rodda et al. (2011). The important variables were tested for cross-correlation by calculating the Pearson correlation coefficient (r) for raster files implemented in ENMtools (Warren & Seifert, 2011). As a further method to assess variable strength and to reduce the cross-correlation effect, three models were run using one of the most important variables, RH, VPD and P, separately. The variables that consistently achieved high percentages for both contribution and permutation importance and that also obtained higher ranks throughout the runs were kept to build the final models. We formulated the final models only with those predictors that were not strongly correlated ($r > 0.5$) among themselves. For each alternative model, 80% of

the occurrence points were used for training the model and 20% were set apart to test its accuracy. The accuracy of the final models was assessed using the AUC, a threshold-independent method (see section 1.3.3.3).

We compared different models using the information criterion-based model selection with a correction term for sample size, AIC_C using ENMTool (Warren & Seifert, 2011) and identified the "best" model in the group. We also calculated the difference between the $AICc_i$ and $AICc_{min}$ (Δ_i) highlighting the actual distance of each model from the best one (i.e. the one with the lowest $AIC_{C,}$ AIC_{Cmin}). Akaike weights ω were used as a further measure of model plausibility (Anderson, 2008).

For the best model, Jackknife (Quenouille, 1949; Tukey, 1958) technique was used to assess variable relative importance. With the Jackknife test, MaxEnt is run with each variable in isolation and without that specific variable. The training gain for the two models was then compared to the training gain of the final model. Individual response curves describing the relationships between the variation in predictor variables and the probability of *C. sonorensis* occurrence were also examined.

3.4.2 *Culicoides sonorensis* under future scenarios

Ten replicates and 500 iterations of the MaxEnt algorithm were run for each pathway and time point, with only hinge features selected.

Seven average probability maps for *C. sonorensis* distribution were developed: a baseline distribution map (2010s), two RCP 2.6 (2030s and 2050s), two RCP 4.5 (2030s and 2050s) and two RCP 8.5 (2030s and 2050s). From these data, the mean change in probability was computed for each cell (1x1 km pixel) scenario from one time point to the following one in order to

identify potential *C. sonorensis* redistributions using ArcGis 10. Pixels with a change value below -0.05 were considered areas at decreased risk of *C. sonorensis* occurrence, whereas pixels with a change value above 0.05 were considered at higher risk of vector occurrence. Cells in which the change in probability of occurrence fell between -0.05 and 0.05 were treated as no change in probability of occurrence was taking place from one time point to the other in that specific cell.

Five presence probability classes were identified: class 1 (0 to 0.2 probability of occurrence), class 2 (0.21 - 0.4), class 3 (0.41-0.6), class 4 (0.61-0.8) and class 5 (0.81-1). A qualitative comparison by class across pathways and time points was therefore performed.

Table 3.1 Variables used to characterize *C. sonorensis* habitat.

Variable	Description	Source
Elevation (E)	Mean, Minimum, Maximum Standard deviation (m)	DEM, University of Calgary
Aspect (A)	Categorical	DEM, University of Calgary
Land cover (LC)	7 Classes: Water Bodies, Evergreen Needle Leaf Forests, Open Shrubland, Grassland, Cropland, Barren and Others	MODIS (2004)
Temperature (T)	Monthly mean: May-October (K)	North American Regional Reanalysis (NARR) project. (1991-2010)
Relative humidity (RH)	Monthly mean: May-October (%)	North American Regional Reanalysis (NARR) project. (1991-2010)
Precipitation (P)	Monthly mean: May-October ($Kg/m^2/s$)	North American Regional Reanalysis (NARR) project. (1991-2010)
Vapour pressure deficit (VPD)	Monthly mean: May-October (KPa)	Computed using T July and RH July from North American Regional Reanalysis (NARR) project. (1991-2010)
Normalized Difference Vegetation Index (NDVI)	Mean and standard deviation: May-October	MODIS (2002-2011)
Middle infrared radiation (MIR)	Mean and standard deviation: May-October	MODIS (2002-2011)

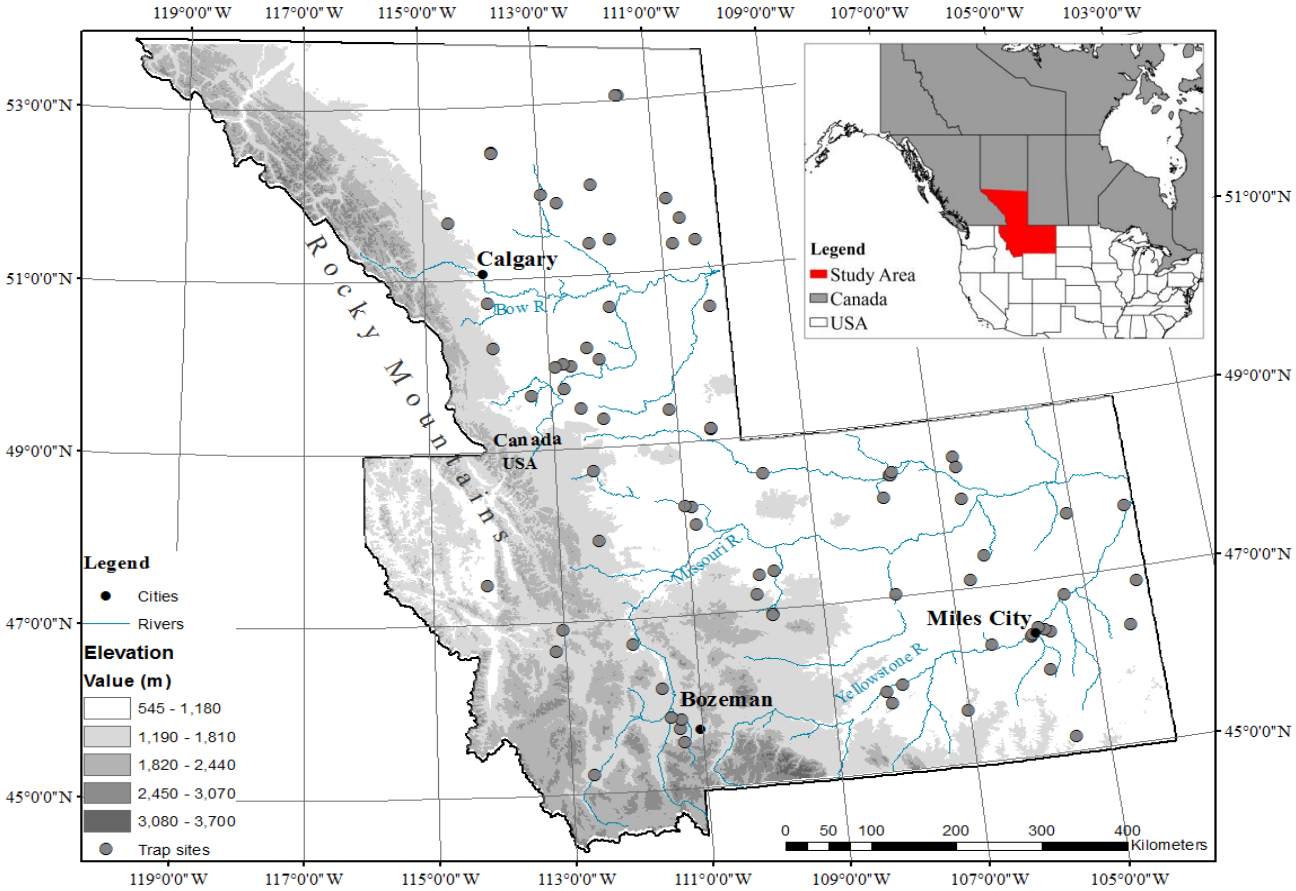


Figure 3.1 Study area (southern Alberta and Montana) with main rivers and elevation gain. Main cities are represented by black dots. Trap site locations (2002-2011) for *C. sonorensis* are represented by grey dots.

RESULTS

4.1 ENTOMOLOGICAL DATA

Culicoides sonorensis was detected in 50 out of 78 trap sites throughout 2002-2011 study period (Figure 4.1).

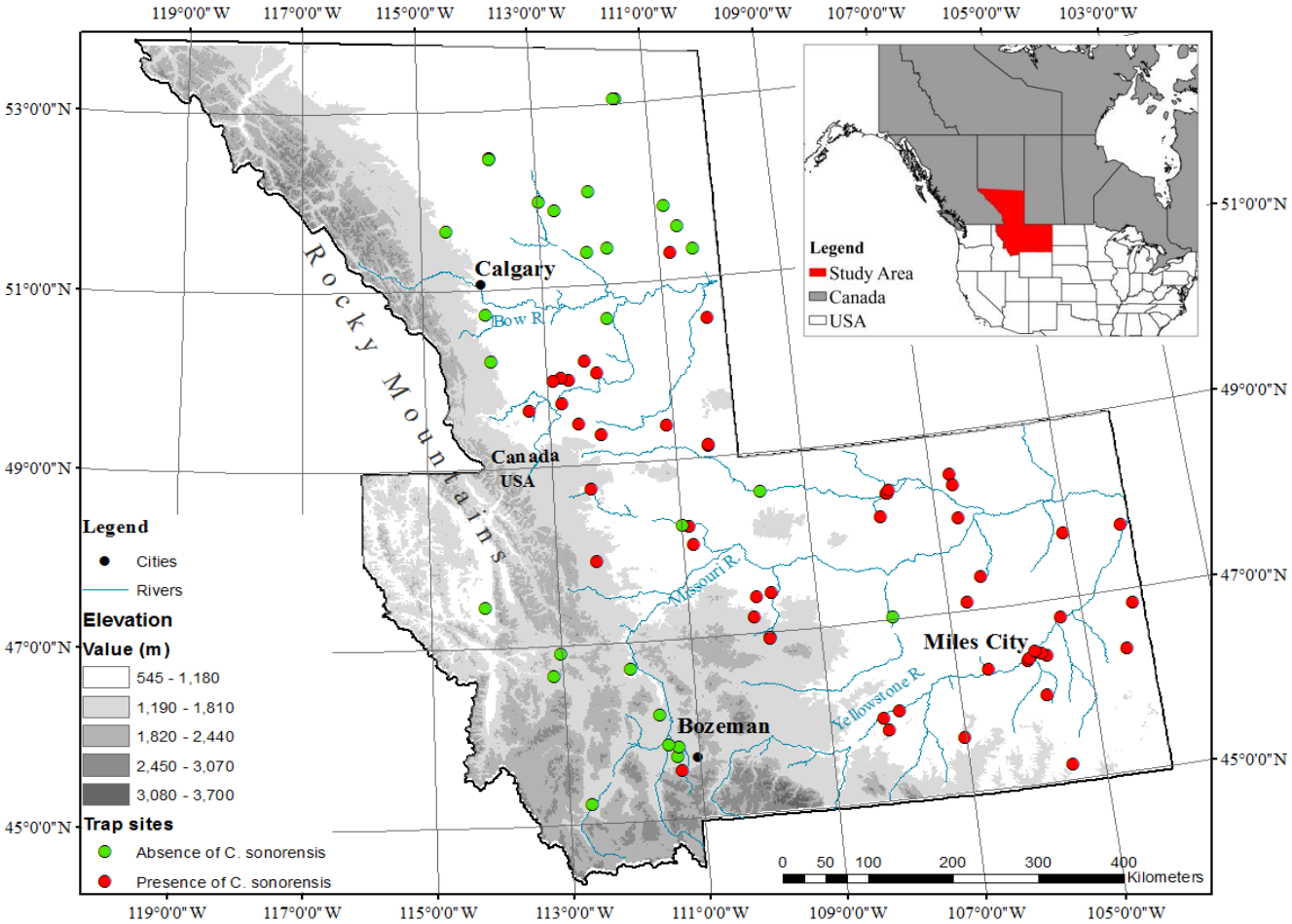


Figure 4.1 *C. sonorensis* presence (red dot) and absence (green dots) data across Montana (USA) and southern Alberta (Canada) collected from 2002 to 2011.

Set AB-1 indicated six positive sites in 2002 after 126 to 154 trap-nights. All trap sites were positive in 2003 following a sampling effort ranging from 161 to 175 trap-nights. Set AB - 2 indicated three positive sites after 140-147 trap-nights. Set AB - 3 indicated seven sites in 2010 after 133-140 trap-nights and in six the following year, after 119-133 trap-nights. After 14 trap-nights, all trap sites of AB - 4 were negative for *C. sonorensis* occurrence (Table 4.1). All negative sites are located in the north and west portions of southern Alberta. Positive sites are clustered in the south-eastern corner of the Province (Figure 4.1).

Table 4.1 Results of *C. sonorensis* collections in southern Alberta.

Group	Source	Year	Sites	Duration	Samples	Trap Night	Positive
AB – 1	**Lysyk 2006**	2002	8	**Weekly**	**18 - 22**	**126 – 154**	6
		2003			**23 - 25**	**161 – 175**	8
AB – 2	**Lysyk (unpubl.)**	2009	7	**Weekly**	**20 - 21**	**140 – 147**	3
AB – 3	**Lysyk (unpubl.)**	2010	10	**Weekly**	**19 - 20**	**133 – 140**	7
		2011			**17 - 19**	**119 – 133**	6
AB – 4	**Lysyk (unpubl.)**	2011	9	**Weekly**	**2**	**14**	0

Dataset MT - 1 indicated 26 positive sites after a sampling effort which ranged from 3 to 27 trap-nights. Set MT - 2 had 15 positive sites after 10 - 20 trap-nights. MT - 3 had seven positive sites after 4 - 8 trap-nights and set MT - 4 had four positive sites after 6 trap-nights and one site showed positivity for *C. sonorensis* after one trap-night (Table 4.2). Positive sites were widespread, ranging from central to eastern Montana (Figure 4.1). *Culicoides sonorensis* was rarely found in southwest Montana and absent from the western part of the state.

Table 4.2 Results of *C. sonorensis* collections in Montana

Group	Source	Year	Sites	Duration	Samples	Trap Night	Positive
MT – 1	**Johnson (unpubl.)**	2002	31	**Nightly**	**1 - 9**	**3 - 27**	26
MT – 2	**Johnson (unpubl.)**	2003	17	**Nightly**	**5 - 10**	**10 - 20**	15
MT – 3	**Johnson (unpubl.)**	2008	8	**Nightly**	**4 - 8**	**4-8**	7
MT – 4	**Johnson (unpubl.)**	2011	12	**Nightly**	**6 ***	**6***	5

*One site (positive) was sampled once.

4.2 CLIMATE AND ENVIRONMENTAL DATA

4.2.1 Climate and environmental data under current conditions

In the first MaxEnt run, 11 variables achieved more than 1% in both contribution and permutation importance. These were: Elevation (standard deviation), Land Cover, NDVI of August (standard deviation), NDVI of October (mean and standard deviation), Precipitation of May, Precipitation of August, Relative Humidity of July and August, Temperature of October and VPD of July (Table 4.3). Cross-correlation was investigated for 10 variables, Land Cover was not included because it is a categorical predictor, calculating the Pearson correlation coefficients (r). Correlation values (r) for the selected variables are shown in Table 4.4. Highly correlated variables ($r \geq 0.75$) included VPD of July, RH of July and August and P of August. Other correlated variables were T of October and VPD of July ($r = 0.78$) and P of May and NDVI of October ($r = 0.56$).

Table 4.3 Variable contribution and permutation importance values achieved during the first MaxEnt run performed with all variables. Variables that obtained values above 1% in both metrics are shown in bold. Variables that scored less than 1% in both metrics are not displayed.

Variable	% contribution	Variable	Permutation importance
RH Jul	**33.7**	**RH Jul**	**26.2**
E sd	**18.9**	**E sd**	**18.5**
LC	**9.8**	**T Oct**	**14.9**
RH Aug	**5.1**	**NDVI Aug sd**	**4.5**
T Oct	**4.9**	**VPD Jul**	**4.3**
NDVI Oct sd	**3.5**	VPD Aug	3.7
NDVI May sd	2.6	**NDVI Oct sd**	**2.9**
NDVI Aug sd	**2.1**	**VPD Sep**	2.6
P Aug	**1.8**	**P Aug**	2.4
NDVI Sep m	1.7	MIR Oct m	2.2
VPD Jul	**1.7**	**LC**	**2.1**
NDVI Oct m	**1.5**	NDVI Jun sd	1.8
E mean	1.5	**RH Aug**	**1.5**
P May	**1.3**	MIR May m	1.4
NDVI Jul sd	1.1	**NDVI Oct m**	**1.3**
MIR Jul sd	1	**P May**	**1.1**

Table 4.4 Correlation matrix showing *r* Pearson coefficient for the 10 predictors that obtained more than 1% contribution and permutation importance in the first preliminary run of MaxEnt.

Variables	**E sd**	**NDVI Aug sd**	**NDVI Oct mean**	**NDVI Oct sd**	**P Aug**	**P May**	**RH Aug**	**RH Jul**	**T Oct**	**VPD Jul**
E sd		-0.373	0.326	0.351	0.065	0.444	0.282	0.261	0.059	-0.232
NDVI Aug sd			-0.377	0.068	-0.045	-0.324	-0.166	-0.140	0.138	0.103
NDVI Oct mean				0.169	0.366	0.560	0.494	0.490	0.195	-0.471
NDVI Oct sd					0.235	0.279	0.258	0.245	0.119	-0.243
P Aug						0.369	0.789	0.764	0.167	-0.836
P May							0.419	0.421	0.452	-0.315
RH Aug								0.984	0.089	-0.953
RH Jul									0.108	-0.955
T Oct										0.782
VPD Jul										

Because of the high correlation values among VPD, RH and P and because of their high ranking in the contribution and permutation values three additional MaxEnt models were built using I) all variables excluding RH and T since they were used to calculate VPD; II) all variables except VPD and RH to evaluate the strength of P; and III) all variables except VPD to evaluate

the strength of RH. The results are shown in tables 4.5, 4.6 and 4.7, respectively. Because of the consistently high contribution and permutation importance of VPD of July (43.9% and 32.1% respectively) and RH of July (46.9% and 23.9% respectively), these variables were considered key predictors for *C. sonorensis* presence and used to build the final models. On the other hand, because of the high correlation of P and RH of August with VPD and RH of July, but overall lower performance, these variables were excluded from subsequent analyses.

Table 4.5 MaxEnt run without T and RH variables. Only the variables that scored more than 1% in both contribution and permutation importance are shown.

Variable	% contribution	Variable	Permutation importance
VPD Jul	**43.9**	**VPD Jul**	**32.1**
E sd	21.8	E sd	17.1
LC	8.5	NDVI Jul m	10.9
NDVI Oct sd	5.1	VPD Sep	7.1
E mean	3.5	NDVI Oct sd	4.8
NDVI Sep m	3.1	E mean	4.2
MIR Jul sd	1.8	LC	3.6
E min	1.8	NDVI Aug sd	3.4
P Aug	1.4	P Jun	2.6
P May	1.3	MIR Jul sd	2.1
MIR Sep m	1.3	P Jul	1.9
NDVI Aug sd	1.3	MIR May mean	1.9
VPD Sep	1.2	P May	1.8
NDVI Jul m	1.2	NDVI May sd	1.8
		NDVI Jun sd	1.7
		NDVI Sep m	1

Table 4.6 MaxEnt run without VPD and RH variables. Only the variables that scored more than 1% in both contribution and permutation importance are shown.

