ANALYSIS OF HISTORICAL AND CURRENT DISTRIBUTION OF POTATO PSYLLID (*BACTERICERA COCKERELLI*) AND THE INDUCED PLANT DISEASE PSYLLID YELLOWS, IN RELATION TO STANDARD CLIMATE INDICES

QING XIA
Bachelor of Science, South China Agricultural University, 2014

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Biology
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

© Qing Xia, 2017

ANALYSIS OF HISTORICAL AND CURRENT DISTRIBUTION OF POTATO
PSYLLID (*BACTERICERA COCKERELLI*) AND THE INDUCED PLANT DISEASE
PSYLLID YELLOWS, IN RELATION TO STANDARD CLIMATE INDICES

QING XIA

Date of Defence: August 16, 2017

Dr. Dan Johnson Supervisor	Professor	Ph.D.
Dr. Daya Gaur Thesis Examination Committee Member	Professor	Ph.D.
Dr. James Byrne Thesis Examination Committee Member	Professor	Ph.D.
Dr. James Thomas Thesis Examination Committee Member	Professor	Ph.D.
Dr. Igor Kovalchuk Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

The biogeography of herbivorous insects often depends on climate and host plants. Potato psyllid, *Bactericera cockerelli* (Sulc), is an insect that presents a threat to the western Canadian potato industry. It is native to the USA and Central America, and currently found at very low numbers on the Canadian Prairies. Potato psyllid can transmit *Candidatus* Liberibacter solanacearum (Lso), causing "zebra chip", severe tuber necrosis. Feeding by potato psyllid nymphs causes psyllid yellows on potato leaves. Historical records of this disease indicate potato psyllid abundance. I used Random Forest model methods to analyze the climate conditions of the historical psyllid yellows records. I used Maxent models to predict the presence probability of potato psyllids. The results indicate that higher temperatures and moderate precipitation are related to larger potato psyllid populations, and suggest the northern expansion of climatically suitable range in western Canada and decrease in the southwestern USA under climate change.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The tomato-potato psyllid, *Bactericera cockerelli* (Sulc) (Hemiptera: Triozidae), is a small flying insect that can transmit *Candidatus* Liberibacter solanacearum (Lso), a plant pathogen (gram-negative unculturable bacterium) known to cause zebra chip disease (ZC) in potato. Infection by Lso causes extensive cell death within the vascular tissue of the potato tuber, resulting in zebra-striped pattern in chips after deep frying (Miles et al. 2010). Infected plants can produce ZC tubers, and the Lso also reduce the seed potato quality, which leads to no germination or very weak plants (Sengoda et al. 2010; Munyaneza 2012). ZC reduces the sugar content within the potato tubers, and the symptoms of ZC can continue to develop during storage (Munyaneza 2015). The pathogen can be transmitted among potato plants by grafting (Crosslin & Munyaneza 2009). The inoculation of one potato psyllid can produce Lso infected plants after feeding for as little as two to six hours. ZC symptoms will develop in tubers about three weeks after inoculation (Munyaneza 2015). Although the ZC tubers can develop very weak or prematurely dead potato plants, the surviving progeny plants can develop ZC symptoms in the daughter tubers (Pitman et al. 2011). Zebra chip has been observed in New Zealand, Norfolk Island (an island between New Zealand and Australia), Texas, Arizona, California, Idaho, Oregon, Washington, Mexico, and Nicaragua (Cameron et al. 2009; Crosslin & Bester 2009; Munyaneza et al. 2009; Brown et al. 2010; Crosslin et al. 2012a; Crosslin et al. 2012b; Munyaneza 2012; Bextine et al. 2013). The above-ground symptoms of zebra chip include leaf yellowing, purpling, curling, upward rolling, aerial tubers, stunted plants and more severe plant death (Munyaneza 2012). Psyllid yellows is caused by feeding of the potato psyllid nymph, not the feeding of potato psyllid adults