Variable	% contribution	Variable	Permutation importance
P Aug	**26.9**	T Aug	32.1
E sd	25.3	E sd	10.6
LC	13.1	NDVI Jul m	10.6
NDVI Oct sd	5.1	T Oct	10.4
MIR Sep sd	3.4	E mean	5.4
NDVI Jul m	3.4	LC	4.9
MIR Jul sd	3.1	**P Aug**	**4.7**
NDVI May sd	2.6	MIR Jul sd	4.1
P May	2.4	NDVI Oct sd	3.1
T Jun	1.9	P Sep	3.1
T Aug	1.9	P Jun	3.1
E mean	1.6	P Jul	3
NDVI Aug sd	1.5	NDVI Aug sd	1.6
MIR Oct m	1.4		
T Oct	1.3		
P Sep	1.1		
E min	1.1		

Table 4.7 MaxEnt run without VPD variables. Only the variables that scored more than 1% in both contribution and permutation importance are shown.

Variable	% contribution	Variable	Permutation importance
RH Jul	**46.9**	**RH Aug**	**35**
E sd	17.5	**RH Jul**	**23.9**
LC	8.4	E sd	7.5
NDVI Oct sd	4.8	NDVI Jul m	6.9
NDVI Jul m	4.7	T Oct	6.4
RH Aug	**2.8**	LC	4
E min	1.6	T Jun	2.4
MIR Oct m	1.5	NDVI Oct sd	2.2
NDVI Aug sd	1.4	NDVI Oct m	2.2
T Jun	1.4	MIR Jul sd	2
NDVI May sd	1.3	E mean	1.8
NDVI Sep m	1.3	MIR Oct m	1.4
T Oct	1.3	P Jun	1.3
MIR Jul sd	1.2		

Nine of the 11 variables that scored above 1% on both contribution and permutation importance in the first MaxEnt run were considered to build the final models.

Ranges of the values are described below and the respective maps are shown in the Appendix.

Elevation (sd) values ranged from 0 m to 380 m. The greatest variability in elevation was located along the Rocky Mountains, in the western part of the study area (Appendix 2a). NDVI mean values in October ranged from -0.14 to 0.84. High values, corresponding to areas with greener vegetation, were found in north-western Montana and in the eastern side of the Albertan Rockies. The icefields, in the western side, had the lowest mean NDVI values for October (Appendix 2c). NDVI standard deviation of October ranged from 0 to 0.46. Greater variability was measured in mountainous areas, with high elevation values (Appendix 2d). The variability of NDVI in August (sd) ranged from 0 to 0.48. Greater values were located in south-eastern Alberta and across the USA-Canadian border (Appendix 2e). Precipitation of May ranged from 1.26 to 2.90 $Kg/m^2/s*100000$ (i.e. 34-76 mm/month). More abundant precipitation events occurred along the Rocky Mountains in Alberta, in the Flathead area of Montana, and along the border of Idaho and Wyoming in Montana (Appendix 2f). RH of July ranged from 36% to 78%. Greater humidity rates were found along the Rocky Mountains, whereas lower values were

found in south-eastern Alberta and central Montana (Appendix 2g). Mean temperatures of October ranged from 282.76˚ K to 291.93˚ K (i.e. 9˚C - 19˚C). Lower temperatures were found in the northern parts of the study areas, whereas greater temperatures were found in southern Montana. A decreasing temperature gradient was observed from south-western Montana to the north-eastern edge of the study area (Appendix 2h). Values for VPD of July ranged from 1.13 kPa to 4.30 kPa. Lower values were found in southern Montana and greater values were located in the northern parts of the study area. The incremental gradient moved from south-east to north-west (Appendix 2i).

4.2.2 Climate data under future scenarios

Precipitation of May and VPD of July were projected using climate change scenarios for 2030`s and 2050`s. Spatial patterns of projected predictors were consistent with their reference layers in 2010, although changes in range values were observed.

RCP 2.6 predicted VPD of July values ranging from 0.62 kPa to 4.43 kPa in the 2030s and from 0.73 kPa to 4.55 kPa in the 2050s. RCP 4.5 predicted VPD of July values ranging from 0.63 kPa to 4.44 kPa in the 2030s and from 0.86 kPa to 4.52 kPa in the 2050s. RCP 8.5 predicted VPD of July values ranging from 0.68 kPa to 4.48 kPa in the 2030s and from 1.05 kPa to 4.72 kPa in the 2050s. Trends of VPD of July from 2010 to 2050s for three locations (i.e. Lake County and Judith Basin in MT and Idamay in AB) within the study area are shown in Figure 4.2. The graphs were developed to illustrate the general trend of a significant variable along the three selected pathways. In RCP 2.6 (Figure 4.2a), VPD of July followed a mitigation scenario with declining emissions forecasted by the end of the century. Therefore its values were expected to remain constant throughout the analyzed timeframe. An intermediate scenario, with

stabilization of VPD values to be reached by 2100, was depicted in RCP 4.5 (Figure 4.2b). A rising radiative forcing pathway leading to higher (i.e. > 3.5kPa) VPD values was predicted by RCP 8.5 (Figure 4.2c).

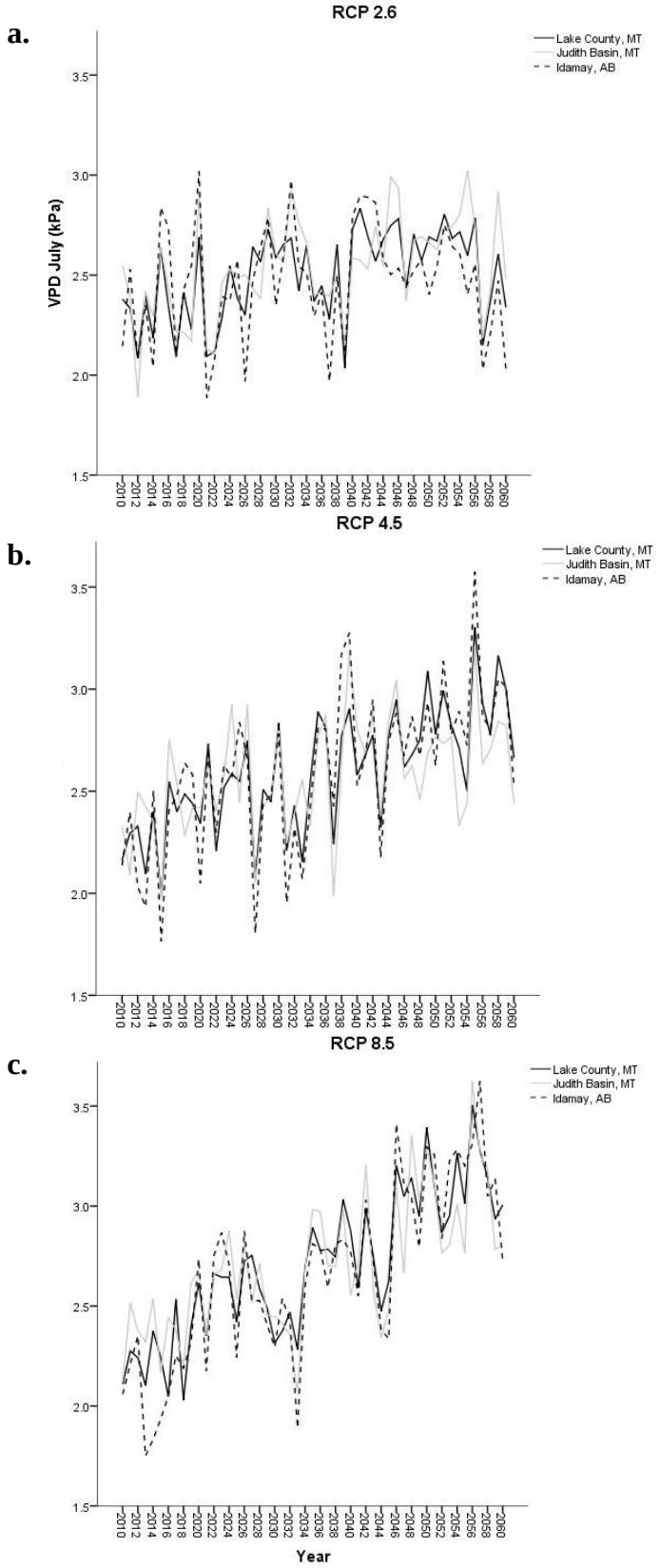


Figure 4.2 Trends for VPD of July from 2010 to 2050s for three locations (Lake County and judith Basin in MT and Idamay in AB) on the basis of Representative Concentration Pathways RCP 2.6 (a), RCP 4.5 (b), RCP 8.5 (c).

Precipitation of May followed similar trends to VPD of July over the three pathways. Mitigation RCP 2.6 predicted P of May ranging from 1.33 to 3.04 $Kg/m^2/s*100000$ in the 2030s and from 1.40 to 3.35 $Kg/m^2/s*100000$ in the 2050s. Intermediate RCP 4.5 predicted P of May ranging from 1.27 to 3.03 $Kg/m^2/s*100000$ in the 2030s and from 1.36 to 3.08 $Kg/m^2/s*100000$ in the 2050s. Extreme RCP 8.5 predicted P of May values ranging from 1.37 to 3.13 $Kg/m^2/s*100000$ in the 2030s and from 1.48 to 3.49 $Kg/m^2/s*100000$ in the 2050s.

4.3 SPATIAL MODELING AND STATISTICAL ANALYSIS

4.3.1 *Culicoides sonorensis* distribution under current conditions

Five final models were built using sub-groups of the nine selected variables on the basis of their biological importance and their cross-correlation values (Table 4.8).

Table 4.8 MaxEnt final models.

Model	Training data AUC	Testing data AUC	Variables	% contribution	Permutation importance	AICc	Δ	ω
A	0.84	0.76	RH Jul	59.7	70.2	1319.99	13.610	0.001
			E sd	22.1	12.9			
			LC	12.3	3.5			
			P May	3.6	8			
			T Oct	2.3	5.4			
B	0.83	0.74	VPD Jul	54.5	72	1306.38	0	0.994
			E sd	27.5	17.3			
			LC	14.9	2.7			
			P May	3	8			
C	0.85	0.79	RH Jul	55.1	62.6	1316.93	10.550	0.005
			E sd	20.9	11			
			LC	12.9	6.5			
			NDVI Oct m	9.1	17.1			
			T Oct	2	2.7			
D	0.85	0.76	VPD Jul	47.4	44.1	1330.37	23.989	0.000
			E sd	25.4	28.9			
			LC	13.5	5.8			
			NDVI Oct sd	5.6	7.5			
			NDVI Aug sd	4.4	8.1			
			NDVI Oct m	3.7	5.6			
E	0.82	0.76	RH Jul	62	66.3	1363.41	57.027	0.000
			E sd	23.1	19.6			
			LC	12.7	4.7			
			T Oct	2.2	9.4			

*Probability maps in Appendix 3

All models had high AUC values (≥ 0.70) in both training and testing sites. Model A (map in Appendix 3a), built with RH of July, E (sd), LC, P of May and T of October, reached an AUC of 0.84 on training sites and an AUC of 0.76 on testing sites. Model B (map in Appendix 3b), built using VPD of July, E sd, LC and P of May, had an AUC on training sites of 0.83 and an AUC of 0.74 on testing sites. Model C (map in Appendix 3c), built using E (sd), LC, RH of July, NDVI of October (m) and T of October, demonstrated the best ability in discriminating presence locations against background points in both training (AUC = 0.85) and testing sites (AUC = 0.79). Model D (map in Appendix 3d), built with VPD of July, E (sd), LC, NDVI of October (m and sd) and NDVI of May (sd), showed an AUC of 0.85 on training sites and an AUC of 0.76 on testing sites. The lowest accuracy in identifying training presence sites was shown in model E (AUC 0.82), built using RH of July, E (sd), LC and T of October. Its accuracy on testing sites reached an AUC value of 0.76 (map in Appendix 3e).

VPD and RH of July had the greatest percentage contribution in constructing *C. sonorensis* probability distributions. VPD of July contribution ranged from 47.4 % in model D to 54.5 % in model B. RH of July ranged from 55.1% in model C, to 59.7% in model A, to 62 % of model E. In each model, the second and the third variables for contribution importance were E (sd) and LC, respectively, accounting for 22.1% and 12.3% in model A, 27.5% and 14.9% in model B, 20.9% and 12.9% in model C, 25.4% and 13.5% in model D and 23.1% and 12.7% in model E. The remaining variables (P of May, T of October, NDVI of October [m and sd] and NDVI of May [sd]) had a contribution ranging from 2% to 9.1%.

When considering probability classes within each model (Table 4.9), classes one, two and three (< 60% probability of occurrence) covered 95% of the study area and 5% of the pixels were classified to be greater than 60% probability of *C. sonorensis* occurrence (probability maps in Appendix 3).

Table 4.9 Percentage of study area sorted by probability of *C. sonorensis* occurrence class using different model.

	Probability Class (%)				
Model	**1 (0-20)**	**2 (21-40)**	**3 (41-60)**	**4 (61-80)**	**5 (81-100)**
A	47.44	20.69	27.64	3.95	0.29
B	45.94	19.89	29.67	4.18	0.32
C	48.67	23.42	23.90	3.75	0.26
D	49.32	25.73	19.97	4.63	0.35
E	45.67	18.76	30.73	4.57	0.27

Because all of the models had high accuracy and probability class percentages and the distribution maps were too similar to select the best model among the final five models, a sample size corrected Information Criterion (AICc) was used for model selection. Model B had the lowest AICc and was therefore preferred to the other models. All of the other four models, compared to model B, have little (Δ_C = 10.55 and Δ_A = 13.61) or no empirical support (Δ_D = 23.989 and Δ_E = 57.027). Model B was considered the most parsimonious among the five models and this compensated for its lower accuracy. As a consequence, model B was selected as a reference *C. sonorensis* current distribution to be projected into future climate scenarios.

The model produced a current probability surface of 618,337 km^2 **(**Figure 4.2**)**. The probability map showed that areas close to the main rivers (i.e. Milk River, Missouri River and Yellowstone River) had a greater probability of suitability for *C. sonorensis*. The north-west part of the study area between Calgary and Edmonton was predicted to have lower probability (i.e. probability class one) of *C. sonorensis* presence. Low probability of occurrence was also predicted along the Rocky Mountains and the higher peaks in both Alberta and Montana.

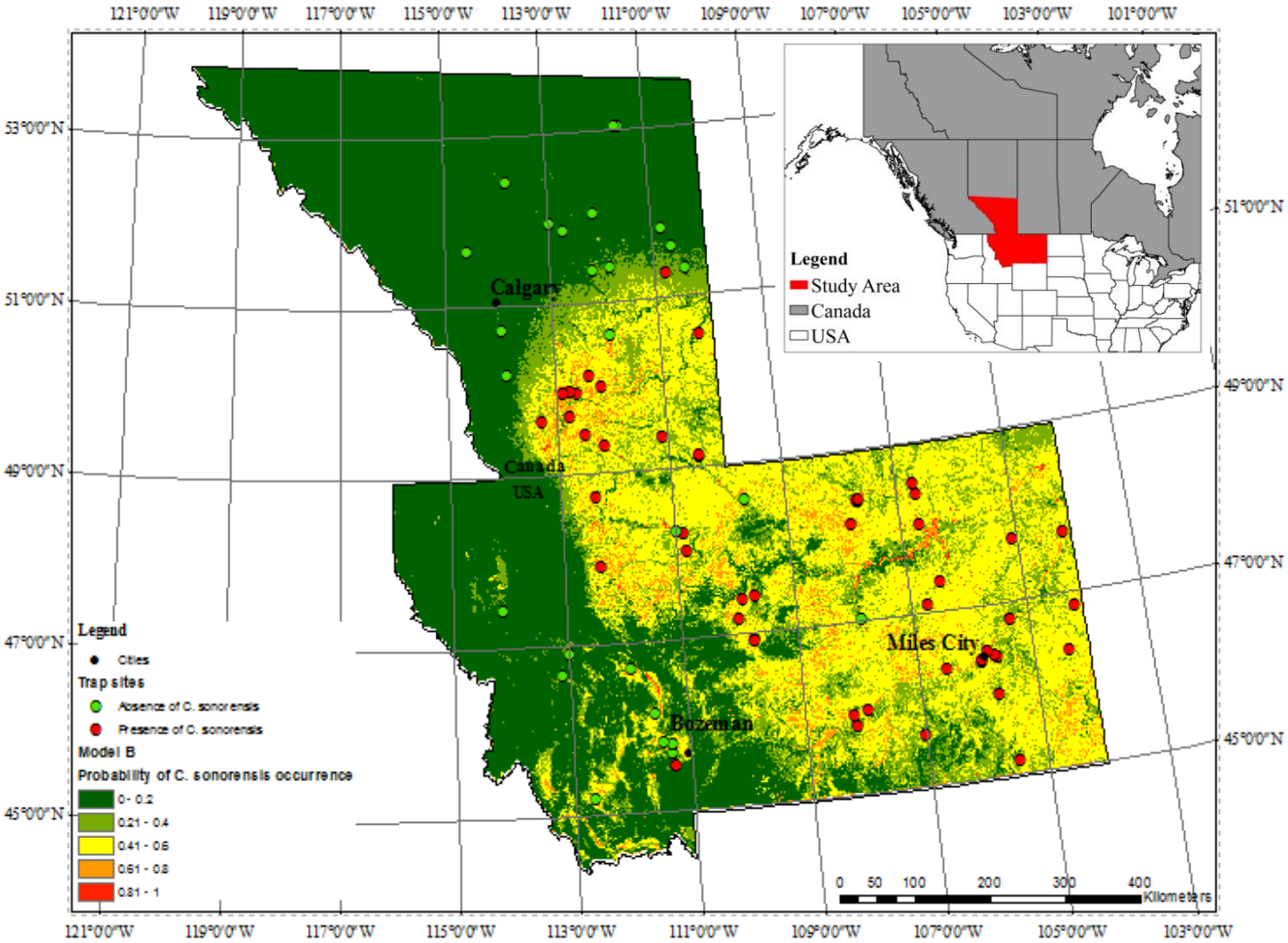


Figure 4.3 Probability distribution for *C. sonorensis* in southern Alberta and Montana. Low probability classes are represented in green (0-40%) and high probability classes in orange (61-80%) and red (81-100%). The intermediate probability class (41-60%) is represented in yellow. Red dots represent trap sites in which the vector was detected, whereas green dots represent sites where no *C. sonorensis* was found.

Variable importance was assessed with the Jackknife test which gave a total training gain of 0.71 (Figure 4.4, red bar at the bottom of the chart). The Jackknife test indicated VPD of July as the variable with the highest gain (0.39, dark blue bar) when used alone and the one decreasing the training gain (0.48, light blue bar) the most when omitted. This demonstrates that VPD of July had the most useful information not explained by other variables to accurately predict *C.*

sonorensis distribution. E (sd) had a training gain of 0.24 (dark blue bar) when used in isolation and caused a gain drop to 0.55 (light blue bar) when omitted. LC had a training gain of 0.23 (dark blue bar) when used in isolation and caused a gain drop to 0.6 (light blue bar) when omitted. P of May had a training gain of 0.19 (dark blue bar) when used alone and decreases the gain the least (0.69, light blue bar) when omitted.

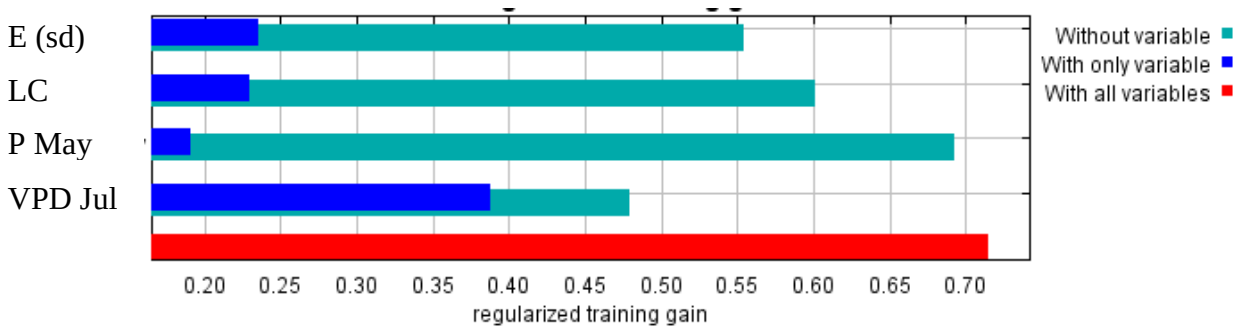


Figure 4.4 Environmental variable contribution to training gain of the final model (model B) for *C. sonorensis* occurrence. Light bars represent AUC values when the model is run without the variables. Dark bars represent AUC values when variables are used in isolation.

The response curves of the probability of *C. sonorensis* occurrence to climate and environmental variables are shown in Figure 4.5. The response curve for VPD of July showed an increased probability of *C. sonorensis* occurrence with increasing values of VPD reaching a plateau above 1.7 kPa. The highest probability of *C. sonorensis* presence (>60%) was observed with low elevation heterogeneity. When E sd exceeded 60m within the same pixel, there was a minimal predicted probability of *C. sonorensis* occurrence. Land cover classes that were positively affecting the probability of *C. sonorensis* occurrence were water bodies (class 1) and barren terrains (class 6). Open shrublands (class 3) were associated with a 0.38 probability of *C. sonorensis* occurrence. Grasslands and croplands (class 4 and 5) predicted *C. sonorensis* presence no better than random. A strong negative correlation was predicted between needle-leaf forests (class 2) and *C. sonorensis* occurrence. *C. sonorensis* was predicted to have a probability

of occurrence ranging from 0.45 to 0.60 up to 2 kg/m^2/s*100000 P of May values. Above this threshold, a steep decrease in probability of occurrence was depicted.

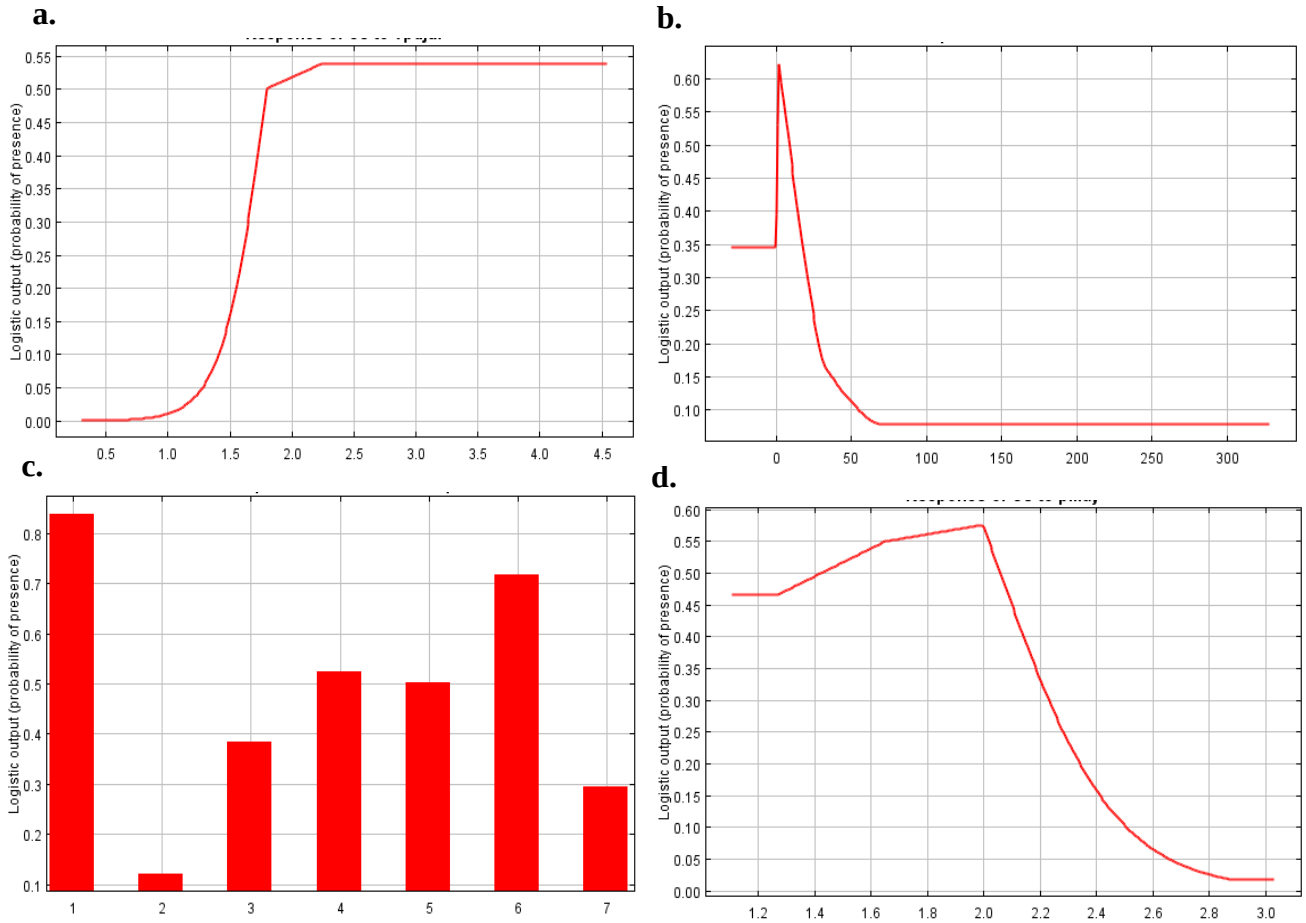


Figure 4.5 Response curves of *C. sonorensis* to top environmental and climatic variables: **a**) VPD of July; **b**) E sd; **c**) LC, and **d**) P of May. The X axis represents the variable value and the Y axis the probability of *C. sonorensis* presence.

4.3.2 *Culicoides sonorensis* distribution under future scenarios

Culicoides sonorensis distribution showed two major trends under climate change scenarios (Appendices 4-12). The 2030s scenarios showed a north-westerly increase in the probability of *C. sonorensis* occurrence compared with the current predicted distribution. The same geographical trend is observed for the 2050s scenarios with an additional increase in the areas at greater risk for *C. sonorensis* occurrence. By 2050s, both RCP 4.5 and RCP 8.5 are predicting

probability of *C. sonorensis* occurrence above the 53rd parallel north which is considered the northern limit of BT distribution worldwide. In the 2030s scenarios, 13.3 % of the pixels of RCP 2.6 recorded an increase in the probability of vector occurrence. An additional 3.9% of the study area was predicted to have increased probabilities of occurrence by 2050. RCP 4.5 forecasted 18.7% of the study area with an increased probability of *C. sonorensis* by the 2030s and 18.4% more by the 2050s. The 8.5 pathway predicts an increase of 19.7% of areas at greater risk for *C. sonorensis* occurrence by the 2030s, and a further increase of 16.3% by the 2050s.

The classes with greater probability of occurrence (i.e. class 4 and 5, Pr > 60%) gradually expanded between 2010 and the 2030s for each scenario. All pathways suggested a strong shift northwards for the high risk probability classes in Montana and north and eastwards in Alberta by the 2050s (Appendix 13). In the 2030s, class 4 and 5 together accounted for 5.9% (RCP 2.6), 5.3% (RCP 4.5) and 6.8 (RCP 8.5) of the study area, whereas in the 2050s they accounted for 6.6% (RCP 2.6), 7.7% (RCP 4.5) and 7.5 % (RCP 8.5) (Table 4.10).

An overall reduction of *C. sonorensis* probability of occurrence is expected in central Montana by the mitigation (i.e. RCP 2.6) and the extreme (i.e. RCP 8.5) pathway in the 2030s (Appendices 6 and 12). By the 2050s, the 2.6 scenario forecasted a reduction of probability of occurrence along an area of the eastern side of the Rocky Mountains between Calgary and Helena. The 8.5 scenario predicted a decreased vector presence probability in the southernmost areas of Alberta surrounding Lethbridge. No probability reduction was forecasted by the RCP 4.5 scenario (Appendix 9).

Table 4.10 Percentage of study area sorted by probability of *C. sonorensis* occurrence class by the reference model (i.e. 2010) and by climate change scenario models.

Scenario		Probability Class (%)				
		1 (0-20)	**2 (21-40)**	**3 (41-60)**	**4 (61-80)**	**5 (81-100)**
	2010	45.30	16.35	34.63	3.49	0.23
2.6	**2030**	42.08	15.75	36.29	5.62	0.26
	2050	42.32	16.89	34.22	6.33	0.24
4.5	**2030**	40.46	16.69	37.54	5.04	0.27
	2050	35.17	17.79	39.35	7.38	0.31
8.5	**2030**	40.95	16.98	35.32	6.48	0.28
	2050	36.60	20.43	35.44	7.23	0.30

DISCUSSION

5.1 ENTOMOLOGICAL DATA

In this study we have used a unique entomological dataset that represents the geographical presence of the arthropod vector *C. sonorensis* from Montana, USA to its northern limit in Southern Alberta, Canada. *Culicoides sonorensis* was detected in 50 out of 78 trap sites over 10 years of sampling (2002-2011). Altitude and latitude seemed to determine two major gradients of vector occurrence. Despite the diversity of the sampling effort, the entomological results confirmed a south to north decline in *C. sonorensis* occurrence and a second visible trend, dictated by the Rocky Mountains, producing an east to west decrease in *C. sonorensis* detectability (Figure 4.1). This pattern of presence/absence is consistent with the current knowledge on *C. sonorensis* distribution in North America (Schmidtmann, Herrero, Green, & Walton, 2011; Wirth & Morris, 1985). The disparity between the number of presence and absence locations in favour of presence sites for *C. sonorensis* occurrence is greater in Montana, where the vector was detected in 36 out of 48 trap-sites. In Montana the sampling was conducted as part of disease outbreak investigations and was therefore biased towards locations with greater probability of *C. sonorensis* presence. In Alberta, during 2002 and 2003, *C. sonorensis* was trapped to study its population dynamics and abundance. Therefore, the sampling plan was also biased towards sites with greater probability of vector presence. The addition of trap-sites in the northern part of the study area during 2009-2011 reduced the disparity between the number of presence (14) and absence (16) locations in Alberta. However, the overall disparity between presence and absence locations in the study area, and the different trapping effort between Alberta and Montana and among sites (Table 4.1 and Table 4.2) led to the choice of using a presence-only method to model *C.*

sonorensis distribution, in order to obtain the best from the available entomological information and to avoid incorrect assumptions on the true meaning of absence points.

5.2 *CULICOIDES SONORENSIS* DISTRIBUTION UNDER CURRENT CONDITIONS

Five models, A-E, were developed with MaxEnt to describe *C. sonorensis* distribution under current climate and environmental conditions. All of them had high performances therefore model B was selected as the best one using AIC_c (Table 4.8). Model B predicted the current distribution of *C. sonorensis* using the correlation between *C. sonorensis* presence data and four environmental and climatic variables (mean values of Vapour Pressure of July, mean values of Precipitation of May, mean standard deviation of Elevation and Land Cover). Statistical models such as MaxEnt, estimate the correlation between variables and species occurrence but cannot be used to infer causative relationships and underlying mechanisms (Rogers, 2006; Dormann, et al., 2012). However, correlative distribution models can provide useful insights on current distributions when they are based on sound ecological knowledge and appropriate data (Dormann, et al., 2012). The main issues are biases in entomological collections and cross-correlation between predictor variables (Murray, et al., 2010). The entomological data used in this study geographically and environmentally covered the ecological range of *C. sonorensis* in Alberta and Montana. The possible bias in trap location selection occurred mainly in the first years of the sampling activity (i.e. 2002-2003) but was corrected by sampling the whole study area during the following years as homogeneously as possible. Predictor variables were selected based on expert opinion and their correlation was assessed. MaxEnt was chosen because it is the best performing presence-only statistical method in species distribution modeling (Elith, et al., 2006) and it is able to fit complex relationships between entomological data and properly transformed predictor variables (Phillips & Dudik, 2008). Although all these points were considered in order to generate an accurate

and simple model, the relationship between the best predictors (i.e. VPD of July, P of May, LC and E sd) and *C. sonorensis* occurrence is complex and hard to interpret to its full extent (Figure 4.5).

Vapour Pressure Deficit was used to predict *Culicoides* spp. distribution in two previous studies (Baylis, Mellor, Wittmann, & Rogers, 2001; Wittmann, Mellor, & Baylis, 2001) and it is considered to describe the combined effects of temperature and relative humidity on the survival of insects (Johnson, 1942; Williams & Brochu, 1969). VPD had a significant contribution in our predictions. Increasing VPD values of July correspond to greater water losses from wet surfaces, and are correlated with *C. sonorensis* presence (Figure 4.5a). A similar relationship calculated with annual values of VPD, was found by Wittmann et al. (2001) when modeling *C. imicola* distribution in Europe. These findings might be related to evaporation rates associated with water bodies that host immature stages of *C. sonorensis* (Mullens, 1989). The higher the VPD values, the higher evaporation rates, consequently the higher the salinity concentration at breeding sites. This is considered a favourable condition for the development of *C. sonorensis* larvae (Schmidtmann, Bobian, & Belden, 2000; Schmidtmann, Herrero, Green, & Walton, 2011). Additionally, in the laboratory study undertaken by Wittmann et al. (2002) it was found that low humidity (40%) resulted in an increased survival of *C. sonorensis* adults at high temperatures (25°C) compared to the same temperature but higher humidity rates (85%). The first case corresponds to VPD values of 1.9 kPa, whereas the second corresponds to VPD values of 0.4 kPa. In our results VPD values above 1.9 kPa predicted a probability of *C. sonorensis* presence above 50%, whereas VPD values of 0.4 kPa corresponded to a minimal probability of *C. sonorensis* occurrence. Wittmann et al. (2002) also reported an increased survival of *C. sonorensis* when temperature was low (15°C), relative humidity was high (85%) and therefore VPD was 0.2 kPa, This was not consistent with our result and it might be due to the different level of complexity on which a laboratory study and large-scale species distribution model focus on.

According to our results, *C. sonorensis* presence was inversely related to the mean Precipitation of May (Figure 4.5d). The highest probability of species occurrence (60%) corresponded to mean P of May values of 2 kg/ m^2/s*100000 (= 54mm) and decreased after that threshold. These findings support previous considerations on VPD and may be explained by the fact that heavy rainfall increases water levels at the breeding sites, decreases VPD values, reduces salts concentration and water contamination. This increase in precipitation can inhibit immature *C. sonorensis* development (Mullens, 1989). Since Albertan *C. sonorensis* emerge in May (Lysyk, 2007), unfavourable conditions at this point in time can affect population dynamics for the whole season.

The standard deviation of elevation is likely associated with the combined effects of altitude and slope, providing more information than the more commonly used mean altitude variable itself (Figure 4.5b). *C. sonorensis* had lower probability of occurrence where values of E sd were greater. The areas with higher terrain variability such as slopes will retain less water, be at higher altitudes and characterized by colder temperatures resulting in their being unsuitable habitats for *C. sonorensis*.

Land cover classes such as water bodies, crop-land and grassland were positively associated with the occurrence of *C. sonorensis* (Figure 4.5c). This might describe suitable breeding sites defined by plain terrains with water resources for natural or agricultural reasons and it is consistent with current knowledge on *C. sonorensis* land cover preferences (Holbrook, Schmidtmann, McKinnon, Bobian, & Grogan, 2000). The needle-leaf forest class is distributed along the Rocky Mountains and is strongly correlated to *C. sonorensis* absence, which might be due not only to shade over *C. sonorensis* potential habitat (Mullens & Rodriguez, 1985) but also to altitudinal and climatic reasons.

When considering the current predicted distribution map obtained using MaxEnt, we observe that *C. sonorensis* has a low (<20%) probability of occurrence in most of southern Alberta. Therefore the risk of occurrence of *Culicoides*-borne disease in southern Alberta, with current climatic and environmental conditions, remains low and mainly located in its southeastern corner as previously stated (Lysyk,

2007). In Montana, areas in the proximities of water streams, lakes and dams are associated with a higher probability (i.e. above 60%) of *C. sonorensis* occurrence.

In Montana, the model outcome is supported by the occurrence of disease outbreaks (e.g. ProMED archive number: 20110820.2529, 2011; Montana Department of Livestock, 2005; Miller, et al., 2010) in the areas predicted to be at higher probability of *C. sonorensis* presence. On the other hand, the areas that experienced the most recent outbreaks of *Culicoides*-borne diseases were a primary target for vector surveillance and might have been over-sampled compared to areas considered at lower risk of disease occurrence. In Alberta, however, the areas predicted to be at low risk for *C. sonorensis* occurrence were sampled and data were presented in set AB 4 (Table 4.1). Absence of *C. sonorensis* at those trap-sites located north from Calgary was information that cannot be used in MaxEnt, as it is a presence-only method, but could validate the model forecast and strengthen its reliability.