(Richards & Blood 1933). The above ground symptoms are similar to zebra chip, including leaves yellowing, purpling and upward rolling. The below ground symptoms include excessive potato tuberization, early sprouting of the immature tubers, knobbed and malformed tubers. This disease significantly reduces potato crop yields, seed potato quality, and more severely, causes plant death (Richards & Blood 1933). Psyllid yellows does not produce tuber necrosis like the zebra chip does (Sengoda et al. 2010). Multiple outbreaks of psyllid yellows with recorded 100% yield loss were documented in the US during the 1930's and 1940's. Psyllid yellows had been observed in Central and Southern Alberta, and Saskatchewan in the late 1930's. Vectoring of the Lso pathogen by potato psyllid is required for spreading the zebra chip disease among the potato plants (Munyaneza 2015). In a recent report from the University of Idaho, it was reported that pesticide control of the zebra chip and potato psyllid can cost farmers from US$170 to US$590 per acre for 6 to 10 applications. Despite using pesticide, 50% yield loss was still reported by a farmer (Guenthner & Greenway 2010). Greenway found that Eastern Idaho would suffer a 55% reduction in returns for potato producers (Greenway 2014). Annual losses in Texas have been reported in the range US $25 to $30 million (Texas A&M AgriLife Research & Extension 2016).

After the ZC outbreaks in Idaho, Washington, and Oregon in 2011 (Crosslin et al. 2012a; Crosslin et al. 2012b), monitoring programs of potato psyllid have been set up in those states. In Idaho, the state potato psyllid monitoring program has been led by Dr. Erik Wenninger, at the Kimberly Research and Extension Center, University of Idaho, Kimberly, Idaho since 2012. In Oregon, Dr. Silvia Rondon, at the Hermiston Agricultural Research and Extension Center, Oregon State University, has been leading the

monitoring program of the potato psyllid population in Oregon since 2012. In Washington, Dr. Bill Snyder, at Washington State University, has been leading a US$2.7 million potato psyllid research and monitoring program since 2015 in Washington.

Potato psyllid can overwinter in the Pacific Northwest on bittersweet nightshade, *Solanum dulcamara*, which significantly increases the cold tolerance of the potato psyllid compared to those feeding on potato plants (Murphy et al. 2013; Horton et al. 2015). Potato psyllid can also feed on Matrimony vine, Silverleaf nightshade, and some other solanaceous plants (Munyaneza 2015; Thinakaran et al. 2015a; Thinakaran et al. 2015b). Potato psyllid has been observed in multiple locations in Southern Alberta since 2015. So far, Lso infection in potatoes or potato psyllid has not been detected in Alberta. A national monitoring program, the Canadian potato Psyllid and Zebra Chip Monitoring Network (2013 - 2017), was organized and coordinated by Dr. Dan Johnson at the University of Lethbridge.

Overview of potato psyllid distribution in Southern Alberta

Different terms were used to described potato psyllids in historical literature and field survey records: Chermidae, tomato psyllid, and jumping plant louse. Its previous scientific names were *Trioza cockerelli* (Sulc), and *Paratrioza cockerelli* (Sulc), currently *Bactericera cockerelli* (Sulc). Potato psyllids appeared temporarily in Edmonton, Alberta, around 1936, but were thought to be unable to overwinter (Strickland 1938). Red Deer, Alberta, was also reported to have had potato psyllid present at least once (Strickland 1953). Potato psyllids had been observed in association with psyllid yellows at various sites in Southern Alberta from 1928 to 1944. Unlike some other psyllid yellows studies and observations during that time in the United States,