The MaxEnt approach has not previously been used in North America, either for the generation of *C. sonorensis* distribution models and for the use of remote sensing and reanalysis products to match vector occurrence. There have been several attempts to model the distribution of other *Culicoides* species in the past decades, especially for countries in the Mediterranean basin as a result of concerns about the northern spread of BTV (Baylis, Meiswinkel, & Venter, 1999; Baylis, Mellor, Wittmann, & Rogers, 2001; Wittmann, Mellor, & Baylis, 2001; De Liberato, Purse, Goffredo, Scholl, & Scaramozzino, 2003; Tatem, et al., 2003; Purse, et al., 2004; Calvete, Estrada, Miranda, Borras, Calvo, & Lucientes, 2008). These models, together with the work done specifically on *C. sonorensis* biology (Schmidtmann, Bobian, & Belden, 2000; Gerry & Mullens, 2000; Wittmann, Mellor, & Baylis, 2002; Lysyk & Danyk, 2007; Boyer, Ward, & Singer, 2010; Mullens, 1989) have been of great value in the selection of biologically meaningful predictors for the species distribution model described in this thesis.

A sound understanding of *C. sonorensis* ecology is necessary to generate reliable distribution models and interpret the relationship between the presence of *C. sonorensis* and specific variables (Tabachnick, Mellor, & Standfast, 1992; Caporale, MacLachlan, Pearson, & Schudel, 2004). Our results can be used as a scientific source of information to identify the potential for any *Culicoides*-borne disease to occur and be sustained in the region. The vector probability distribution maps would need to be integrated with vectorial capacity parameters in order to assess disease transmission risk in the region. However, MaxEnt models, which describe *C. sonorensis* fundamental niche and therefore over-estimate the actual vector distribution, can be directly used to help inform future vector surveillance plans and can be seen as an additional precautionary strategy for disease preparedness.

5.3 *CULICOIDES SONORENSIS* DISTRIBUTION UNDER FUTURE SCENARIOS

Climate change is expected to have an impact on arthropod vectors since they are particularly sensitive to climate (Martens, Jetten, Rotmans, & Niessen, 1995; Kovats, Campbell-Lendrum, McMichael, Woodward, & Cox, 2001). Among them, *Culicoides* vectors are believed particularly responsive to global warming (Purse, Mellor, Rogers, Samuel, Mertens, & Baylis, 2005; Purse, Brown, Harrup, Mertens, & Rogers, 2008). In Europe, it was suggested that higher temperatures caused an expansion in *C. imicola* distribution and an increased importance of novel vector species (Rogers & Randolph, 2006). While the latter hypothesis has been confirmed, several studies claimed that *C. imicola* was not undergoing a range expansion (Conte, Gilbert, & Goffredo, 2009; Acevedo, et al., 2010). Purse et al. (2008) suggested that regional differences (i.e. different climatic and environmental variable interactions) might be considered when analyzing changes in *C. imicola* distribution since temperature alone might not be sufficient to draw future vector distributions.

Keeping the European experience in mind, *C. sonorensis* range projections under three different climate change scenarios were undertaken. This study was of particular interest not only because it was the first of its kind for *C. sonorensis* in North America but also because it considered a geographical area where climate changes will have more impact (Root, Price, Hall, Rosenzweig, & Pounds, 2003) and therefore responses of the species might be greater. The climate change scenarios selected are the most up-to-date climate projections produced under the IPPC mandate (van Vuuren, Edmons, Kainuma, Riahi, & et al, 2011). They also included land use data which was an important piece of missing information in previous scenarios and a key variable in our models.

Thus, *C. sonorensis* distribution projections onto the 2030s and 2050s were built using changes in VPD of July and P of May values obtained for three pathways (i.e. RCP 2.6, RCP 4.5, RCP 8.5). The projections for both time steps predicted novel parts of the study region, mainly at northern latitudes and higher latitudes, at greater probability of *C. sonorensis* occurrence. At the same time, certain areas mainly located in the center of the study region, were forecasted at a decreased probability of occurrence in RCP 2.6 and RCP 8.5, suggesting that climate change might shift northwards more than generally expand *C. sonorensis* presence in the region. This finding is supported by existing literature on climate change effects on species distribution. It is generally expected that changes in climate will produce distribution shifts more than expansions, since species have not only lower but also upper thresholds of climate tolerance (Lafferty, 2009).

Moreover, the number of areas that were predicted to be at high probability of *C. sonorensis* occurrence (>60%) increased from the baseline period to the 2030s and from the 2030s to the 2050s for each RCP. The spatial distribution of these areas followed the northward directional shift of the overall *C. sonorensis* occurrence. This forecast suggested that, independently of the scenario, climatic and environmental conditions able to sustain vector presence are predicted to move northwards, delineating new northern limits above the 53rd parallel North for *C. sonorensis* distribution. This forecast suggests

that there might be a risk of *C. sonorensis*-transmitted diseases at those latitudes. The presence of a competent vector is a conditio sine qua non for disease transmission, but it is not the only one and vectorial capacity would need to be estimated under climate change circumstances to assess disease transmission risk.

The reliability of species distribution models in future climates strongly depends on the quality of the underlying distribution model built under current climatic conditions. MaxEnt assumes that the relationship between variables and species remains the same through time, therefore it is essential that the selected predictor variables are biologically meaningful and presence points fairly represent the whole environmental range in which *C. sonorensis* exist. Even if all these precautions were undertaken, modeling species distributions under future climate is a delicate art (Elith, Kearney, & Phillips, 2010) that can generate useful but still uncertain results.

5.4 Conclusions and future directions

The OIE recently developed standards for the implementation of surveillance plans for listed and emerging animal disease vectors, in order to gather information and monitor changes on their distribution. Arthropods are key elements for pathogen transmission and accurate data are required to develop a reliable risk assessment. An understanding of the ecology of vector borne diseases, especially of the factors that determine the distribution and competence of arthropod vectors, is an important pre-requisite for the development of regional risk profiles for diseases such as BT and EHD. MaxEnt predictions, with all the known limitations of statistical models, should be considered as a reliable source of information on *C. sonorensis* distribution on un-sampled areas on the basis of its correlation to biologically relevant environmental and climatic variables on sampled locations. Moreover, the extrapolation of these correlations to future climatic scenarios represents to our knowledge, the first

attempt to quantify at a broad scale and coarse resolution the potential spread of *C. sonorensis* under climate change conditions.

The outcomes of this work indicate the potential ecological requirements of *C. sonorensis* at a regional scale and predict its northernmost distribution edge. The maps generated with MaxEnt can be used to set up future vector surveillance, targeting areas within the known limits of *C. sonorensis* geographical occurrence under current and future climatic conditions. This would reduce the costs of an extensive surveillance plan while enhancing its efficacy. Furthermore, *C. sonorensis* probability distribution maps can be used as a starting point in the assessment of disease transmission risk, given the availability of information on vectorial capacity components (e.g. vector competence and host density).

In summary, the results of this study can inform the development of *C. sonorensis* surveillance in Alberta and enhance the Province`s capability to meet OIE health certification requirements. In the case of Montana, MaxEnt models demonstrated to be able to accurately identify areas at risk for emerging and re-emerging *Culicoides*-borne diseases such as BT, EHD and VS. The developed maps can therefore support decision-making on vector surveillance and disease prevention in specific areas.

Despite the potential value of the presented *C. sonorensis* probability maps, improvements are required. Additional predictor variables might need to be considered during future investigations. One might be soil composition, which has a great impact on larval stages and might improve model predictions. This variable was not included because it was not available at the required resolution for both Alberta and Montana. Future investigations should also incorporate viruses and host data to assess disease transmission risk in current and future climatic conditions. This step would produce interesting and reliable results since vectorial capacity components are available specifically for *C. sonorensis* (Mullens, Gerry, Lysyk, & Schmidtmann, 2004; Lysyk & Danyk, 2007; Mellor, Baylis, & Mertens, 2009). Secondly, the application of other modeling methods such as logistic regression or discriminant

analysis would test the predictions made with MaxEnt and set an interesting discussion platform for the species distribution modeling community. Lastly, collection of long term entomological data would help monitoring predicted *C. sonorensis* range shifts and ensure a robust early warning system for *Culicoides*-transmitted pathogens.

To conclude, vector ecology and distribution are key aspects to consider when examining the likely occurrence and maintenance of vector-borne diseases in a particular region. The results of this work can provide the first necessary piece of information to help address international vector surveillance requirements as well as to help inform risk assessment for emerging and re-emerging *Culicoides*-transmitted pathogens in Alberta and Montana.

BIBLIOGRAPHY

Acevedo, P., Ruiz-Fons, F., Estrada, R., Marquez, A., Miranda, M., Gortazar, C., et al. (2010). A broad assessment of factors determining Culicoides imicola abundance: modeling the present and forecasting its future in climate change scenarios. *PlosOne* , 5(12), e14236. doi:10.1371/journal.pone.0014236.

Akaike, H. (1973). Information theory as an extention of the maximum likelihood principle. *Second International Symposium on Information Theory*, (pp. 267-281). Budapest.

Alberta Agriculture and Rural Development. (2008). Animal Health Act. *Alberta Regulation 209* .

Allison, A. (2010). Detection of a novel reassortant epizootic hemorrhagic disease virus (EHDV) in the USA containing RNA segments derived from both exotic (EHDV-6) and endemic (EHDV-2) serotypes. *The Journal of General Virology* , 91, 430-439.

Anderson, D. (2008). *Model Based Inference in the Life Sciences: a primer on evidence.* New York: Springer.

Anthony, S., Maan, N., Maan, S., Sutton, G., Attoui, H., & Mertens, P. (2009). Genetic and phylogenetic analysis of the non-structural proteins NS1, NS2 and NS3 of epizootic haemorrhagic disease virus (EHDV). *Virus Research* , 145 (2), 211-219.

Atchley, W. (1967). The Culicoides of New Mexico. *the University of Kansas Science Bulletin* , 46(22), 937-1022.

Attoui, H., Maan, S., Anthony, S., & Mertens, P. (2009). Bluetongue virus, other orbiviruses and other reoviruses: Their relationship and taxonomy. In P. Mellor, M. Baylis, & P. Mertens, *Bluetongue* (pp. 23-52). London: Elsevier.

Austin, M. P. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* , 157, 101-118.

Austin, M., & Cunningham, R. (1981). Observational analysis of environmental gradients. *Proceedings of the Ecological Society of Australia* , 11, 109-119.

Austin, M., & Heyligers, P. (1989). Vegetation survey design for conservation: gradset sampling of forests in North-east New South Wales. *Biological Conservation* , 50, 13-32.

Barber, T. (1979). Temporal appearance, geographical distribution and species of origin of bluetongue virus serotypes in the United States. *American Journal of Veterinary Research* , 40, 1654-1656.

Barnard, D., & Jones, R. (1980). Culicoides variipennis: Seasonal Abundance, Overwintering and Voltinism in Northeastern Colorado . *Environmental Entomology* , 9 (5), 709-712.

Baylis, M., Meiswinkel, R., & Venter, G. (1999). A preliminary attempt to use climate data and satellite imagery to model the abundance and distribution of Culicoides imicola (Diptera: Ceratopogonidae) in southern Africa. *Journal of the South African Veterinary Association* , 70, 80-89.

Baylis, M., Mellor, P., Wittmann, E., & Rogers, D. (2001). Prediction of areas around the Mediterranean at risk of bluetongue by modelling the distribution of its vector using satellite imaging. *Veterinary Record* , 149(21), 639 – 643 .

Baylis, M., O`Connell, L., & Purse, B. (2004). Modelling the distribution of Bluetongue vectors. *Veterinaria Italiana* , 40 (3), 176-181.

Berger, A., Della Pietra, S., & Della Pietra, V. (1996). A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics* , 22 (1), 39-71.

Borkent, A., & Grogan, W. (2009). Catalog of the New World Biting Midges North of Mexico. *Zootaxa* , 2273, 1-48.

Boyer, T., Ward, M., & Singer, R. (2010). Climate, Landscape, and the Risk of Orbivirus Exposure in Cattle in Illinois and Western Indiana. *American Journal of Tropical Medical and Hygiene* , 83(4), 789–794.

Braverman, J. (1994). Nematocera (Ceratopogonidae, Psychodidae, Simulidae and Culicidae) and control methods. *Revue Scientifique et Technique, International Office of Epizootics* , 13 (4), 1175-1199.

Breiman, L. (2001). Random Forests. *Machine Learning* , 45, 15-32.

Calvete, C., Estrada, R., Miranda, M., Borras, D., Calvo, H., & Lucientes, J. (2008). Modelling the distribution and spatial coincidence of bluetongue vectors Culicoides imicola and Culicoides obsoletus group throughout the Iberian peninsula. *Medical and Veterinary Entomology* , 22, 124-134.

Capela, R., Purse, B., Pena, I., Wittmann, E., Margarita, Y., Capela, M., et al. (2003). Spatial distribution of Culicoides species in Portugal in relation to the transmission of African horse sickness and bluetongue viruses. *Medical and Veterinary Entomology* , 17, 165-177.

Caporale, V., MacLachlan, N., Pearson, J., & Schudel, A. (2004). Conclusions of the third international symposium on Bluetongue. *Proceedings of the Third International Symposium on Bluetongue* , 40(4), 708-721.

CFIA. (2007). Directive DC 2007-1-2.

CFIA. (2009). Directive DC 2009-08-10.

CFIA. (2013, 01 10). *Federally Reportable Diseases for Terrestrial Animals in Canada.* Retrieved from www.inspection.gc.ca: http://www.inspection.gc.ca/animals/terrestrial-animals/diseases/reportable/2012/eng/1329499145620/1329499272021

CFIA. (2011, 05 17). *Immediately Notifiable Diseases.* Retrieved from www.inspection.gc.ca: http://www.inspection.gc.ca/animals/terrestrial-animals/diseases/immediately-notifiable/eng/1305670991321/1305671848331

CFIA. (2010). Regulations amending the Health of Animals Regulations. *Canada Gazette* , 144 (10).

CFIA. (2010). TAHD-DSAT-FADEM-2010-04-1.

Conte, A., Gilbert, M., & Goffredo, M. (2009). Eight years of entomological surveillance in Italy show no evidence of Culicoides imicola geographical range expansion. *Journal Applied Ecology* , 46, 1332-1339.

Couvillion, C., Nettles, V., Davidson, W., Pearson, J., & Gustafson, G. (1981). Hemorrhagic disease among white tailed deer in the Southeast from 1971 through 1980. *Proceedings, Annual Meeting of the US Animal Health Association* , 85, 522-537.

De Liberato, C., Farina, F., Magliano, A., Rombola`, P., Scholl, F., Spallucci, V., et al. (2010). Biotic and Abiotic Factors Influencing Distribution and Abundance of C. obsoletus Group (Diptera: Ceratopogonidae) in Central Italy. *Journal of Medical Entomology* , 47(3), 313-318.

De Liberato, C., Purse, B., Goffredo, M., Scholl, F., & Scaramozzino, P. (2003). Geographical and seasonal distribution of the bluetongue disease vector, Culicoides imicola, in central Italy. *Medical and Veterinaty Entomology* , 17, 388-394.

De Regge, N., Deblauwe, I., De Deken, R., Vantieghem, P., Madder, M., Geysen, D., et al. (2012). Detection of Schmallenberg virus in different Culicoides spp. by real-time RT-PCR . *Transboundary and Emerging Diseases* , In press.

de Vos, C., Hoek, M., Fisher, E., de Koeijer, A., & Bremmer, J. (2010). *Risk Assessment Framework for Emerging Vector-Borne Livestock Diseases.* Lelystad: Central Veterinary Institute-Wageningen UR.

Deer and Elk Farmers' Information Network,. (2003). *Farmed Deer Statistics.* Retrieved from http://www.deer-library.com/artman/publish/article_92.shtml

DeHaven, W., del Valle Molina, J., & Evans, B. (2004). Bluetongue viruses and trade issues: a North American perspective. *Veterinaria Italiana* , 40(4), 683-687.

Donner, L., Schubert, W., & Somerville, R. (2011). *The Development of Atmospheric General Circulation Models. Complexity Synthesis and Computation.* Cambridge: Cambridge University Press.

Dormann, C. F., Schymanski, S., Cabral, J., Chuine, I., Graham, C., Hartig, F., et al. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* , 39, 2119–2131.

Downes, J. (1978). The Culicoides variipennis complex: a necessary realignment of nomenclature (Diptera: Ceratopogonidae). *Canadian Entomologist* , 110: 63-6.

Dragonas, P. (1981). Evolution of bluetongue in Greece. *Office International des Epizootiques, monthly epizootic circular* , 9, 10-11.

Drolet, B., Campbell, C., M.A., S., & Wilson, W. (2005). Vector competence of Culicoides sonorensis (Diptera: Ceratopogonidae) for vesicular stomatitis virus. *J. Med. Entomol.* , 42(3), 409-418.

Dulac, G., Dubuc, C., Afshar, A., Meyers, D., Bouttard, A., Shapiro, J., et al. (1988). Consecutive outbreaks of epizootic hemorrhagic disease of deer and bluetongue. *Veterinary Record* , 122, 340.

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across space and Time. *Annual Review of Ecology, Evolution and Systematics* , 40, 677-697.

Elith, J., Graham, C. A., Dudık, M., Ferrier, S., Guisan, A., Hijmans, R., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* , 29, 129-151.

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution* , 1, 330-342.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* , 17, 43-57.

ESRI. (2011). ArcGIS Desktop: Release 10. Redlands, California.

European Food Safety Authority, A. P. (2012). Schmallenberg virus: analysis of the epidemiological data.

European Food Safety Authority, A. P. (2009). Scientific Opinion on Epizootic Hemorrhagic Disease. *EFSA Journal* , 7(12).

Everitt, B., & Dunn, G. (1991). *Applied Multivariate Data Analysis.* London: Edward Arnold.

FAO. (1998). Chapter3 - Meteorological data. In *Crop evapotranspiration - Guidelines for computing crop water requirements.* Rome: FAO.

Fielding, A., & Bell, J. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* , 24, 38-49.

Flory, A., Kumar, S., Stohlgren, T., & Cryan, P. (2012). Environmental conditions associated with bat white-nose syndrome mortality in the north-eastern United States . *Journal of Applied Ecology* , 49, 680–689.

Foster, N., Breckon, R., Luedke, A., & Jones, R. (1977). Transmission of two strains of epizootic hemorrhagic disease virus in deer by Culicoides variipennis. *Journal of Wildlife diseases* , 13 (1), 9-16.

Franklin, J. (2009). *Mapping Species Distributions.* Cambridge: University Press.

Gambles, R. (1949). Bluetongue of sheep in Cyprus. *J. Comp. Pathol.* , 176-190.

Gerry, A. C., & Mullens, B. (2000). Seasonal abundance and survivorship of Culicoides sonorensis (Diptera: Ceratopogonidae)at a southern California dairy, with reference to potential bluetongue virus transmission and persistence. *J. Med. Entomol.* , 37, 675-688.

Gillison, A. N., & Brewer, K. R. (1985). The use of gradient directed transects or gradsect in natural resources survey. *Journal of Environmental Management* , 20, 103-127.

Goller, K., Höper, D., Schirrmeier, H., Mettenleiter, T. C., & Beer. (2012). Schmallenberg Virus as Possible Ancestor of Shamonda Virus. *Emerging Infectious Diseases* , 18(10), 1644-1646.

Government of Canada. (2012, June 08). *Bovine Serological Survey 2007-2008, Executive Summary.* Retrieved from Science.gc.ca: http://www.science.gc.ca/default.asp?lang=En&xml=E8DF3735-6DDE-4908-A2DD-C39D32A22813

Grinnell, J. (1917). The niche-relationship of the California Thrasher. *The Auk* , 34, 427-433.

Grogan, W., & Phillips, R. (2008). A new species of biting midge in the subgenus Monoculicoides of Culicoides from Utah (Diptera: Ceratopogonidae). *Proceedings of the Entomological Society of Washington* , 110 (1), 196-203.

Guisan, A., & Zimmermann, N. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* , 135: 147-186.

Guisan, A., Edwards, T., & Hastie, T. (2002). General linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* , 157, 89-100.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and the use of the area under a receiver operating characteristics curve. *Radiology* , 143, 29-36.

Hanski, I. (1999). *Metapopulation ecology.* Oxford: Oxford University Press.

Hanson, R. (1952). The natural history of vesicular stomatitis. *Bacteriological Reviews* , 16, 179-204.

Hardy, W., & Price, D. (1952). Soremuzzle of sheep. *Journal of the American Veterinary Medical Association* , 120, 23.

Hay, S. (2000). An overview of remote sensing and geodesy for epidemiology and public health application. *Advances in Parasitology* , 47, 1-35.

Hernandez, P., Graham, C. H., Master, L., & Albert, D. (2006). The effect of sample size and species charateristics on performance of different species distribution modeling methods. *Ecography* , 29, 773-785.

Hirtzel, A., & Le Lay, G. (2008). Habitat suitability modeling and niche theory. *Journal of Applied Ecology* , 45, 1372-1381.

Holbrook, F. T., Schmidtmann, E., McKinnon, C., Bobian, R., & Grogan, W. (2000). Sympatry in the Culicoides variipennis Complex (Diptera: Ceratopogonidae): a taxonomic reassessment. *Journal of Medical Entomology* , 37(1), 65-76.

Homan, E., Mo, C., Thompson, L., Barreto, C., Oviedo, M., Gibbs, E., et al. (1990). Epidemiologic study of bluetongue viruses in Central america and the Caribbean: 1986-1988. *American Journal of Veteterinary Research* , 51, 1089-1094.

Hourrigan, J. L., & Klingsporn, A. (1975). Epizootiology of bluetongue: the situation in the United States of America. *Australian Veterinary Journal* , 51, 203-208.

Howerth, E., Greene, C., & Prestwood, A. (1988). Experimentally induced bluetongue virus infection in white-tailed deer: coagulation, clinical pathologic, and gross pathologic changes. *American Journal of Veterinary Research* , 49(11), 1906-1913.

Hutchinson, G. (1957). Concluding remarks. *Cold Spring Harbour Symposium of Quantitative Biology* , 22, 415-427.

IPCC. (2011, November 28). *Climate Scenario Data.* Retrieved from Data Distribution Centre: http://www.ipcc-data.org/ddc_change_field.html

IPCC. (1990). *The IPCC First Assessment Report.* Cambridge: Cambridge University Press.

IPCC. (2007). *The IPCC Fourth Assessment Report.* Cambridge: Cambridge University Press.

IPCC. (1996). *The IPCC Second Assessment Report.* Cambridge: Cambridge University Press.

IPCC. (2001). *The IPCC Third Assessment Report.* Cambridge: Cambridge University Press.

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *The Physical Review* , 106, 620-630.

Jennings, M., & Mellor, P. (1989). Culicoides: Biological Vectors of Akabane Virus. *Veterinary Microbiology* , 21, 125-131.

Jimenez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modeling. *Diversity Distrib* , 14, 885-890.

Johnson, C. (1942). Insect survival in relation to the rate of water loss. *Biological Reviews* , 17(2), 151-177.

Johnson, D. (2007). Identification of new United States bluetongue types. *Proceedings of the Annual Meeting of the US Animal Health Association* , 111, 209-210.

Jorgensen, N. (1969). The sistematics, occurence and host preference of Culicoides (Diptera:Ceratopogonidae) in Southeastern Washington. *Melanderia* , 3, 1-47.

Kadmon, R., Farber, O., & Danin, A. (2003). A systematic analysis of factors affecting the performance of climate envelope models. *Ecological Applications* , 13, 853-867.

Kilpatrick, A. M., & Randolph, S. E. (2012). Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* , 380, 1946-1955.

Kovats, R., Campbell-Lendrum, D., McMichael, A., Woodward, A., & Cox, J. S. (2001). Early effects of climate change: do they include changes in vector-borne disease? *Philosophical Transactions of the Royal Society B* , 356, 1057-1068.

Kramer, W. L., Jones, R. H., Holbrook, F. R., Walton, T. E., & Calisher, C. H. (1990). Isolation of Arboviruses from Culicoides Midges (Diptera: Ceratopogonidae) in Colorado During an Epizootic of Vesicular Stomatitis New Jersey. *Journal of Medical Entomology, Volume 27, Number 4, July 1990 , pp. 487-493(7)* , 27(4), 487-493.

Lafferty, K. D. (2009). The ecology of climate change and infectious diseases. *Ecology* , 90(4), 888-900.

Levin, S. (1992). The Problem of Pattern and Scale in Ecology. *Ecology* , 73(6), 1943-1967.

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

Lillie, T., Marquart, W., & Jones, R. (1981). The flight range of Culicoides variipennis (Diptera: Ceratopogonidae). *Canadian Entomologist* , 113, 419-426.

Lysyk, T. (2006). Abundance and species composition of Culicoides (Diptera: Ceratopogonidae) at cattle facilities in southern Alberta, Canada. *Journal of Medical Entomology* , 43(5), 840-849.

Lysyk, T. (2007). Seasonal abundance, Parity and Survival of Adult Culicoides sonorensis (Diptera: Ceratopogonidae) in Southern Alberta, Canada. *Journal of Medical Entomology* , 44 (6), 959-969.

Lysyk, T., & Danyk, T. (2007). Effect of Temperature on Life History Parameters of Adult Culicoides sonorensis (Diptera: Ceratopogonidae) in Relation to Geographic Origin and Vectorial Capacity for Bluetongue Virus. *Journal of Medical Entomology* , 44(5), 741-751.

Maan, S., Maan, S., Belaganahalli, M. O., Johnson, D., Nomikou, N., & Mertens, P. (2012). Identification and Differentiation of the Twenty Six Bluetongue Virus Serotypes by RT–PCR Amplification of the Serotype-Specific Genome Segment 2. *PLoS ONE* , 7(2).

Maclachlan, N. (2010). Global Implications of the Recent Emergence of Bluetongue Virus in Europe. *Veterinary Clinics of North America: Food Animal Practice* , 26, 163-171.

MacLachlan, N., & Guthrie, A. (2010). Re-emergence of bluetongue, African horse sickness and other Orbivirus diseases. *Veterinary Research* , 41: 35-46.

Magori, K., & Drake, J. (2013). The population dynamics of Vector borne-Diseases. *Nature Education Knowledge* , 4 (4), 14.

Manso-Ribiero, J., Rosa-Azevedo, J., Noronha, F., Branco-Forte-Junior, M., Grave-Periera, C., & Vasco-Fernandes, M. (1957). Fievre catarrhale du mouton (blue-tongue). *Bulletin, Office International des Epizooties* , 48, 350-367.

Martens, W. M., Jetten, T., Rotmans, J., & Niessen, L. (1995). Climate change and vector-borne diseases: a global modeling perspective. *Global Environmental Change* , 5(3), 195-209.

McKercher, D., McGowan, B., Howarth, J., & Saito, J. (1953). A preliminary report on the isolation and identification of the bluetongue virusfrom sheep in California. *Journal of the American Veterinary Medical Association* , 122, 300-301.

McMullen, R. (1978). Culicoides (Diptera: Ceratopogonidae) of the south Okanagan area of British Columbia. *Canadian Entomologist* , 110, 1053-1057.

Mecham, J., & Johnson, D. (2005). Persistence of bluetongue virus serotype 2 (BTV-2) in the southeast United States. *Virus Research* , 113, 116-122.

Mellor, P., Baylis, M., & Mertens, P. (2009). *Bluetongue.* London: Elsevier Academic Press.

Mellor, P., Boorman, J., & Baylis, M. (2000). Culicoides Biting Midges: Their Role as Arbovirus Vectors. *Annual Review of Entomology* , 45: 307-340.

Meyer, C., & Thuiller, W. (2006). Accuracy of resource selection functions across spatial scales. *Diversity and Distribution* , 12, 288-297.

Miller, M. M., Brown, J., Cornish, T., Johnson, G., Mecham, J. O., Reeves, W. K., et al. (2010). Investigation of a bluetongue disease epizootic caused by bluetongue virus serotype 17 in sheep in Wyoming. *Journal of the American Veterinary Medical Association* , 237 (8), 955-959.

Montana Department of Livestock. (2005). *MT Vesicular Stomatitis Situation Report.*

Moss, R., Edmonds, J., Hibbard, K., Manning, M., Rose, S., van Vuuren, D., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature* , 463, 747–756.

Mullen, G. (2009). Biting Midges (Ceratopogonidae). In G. Mullen, & D. L.A., *Medical and Veterinary Entomology* (pp. 169-188). London: Academic Press.

Mullens, B. (1989). A Quantitative Survey of Culicoides variipennis (Diptera: Ceratopogonidae) in Dairy Waste water Ponds in Southern California. *Journal of Medical Entomology* , 26 (6); 559-565.

Mullens, B. (1989). A Quantitative Survey of Culicoides variipennis (Diptera: Ceratopogonidae) in Dairy Wastewater Ponds in Southern California. *Journal of Medical Entomology* , 26(6), 559 - 565.

Mullens, B., & Lii, K. (1987). Larval Population Dynamics of Culicoides variipennis (Diptera: Ceratopogonidae) in Southern California. *Journal of Medical Entomology* , 24 (5), 566-573.

Mullens, B., & Holbrook, F. (1991). Temperature effects on the gonotrophic cycle of Culicoides variipennis (Diptera: Ceratopogonidae). *Journal of the American Mosquito Control Association* , 7 (4), 588-591.

Mullens, B., & Rodriguez, J. (1985). Effect of experimental habitat shading on the distribution of Culicoides variipennis (Diptera: Ceratopogonidae) larvae. *Environmental Entomology* , 14, 749-754.

Mullens, B., Gerry, A., Lysyk, T., & Schmidtmann, E. (2004). Environmental effects on vector competence and virogenesis of bluetongue virus in culicoides: interpreting laboratory data in a field context. *Veterinaria Italiana* , 40(3), 160-166.

Murray, K., Retallick, R., Puschendorf, R., Skerratt, L., Rosauer, D., McCallum, H., et al. (2010). Issues with modeling the current and future distribution of invasive pathogens. *Journal of Applied Ecology* , 1-4.

Nakicenovic, N., Swart, R., & al, e. (2000). *Special Report on Emission Scenarios (SRES).* Cambridge: Cambridge University Press.

NASA Land Processes Distributed Active Archive Center, L. D. (2001). Sioux Falls, South Dakota.

Nettles, V., Hylton, S. A., Stallknecht, D., & Davidson, W. (1992). Epidemiology of Epizootic Hemorrhagic Disease viruses in wildlife in the USA . *Bluetongue, African horse sickness, and related orbiviruses: Proceedings of the Second International Symposium* , 238-248.

Nunamaker, R., Peréz De León, A., Campbell, C., & Lonning, S. (2000). Oral infection of Culicoides sonorensis (Diptera: Ceratopogonidae) by vesicular stomatitis virus. *Journal of Medical Entomology* , 37(5),784-6.

OIE. (2011). Bluetongue. *Terrestrial Animal Health Code* .

OIE. (1968). International Zoo-Sanitary Code. 1st ED.

OIE. (2013, February). *OIE Technical Factsheet, Schmallemberg Virus.* Retrieved from http://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/A_Schmallenberg_virus.pdf

OIE. (2012). Risk Analysis. *Terrestrial Animal Health Code* , Section 2.

OIE. (2012). Surveillance for arthropods vectors of animal diseases. In *Terrestrial Animal Health Code* (p. Chapter 1.5). Paris: OIE.

OIE. (2009). *Terrestrial Animal Health Code.* Paris: OIE.

Olden, J. D., Lawler, J. J., & Poff, N. (2008). Machine Learning Methods Without Tears: A primer for ecologists. *Quarterly Review of Biology* , 83, 171-193.

Omori, T., Inaba, Y., Morimoto, T., Tanaka, Y., & Ishitani, R. (1969). Ibaraki virus, an agent of epizootic disease of cattle resembling bluetongue. I. Epidemiologic, clinical and pathologic observations and experimental transmission to calves. *Japanese Journal of Microbiology* , 13 (20), 139-157.

Pedgley, D. E. (1983). Windborne Spread of Insect-Transmitted Diseases of Animals and Man. *Philosophical Transactions of the Royal Society of London, B* , 302, 463-470.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). *Updated world map of the Köppen-Geiger climate classification.* 11: 1633-1644: Hydrology and Earth System Sciences.

Perez de Leon, A., & Tabachnick, W. (2006). Transmission of vesicular stomatitis New Jersey virus to cattle by the biting midge Culicoides sonorensis (Diptera: Ceratopogonidae). *Journal of Medical Entomology* , 43(2),323-329.

Peters, J., De Baets, B., Van doninck, J., Calvete, C., Lucientes, J., De Clercq, E., et al. (2011). Absence reduction in entomological surveillance data to improve niche-based distribution models for Culicoides imicola. *Preventive Veterinary Medicine* , 100 (1), 15-28.

Pfeiffer, R. (1960). Dynamics of climate: the Proceedings of the Conference on the Application of Numerical IntegrationTechniques to the Problem of the General Circulation. Pergamon Press.

Phillips, S., & Dudik, M. (2008). Modeling species distributions with MaxEnt: new extentions and a comprehensive evaluation. *Ecography* , 31, 161-175.

Phillips, S., Anderson, R., & Schapire, R. (2006). Maximum entropy modeling of species geographic distribution. *Ecological Modeling* , 190, 231-259.

ProMED archive number: 20110820.2529. (2011, August 20). Retrieved from ProMED-mail: http://www.promedmail.org/direct.php?id=20110820.2529

ProMED archive number: 20121010.1334173. (2012, October 05). Retrieved from ProMED-mail: http://www.promedmail.org/direct.php?id=20121010.1334173

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters* , 3, 349-361.

Pulliam, H. R. (1988). Source, sinks, and population regulations. *The American Naturalist* , 132, 652-661.

Purse, B., Brown, H. E., Harrup, L., Mertens, P., & Rogers, D. (2008). Invasion of bluetongue and other orbivirus infections into Europe: the role of biological and climatic processes. *Revue scientifique et technique, International Office of Epizootics* , 27(2), 427-442.

Purse, B., Tatem, A., Caracappa, S., Rogers, D., Mellor, P. S., Baylis, M., et al. (2004). Modelling the distributions of Culicoides bluetongue virus vectors in Sicily in relation to satellite-derived climate variables. *Medical and Veterinary Entomology* , 18 (2), 90-101.

Purse, V. B., Mellor, P., Rogers, D., Samuel, A., Mertens, P. C., & Baylis, M. (2005). Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology* , 3, 171-181.

Quenouille, M. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society Series B* , 11, 68-84.

Rainwater-Lovett, K., Pauszek, S., Kelley, W., & Rodriguez, L. (2007). Molecular epidemiology of vesicular stomatitis New Jersey virus from the 2004-2005 US outbreak indicates a common origin with mexican strains . *Journal of General Virology* , 88, 2042-2051.

Reisen, W. K. (1989). Estimation of vectorial capacity: introduction. *Bulletion of the Society for Vector Ecology* , 14, 39-40.

Rodda, G., Jarnevich, C., & Reed, R. (2011). Challenges in Identifying Sites Climatically Matched to the Native Ranges of animal Invaders. *PLoS ONE* , 6(11), 1-18.

Rogers, D. (2006). Models for Vectors and Vector-Borne Diseases. *Advances in Parasitology* , 62: 1-35.

Rogers, D., & Randolph, S. (2006). Climate change and vector-borne diseases. *Advances in Parasitology* , 62, 345-381.

Rolston, M., & Johnson, G. (2012, July). Biology and Impact of Biting Midges in Montana. *Montana State University Extension, MontGuide* , MT201205AG New 07/12.

Root, T., Price, J., Hall, K. S., Rosenzweig, C., & Pounds, J. (2003). Fingerprints of global warming on wild animals and plants. *Nature* , 42, 57-60.

Rushton, S. P., Ormerod, S. J., & Kerby, G. (2004). New paradigms for modelling species` distributions? *Journal of Applied Ecology* , 41, 193-200.

Rykiel, E. J. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling* , 90, 229-244.

Savini, G., Afonso, A., Mellor, P., Aradaib, I., Yadin, H., Sanaa, M., et al. (2011). Epizootic heamorragic disease. *Research in Veterinary Science* , 91, 1-17.

Schmidtmann, E. T., Mullens, B., Schwager, S., & Spear, S. (1983). Distribution, abundance and a probability model for larval Culicoides variipennis (Diptera: Ceratopogonidae) on dairy farms in New York state. *Environmental Entomology* , 12 (3), 768 - 773.

Schmidtmann, E., Bobian, R., & Belden, R. (2000). Soil chemistries deÞne aquatic habitats with immature populations of the Culicoides variipennis complex (Diptera: Ceratopogonidae). *Journal of Medical Entomology* , 37, 58-64.

Schmidtmann, E., Herrero, M., Green, A. D., & Walton, T. (2011). Distribution of Culicoides sonorensis (Diptera: Ceratopogonidae) in Nebraska, South Dakota, and North Dakota: Clarifying theEpidemiology of Bluetongue Disease in the Northern Great Plains Region of the United States. *Journal of Medical Entomology* , 48(3), 634 - 643.

Schmidtmann, E., Tabachnick, W., Hunt, G., Thompson, L., & Hurd, H. (1999). 1995 Epizootic of Vesicular Stomatitis (New Jersey Serotype) in the Western United States: an Entomologic Perspective. *Journal of Medical Entomology* , 36(1),1-7.

Shannon, C. E. (1948). The Mathematical Theory of Communication. *Bell System Technical Journal* , 27, 379-423, 623-656.

Shope, R., MacNamara, L., & Mangold, R. (1955). Report on deer mortaity, epizootic hemorrhagic disease of deer. *New Jersey Outdoors* , 6(5), 16-21.

Soberon, J., & Peterson, A. T. (2005). Interpretaion of models of fundamental ecological niches and species` distributional areas. *Biodiversity Informatics* , 2, 1-10.

Spreull, J. (1905). Malarial catarrhal fever (bluetongue) of sheep in South Africa. *Journal of Comparative Pathology* , 18, 321-337.

Stallknecht, D., Blue, J., Rollor, E., Nettles, V., Davidson, W., & Pearson, J. (1991). Precipitating antbodies to epizootic hemorrhagic disease and bluetongue viruses in white-tailed deer in South-eastern United States. *Journal of Wildlife Diseases* , 27, 238-247.

Statistics Canada. (2010). CATSNET Analytics.

Sterritt, W., & Dulac, G. (1992). Evolving perceptions of bluetongue: a challenge for government and industry. *Canadian Veterinary Journal* , 33(2): 109-111.

Tabachnick, B., & Fidell, L. (2013). Discriminant Analysis. In *Using Multivariate Statistics, 6th ed.* (p. Chapter 9). Upper Saddle River, NJ: Perason Education.

Tabachnick, W. (2010). Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *The Journal of Experimental Biology* , 213: 946-954.

Tabachnick, W. (2004). Culicoides and the global epidemiology of bluetongue virus infection. *Veterinaria Italiana* , 40, 145-150.

Tabachnick, W. (1992). Genetic differentiation among populations of Culicoides variipennis (Diptera:Ceratopogonidae), the North American vector of Bluetongue Virus. *Annals of the Entomological Society of America* , 85(2), 140-147.

Tabachnick, W., Mellor, P., & Standfast, H. (1992). Working team report on vectors: recommandations for research on Culicoides vector biology. *Bluetongue, African Horse Sickness, and Related Orbiviruses: Proceedings of the Second International Symposium* , 977-981.

Tatem, A., Baylis, M., Mellor, P., Purse, B., Capela, R., Pena, I., et al. (2003). Prediction of bluetongue vector distribution in Europe and North Africa using satellite imagery. *Veterinary Microbiology*, 97 (1-2), 13-29.

Temizel, E., Yesilbag, K., Batten, C., Senturk, S., Maan, N., Mertens, P., et al. (2009). Epizootic hemorrhagic disease in cattle, Western Turkey. *Emerging Infectious Diseases Journal* , 15, 317-319.

The MathWorks Inc. (2012). MATLAB R2012b, 8.0.

Thomas, F., Skinner, D., & Samagh, B. (1982). Evidence for bluetongue virus in Canada: 1976-1979. *Canadian Journal of Comparative Medicine* , 46(4), 350-353.

Tilman, D. (1982). *Resource Competition and Community Structure.* Princeton, New Jersey: Princeton University Press.

Tukey, J. (1958). Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics* , 29, 614.

van Vuuren, D., Edmons, J., Kainuma, M., Riahi, K., & et al. (2011). The Representative Concentration Pathways: an overwiev. *Climatic Change* , 109, 5 - 31 .

Veronesi, E., Henstock, M., Gubbins, S., Batten, C., Manley, R., Barber, J., et al. (2013). Implicating Culicoides Biting Midges as Vectors of Schmallenberg Virus Using Semi-Quantitative RT-PCR. *PLoS ONE* , 8(3).

Warren, D., & Seifert, S. (2011). Environmental niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* , 21(2), 335-342.

Williams, G., & Brochu, J. (1969). Vapour pressure deficit vs. relative humidity for expressing atmospheric moisture content. *Naturaliste can.* , 96, 621-636.

Wirth, W., & Jones, R. (1957). The North American subspecies of Culicoides variipennis (Diptera: Heleidae). *U.S. Department of Agriculture Technical Bulletin* , 1170.

Wirth, W., & Morris, C. (1985). The Taxonomic Complex, Culicoides Variipennis. In T. Barber, & M. Jochim, *Bluetongue and related orbiviruses* (pp. 165-175). New York: Wiley Liss.

Wittmann, E. J., Mellor, P. S., & Baylis, M. (2002). Effect of temperature on the transmission of orbiviruses by the biting midge, Culicoides sonorensis. *Medical and Veterinary Entomology* , 16, 147-156.

Wittmann, E., Mellor, P., & Baylis, M. (2001). Using climate data to map the potential distribution of Culicoides imicola (Diptera:Ceratopogonidae) in Europe. *Revue scientifique et technique, International Office of Epizootics* , 20 (3), 731-740.

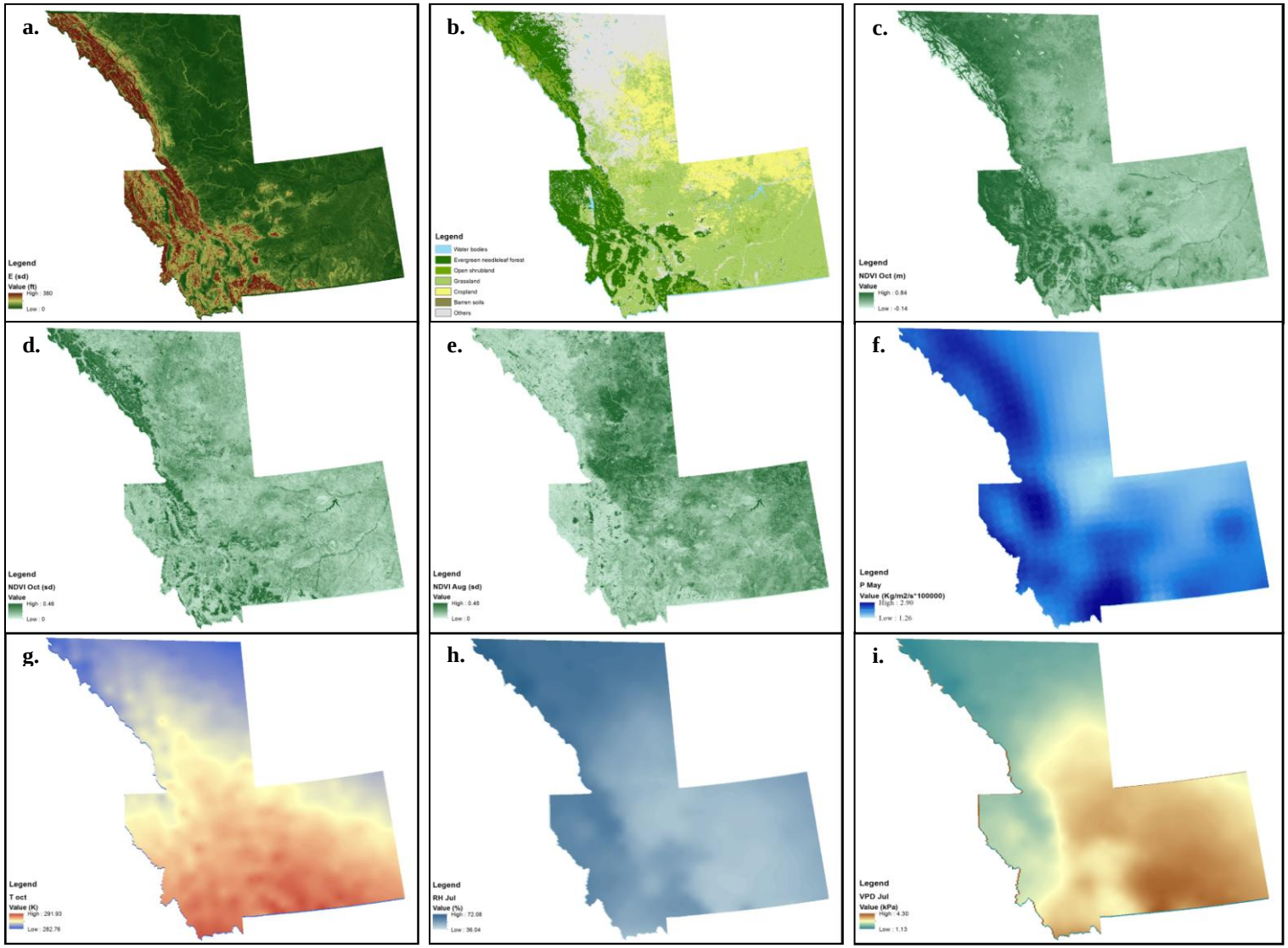
Yadin, H., Brenner, J., Bumbrov, V., Oved, Z., Stram, Y., Klement, E., et al. (2008). Epizootic haemorrhagic disease virus type 7 infection in cattle in Israel. *Veterinary Record* , 162, 53-56.

Yee, T. W., & Mitchell, N. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science* , 2, 587-602.

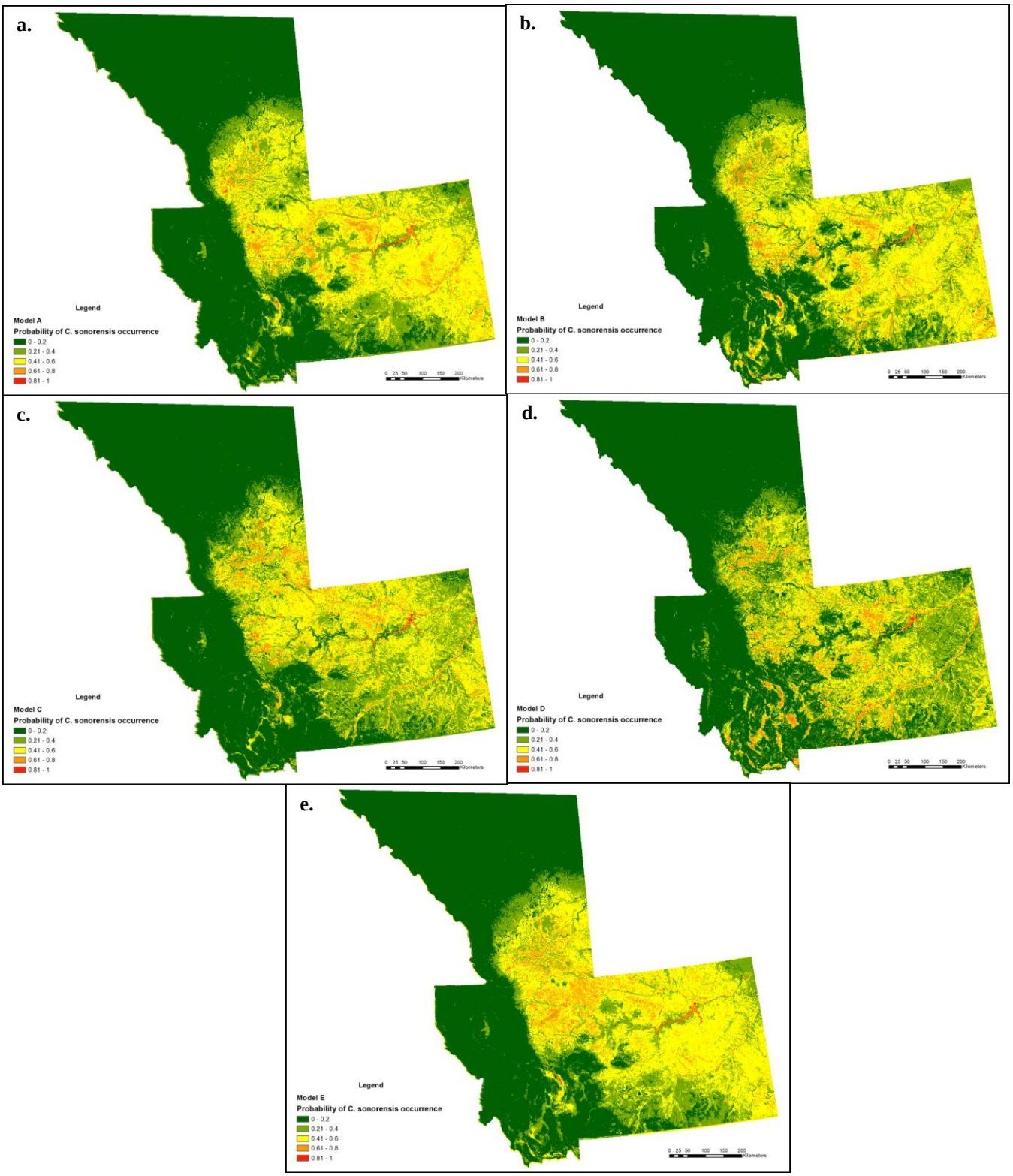
APPENDICES

Appendix 1 Land Cover Type 1, classification scheme of the International Geosphere Biosphere Programme (IGBP) and classification scheme used in the present study

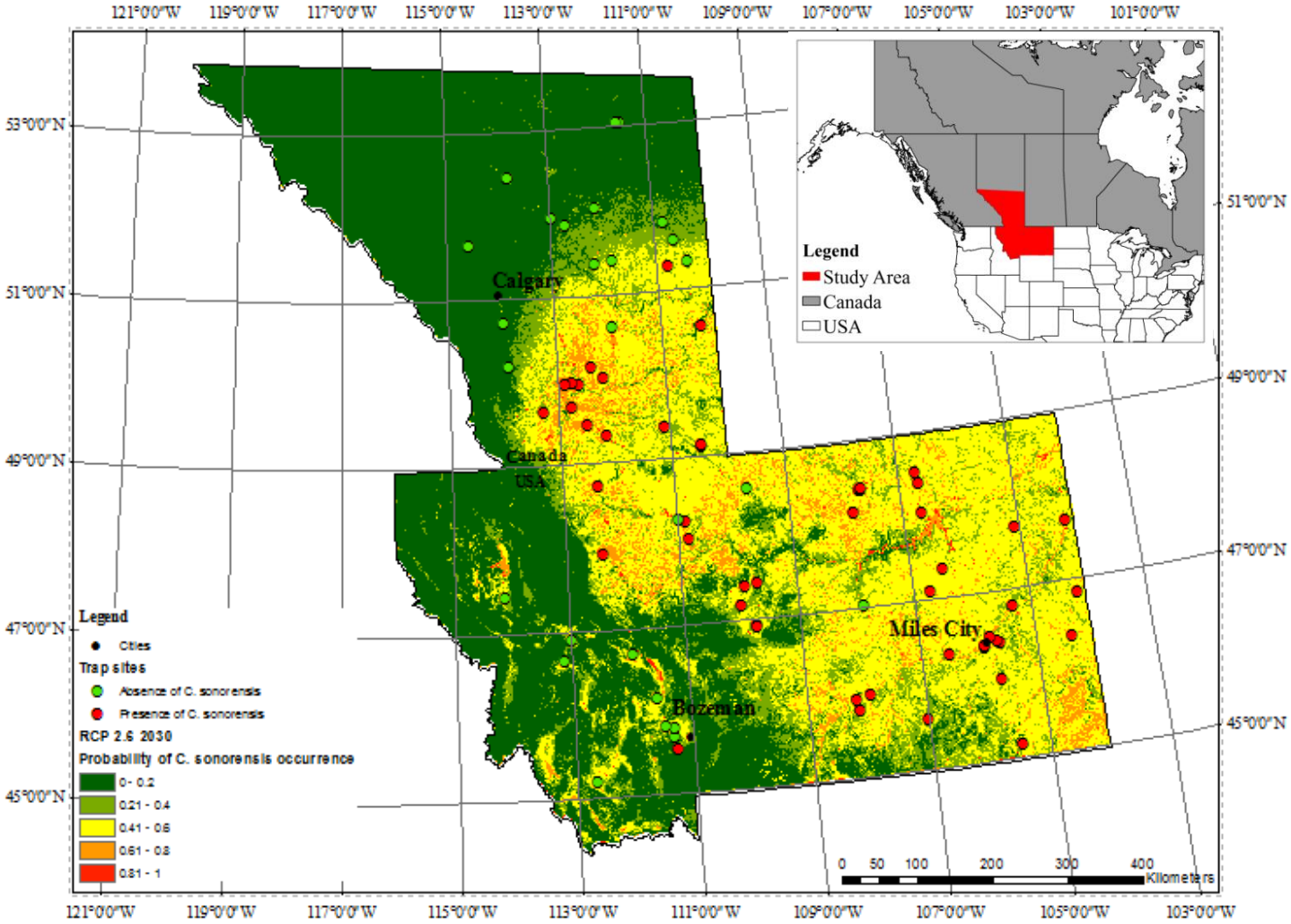
Class	IGBP (Type 1)	Description	Land Cover classification for the present study
0	Water Bodies	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies.	Water Bodies (class 1)
1	Evergreen Needleleaf forest	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.	Evergreen Needleleaf forest (class 2)
2	Evergreen Broadleaf forest	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Almost all trees remain green all year. Canopy is never without green foliage.	Other (class 7)
3	Deciduous Needleleaf forest	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	Other (class 7)
4	Deciduous Broadleaf forest	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	Other (class 7)
5	Mixed forest	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 meters. Consists of tree communities with interspersed mixtures or mosaics of the other four forest cover types. None of the forest types exceeds 60% of landscape.	Other (class 7)
6	Closed shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is >60%. The shrub foliage can be either evergreen or deciduous.	Other (class 7)
7	Open shrublands	Lands with woody vegetation less than 2 meters tall and with shrub canopy cover is between 10-60%. The shrub foliage can be either evergreen or deciduous.	Shrublands (class 3)
8	Woody savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 30-60%.The forest cover height exceeds 2 meters.	Other (class 7)
9	Savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 10-30%.The forest cover height exceeds 2 meters.	Other (class 7)
10	Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%.	Grasslands (class 4)
11	Permanent wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water.	Other (class 7)
12	Croplands	Lands covered with temporary crops followed by harvest and a bare soil period	Croplands (class 5)
13	Urban and built-up	Land covered by buildings and other man-made structures.	Other (class 7)
14	Cropland/Natural vegetation mosaic	Lands with a mosaic of croplands, forest, shrublands, and grasslands in which no one component comprises more than 60% of the landscape.	Other (class 7)
15	Snow and ice	Lands under snow and/or ice cover throughout the year.	Other (class 7)
16	Barren or sparsely vegetated	Lands exposed soil, sand, rocks, or snow and never has more than 10% vegetated cover during any time of the year.	Barren or sparsely vegetated (class 6)
254	Unclassified		
255	Fill Value		



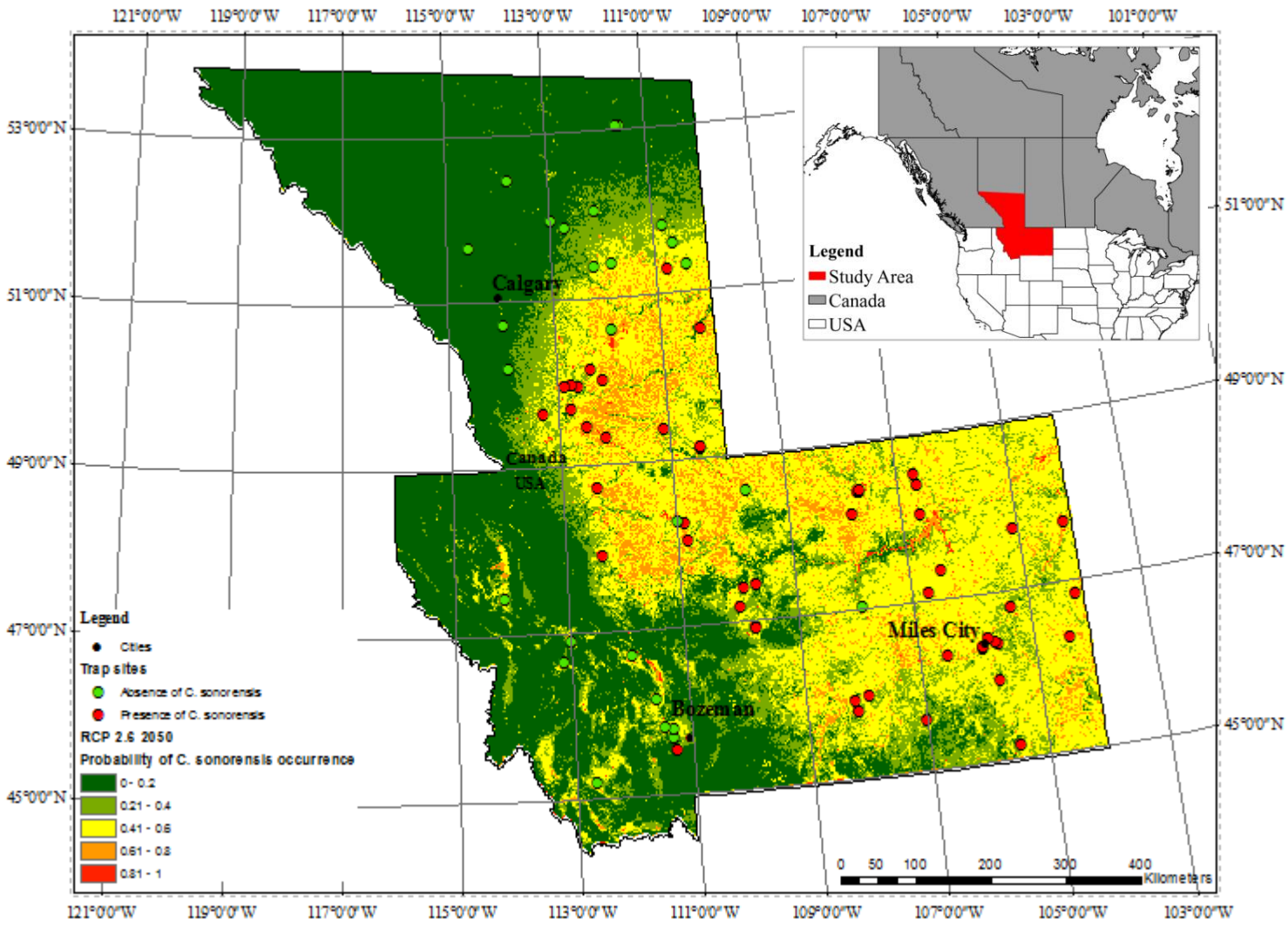
Appendix 2 Best predictors for *C. sonorensis* distribution modeling shown as geographic layers: a) E (sd); b) LC; c) NDVI Oct (m); d) NDVI Oct (sd); e) NDVI Aug (sd); f) P May; g) T Oct; h) RH Jul; and i) VPD Jul.



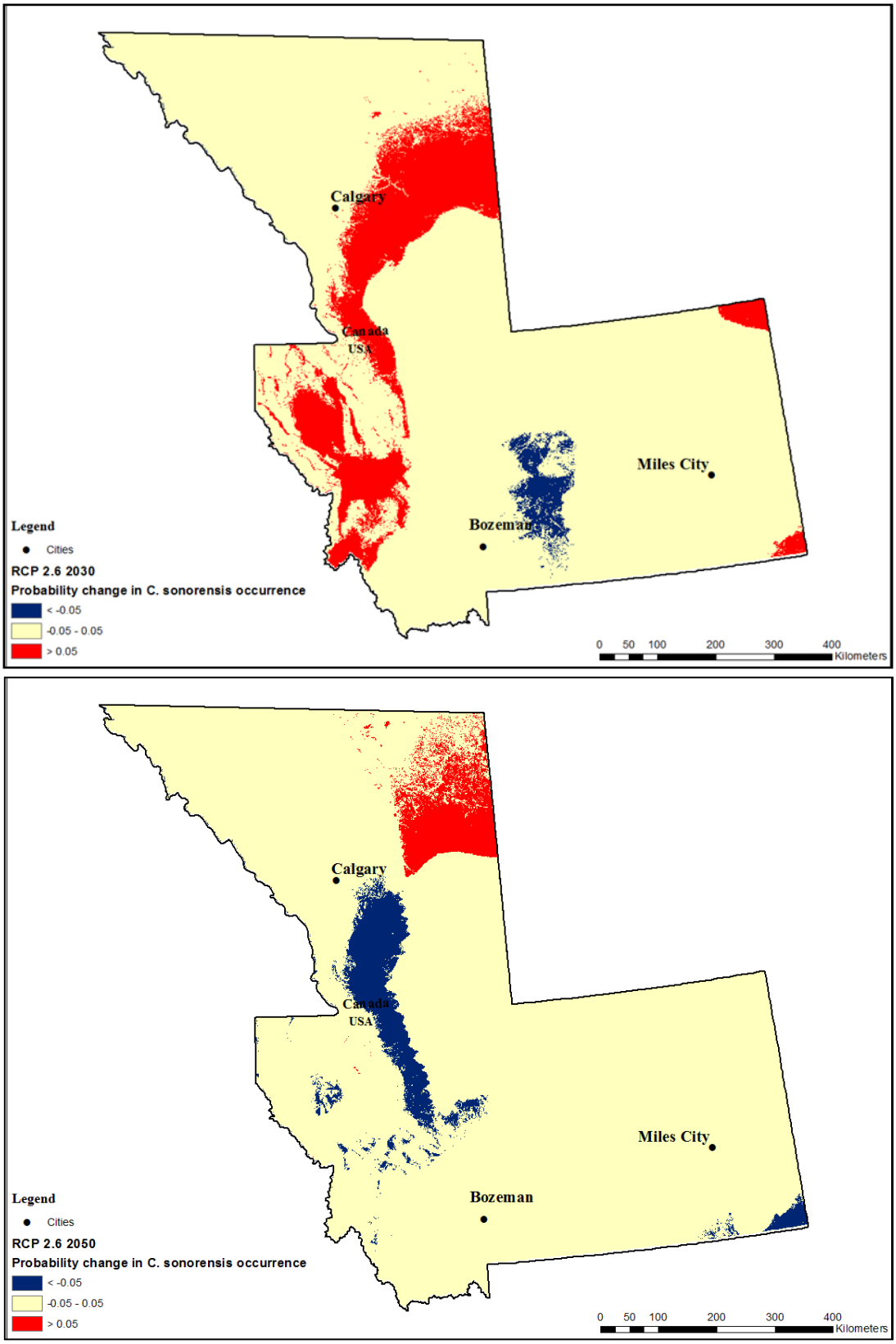
Appendix 3 Probability distribution maps for *C. sonorensis* developed by **a)** model A using RH Jul, E sd, LC, P May and T Oct; **b)** model B using VPD Jul, E sd, LC and P May; **c)** model C using RH Jul, E sd, LC, NDVI Oct m and T Oct ; **d)** model D using VPD Jul, E sd, LC, NDVI Oct sd, NDVI May sd, NDVI Oct m and **e)** model E using RH Jul, E sd, LC and T Oct.



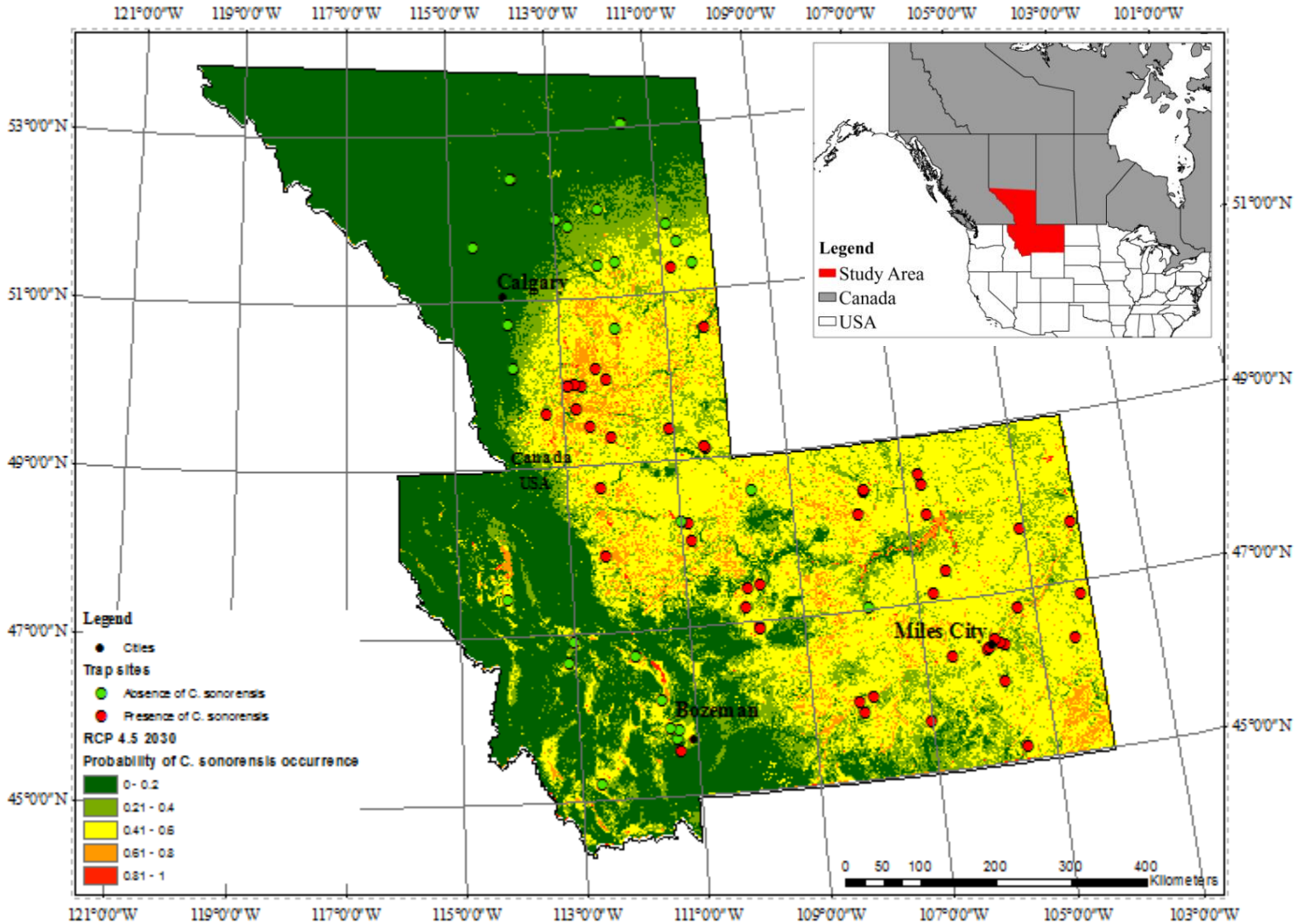
Appendix 4 Probability distribution map for *C. sonorensis* under RCP 2.6 scenario in 2030s (2020-2040).



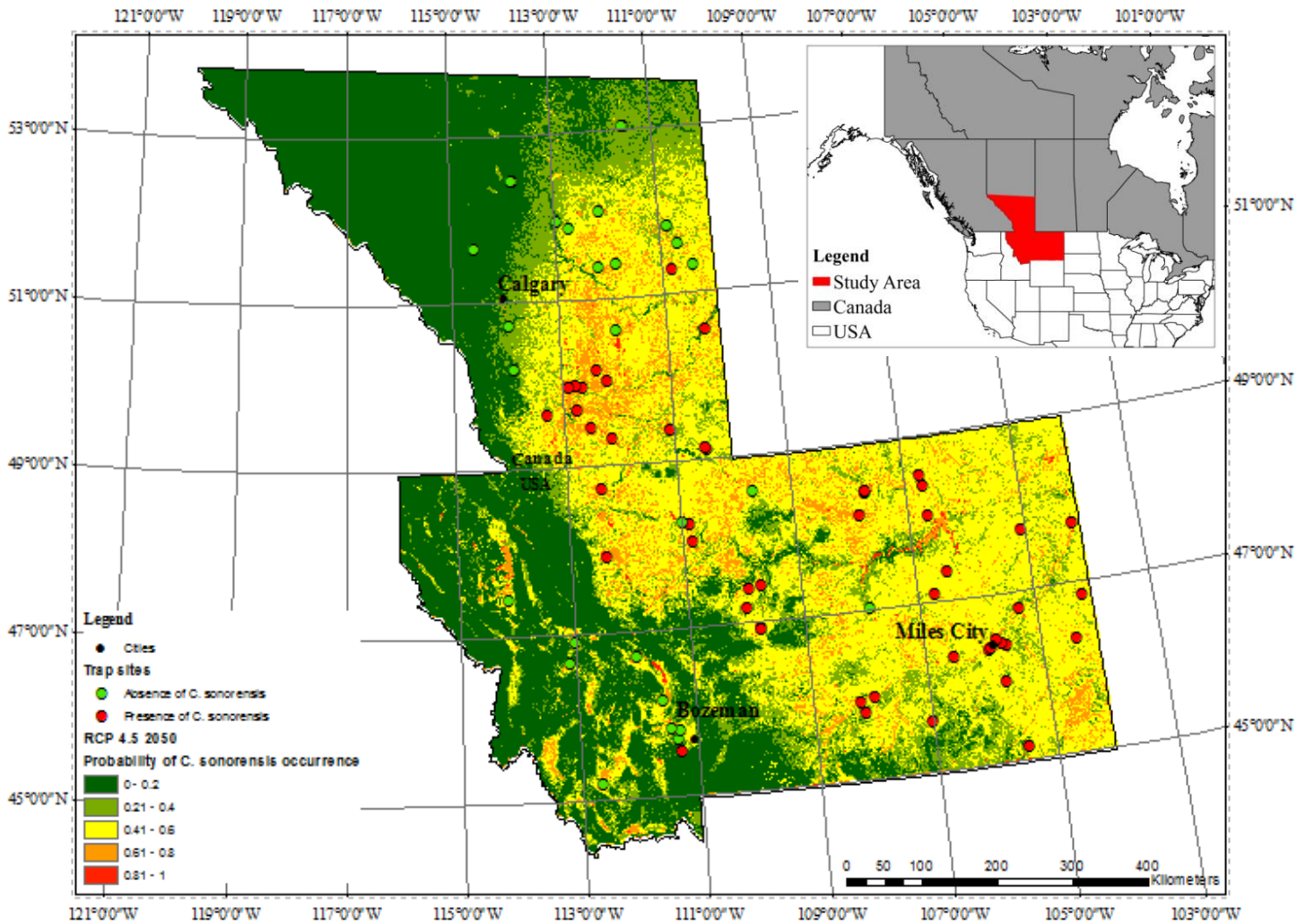
Appendix 5 Probability distribution map for *C. sonorensis* under RCP 2.6 scenario in 2050s (2040-2060).



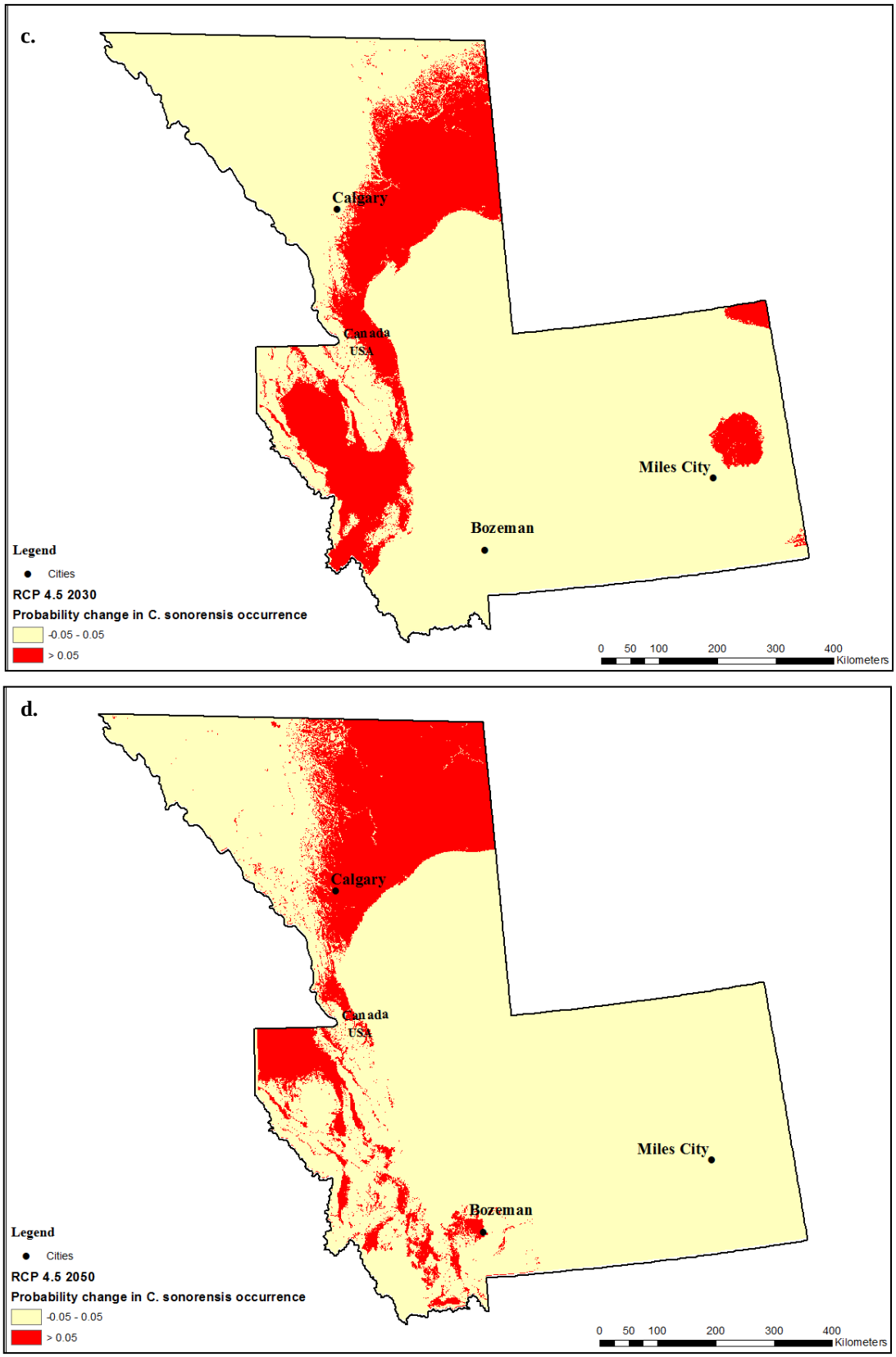
Appendix 6 Probability change in *C. sonorensis* occurrence from the baseline (model B) to **a)** 2030 and **b)** 2050 according to RCP 2.6. Blue areas showed a decreased probability of occurrence, whereas red areas showed an increased probability of *C. sonorensis* occurrence.



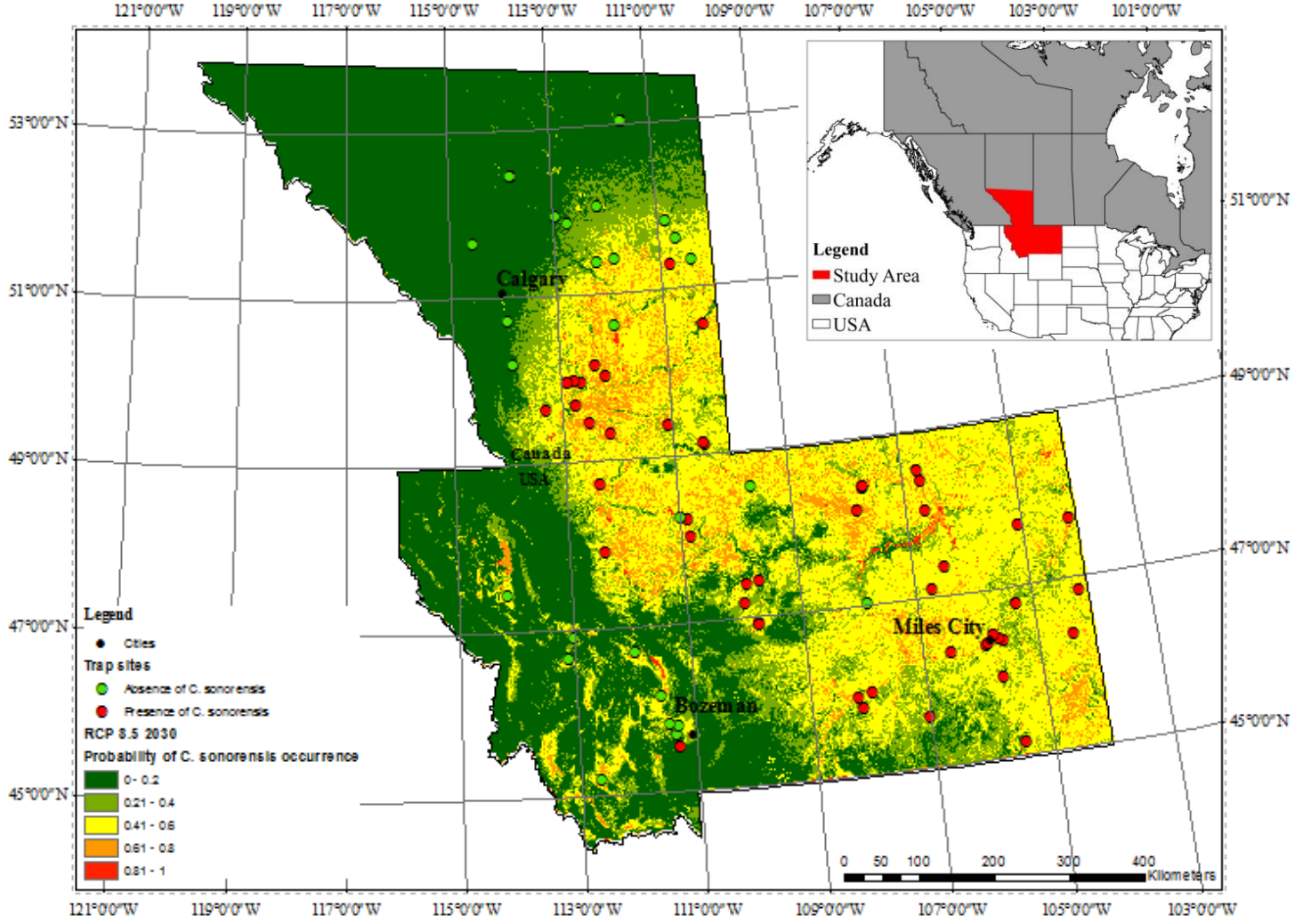
Appendix 7 Probability distribution map for *C. sonorensis* under RCP 4.5 scenario in 2030s (2020-2040).



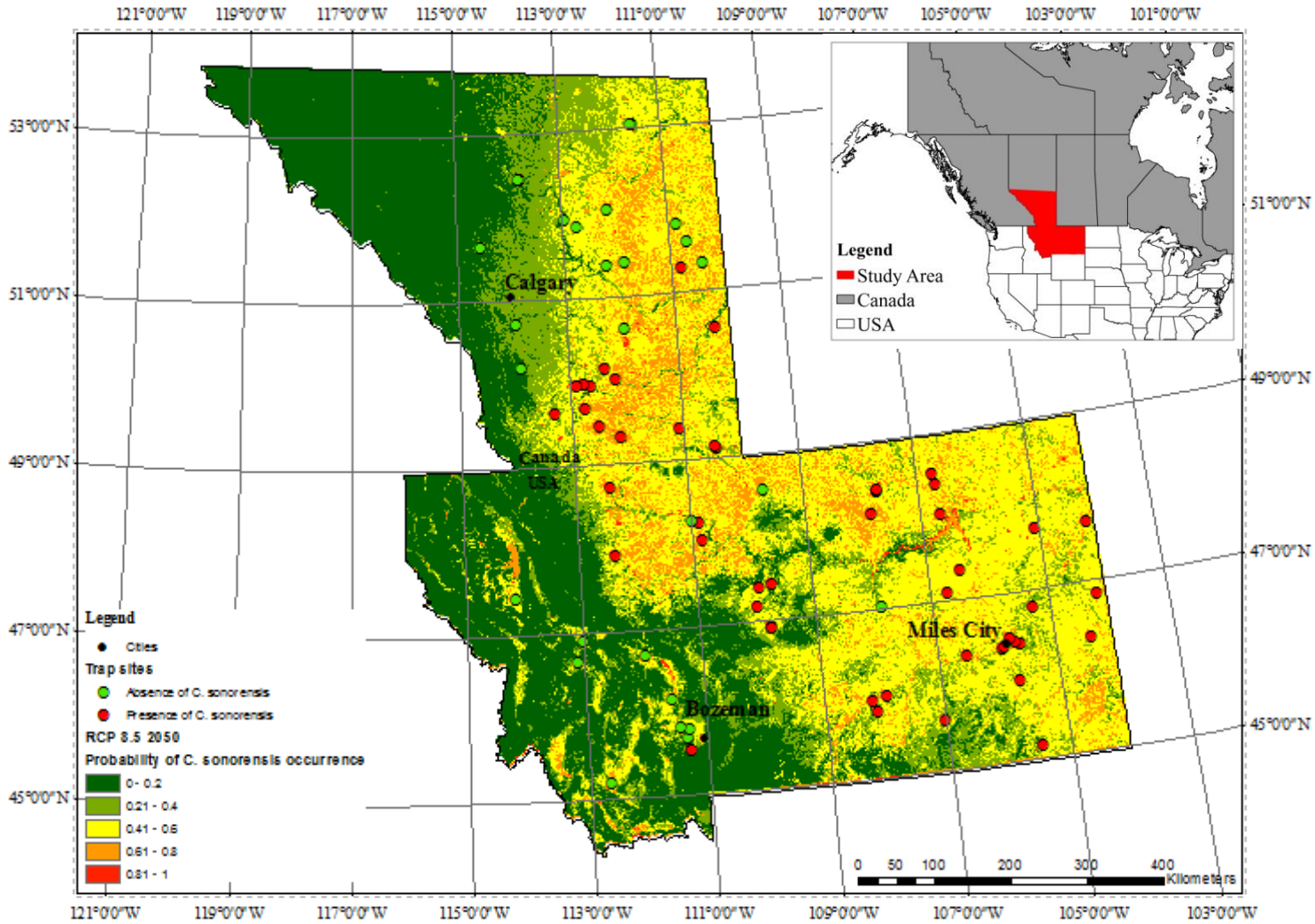
Appendix 8 Probability distribution map for *C. sonorensis* under RCP 4.5 scenario in 2050s (2040-2060).



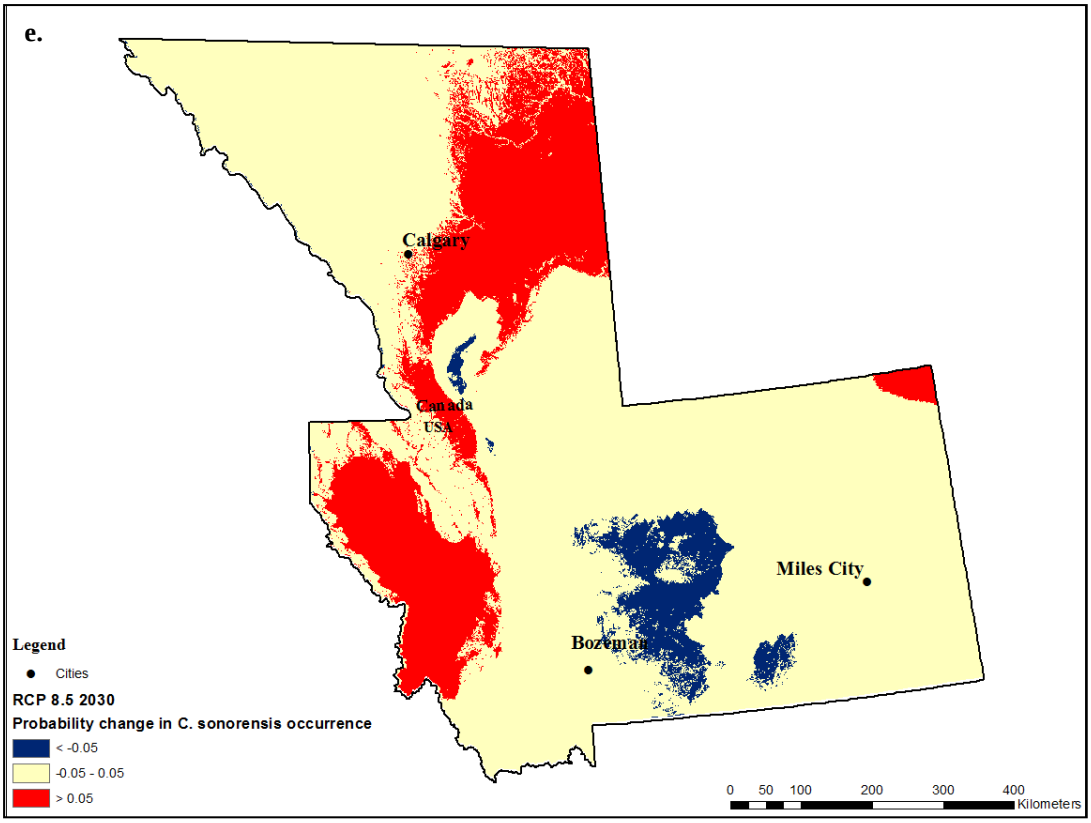
Appendix 9 Probability change in *C. sonorensis* occurrence from the baseline (model B) to **c)** 2030 and **d)** 2050 according to RCP 4.5. Red areas showed an increased probability of *C. sonorensis* occurrence.

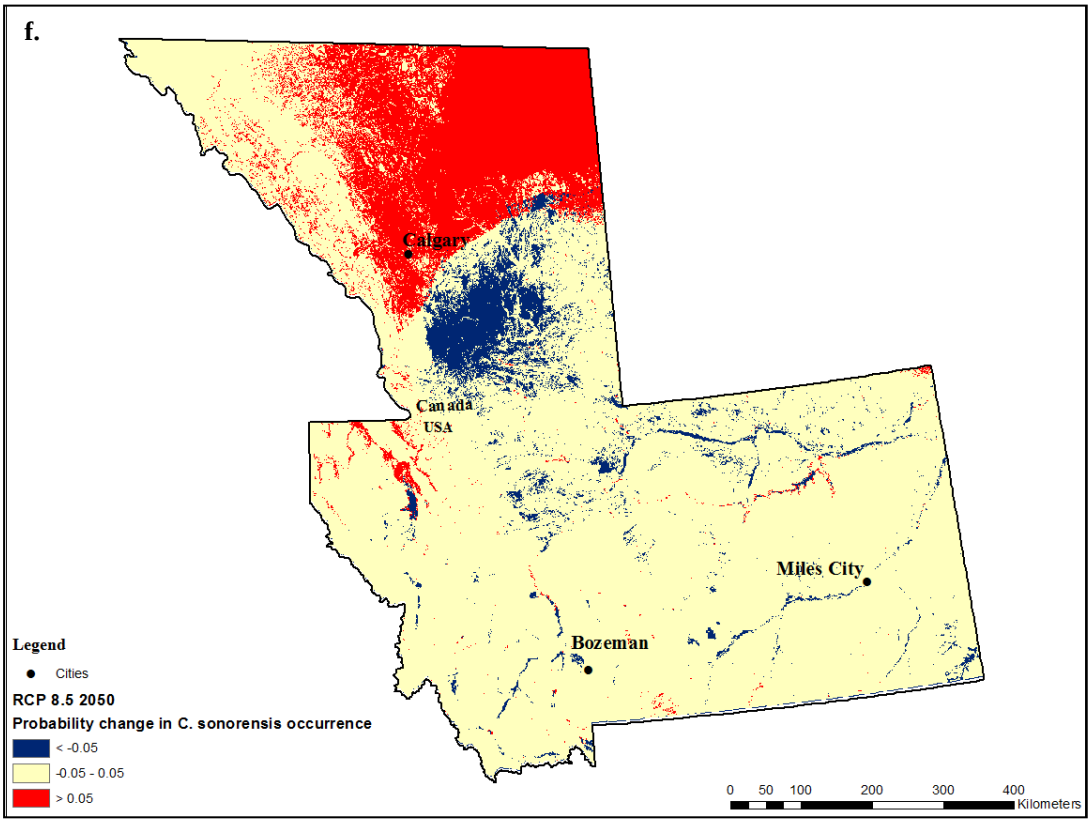


Appendix 10 Probability distribution map for *C. sonorensis* under RCP 8.5 scenario in 2030s (2020-2040).



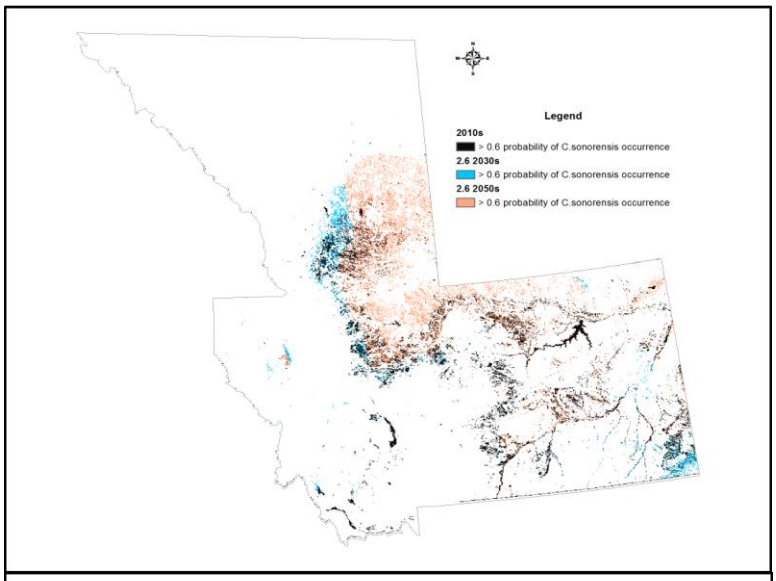
Appendix 11 Probability distribution map for *C. sonorensis* under RCP 8.5 scenario in 2050s (2040-2060).



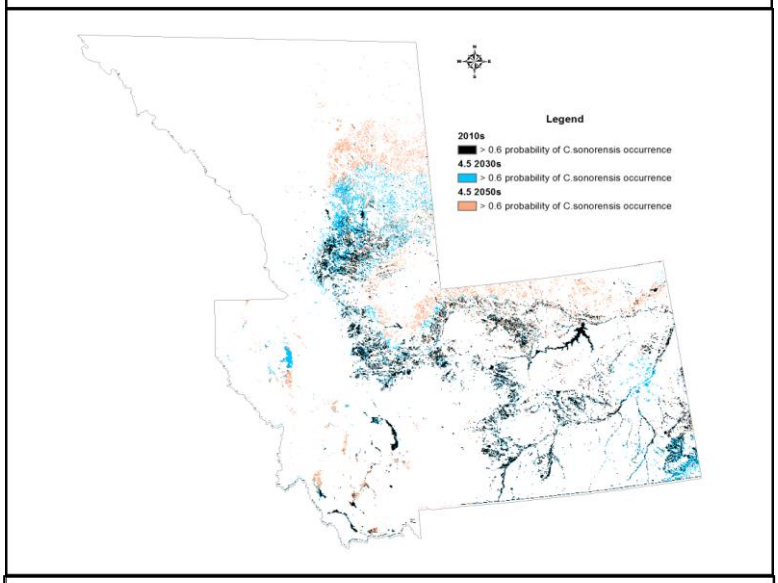


Appendix 12 Probability change in *C. sonorensis* occurrence from the baseline (model B) to **e)** 2030 and **f)** 2050 according to RCP 8.5. Blue areas showed a decreased probability of occurrence, whereas red areas showed an increased probability of *C. sonorensis* occurrence.

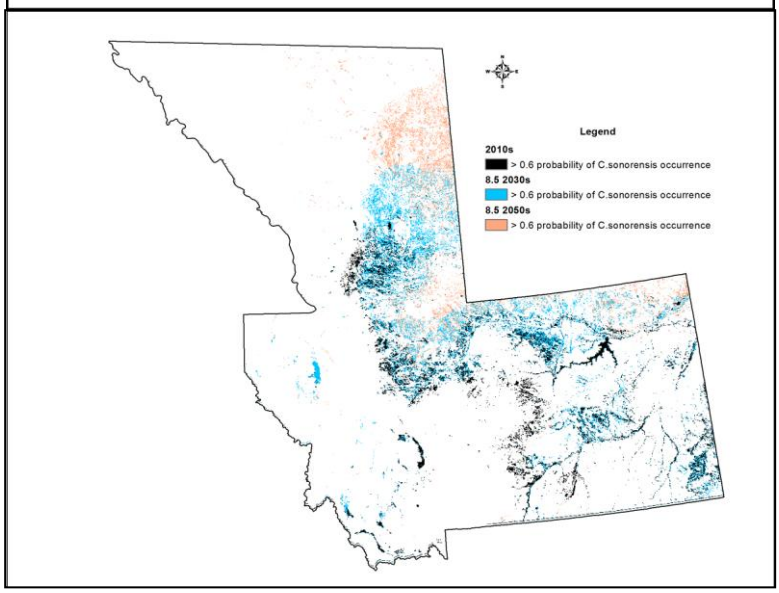
a.



b.



c.



Appendix 13 Areas at probability of *C. sonorensis* occurrence > 60% for a) RCP 2.6; b) RCP 4.5 and c) RCP 8.5. Baseline year (i.e.2010) is represented in black. projections for 2030s are depicted in blue and those for 2050s in orange.