severe potato tuber phloem necrosis sometimes accompanied potato psyllid infestations observed in Lethbridge, Medicine Hat, and Calgary in 1938, and continuing to 1939 and 1940. Psyllid yellows usually did not produce necrotic potato tuber (Richards & Blood 1933; Sengoda et al. 2010), a condition described as abnormal psyllid yellows (Sanford 1952). During the same years, extensive and severe psyllid yellows outbreak took place in Montana, Wyoming, Nebraska, and Colorado. Lethbridge Research Station had reared potato psyllid population in the past. However, it perished during the winter from 1943-1944 in the greenhouse under critical low temperature (Sanford 1952). The table below summarizes the years and locations of observations of potato psyllids in southern Alberta. The locations include Medicine Hat, Calgary, Edmonton, Lethbridge, Drumheller, Barnwell, Taber, Brooks, Cowley, and Olds. Psyllid yellows outbreaks were observed in those locations during different years.

Table 1.1. Potato psyllids records in southern Alberta from 1928 to 1944 (Marritt 1935, 1936, 1937, 1938; Strickland 1938; Marritt 1939; Strickland 1939; Marritt 1940; Sanford 1941; Manson 1944; Sanford 1952)

City or town in Alberta	Year
Barnwell	1938
Brooks	1939
Calgary	1936, 1937, 1938, 1939
Cowley	1939
Drumheller	1937, 1944
Edmonton	1936, 1938, 1939
Lethbridge	1936, 1937, 1938, 1939
Medicine Hat	1928, 1932, 1935, 1936, 1937, 1938
Olds	1939
Taber	1938

.

During 2013-2017, a national and provincial monitoring program has been managed by national coordinator and research P.I. Dr. Dan Johnson from the University of Lethbridge in conjunction with research on sampling, natural enemies, forecasting, and genetic characteristics (the "Canadian Potato Psyllid and Zebra Chip Monitoring", 2013-2017).

The methods used include mainly yellow sticky traps staked in crop within 10 m of the potato field margin, sweepnet sampling, portable vacuum sample, and leaf examination. Traps are changed weekly, and examined for potato psyllid under dissecting scope. Being part of the sampling network and Johnson lab at the University of Lethbridge, I participated in field sampling during the Summer in 2015, as well as study design discussions and management. During 2013 and 2014, no potato psyllids were found in the sampling sites in Alberta. During the years from 2015-2017, additional permanent field sites (up to 41 per year) were established in potato fields and checked regularly by University research personnel, as well as additional sites at known locations managed by Alberta Agriculture & Forestry (Scott Meers and Shelley Barkley), and in 2016 by the Potato Growers of Alberta (Jay Anderson). Additional sampling at unknown locations was conducted by private scouting services but were not made available to the network database or psyllid mapping project. Potato psyllids were observed mainly near Coaldale, Bow Island, Pincher Creek, Taber, Vauxhall, Lethbridge, and Picture Butte, Alberta. With an expansion of sampling, potato psyllids were found near Edmonton (a site operated by Tina Lewis and collaborators), Manitoba (Vikram et al. 2016), and Saskatchewan (Jazeem Wahab). All specimens were subjected to analysis of cytochrome c oxidase subunit 1 (COI) of individual potato psyllids at the L. Kawchuk lab, Lethbridge Research Centre. So far, PCR analysis indicated that all potato psyllids were negative for

Lso, the main concern of the sampling network. This thesis is a study of the relationship

between weather and climate to psyllid observations, primarily in historical literature, in

support of understanding of the biogeography of the potato psyllid and its impact on

sustainable agriculture.

Current sampling effort: sample cards examined at the University of Lethbridge in 2016

(team effort. Most cards during 2014-2017 were examined by Dan Johnson, Sampath

Walgama, and Christian Sapsford)

Table 1.2. Number of yellows sticky cards collected from each province in 2016

Province	Cards	Percentage (%)
Alberta	1384	66
British Columbia	73	4
Manitoba	113	5
Québec	3	<1
Ontario	Sampling planned for 2017	NA
Newfoundland and Labrador	3	<1
Prince Edward Island	8	<1
Saskatchewan	14	1
New Brunswick	486	23
Total	2084	100
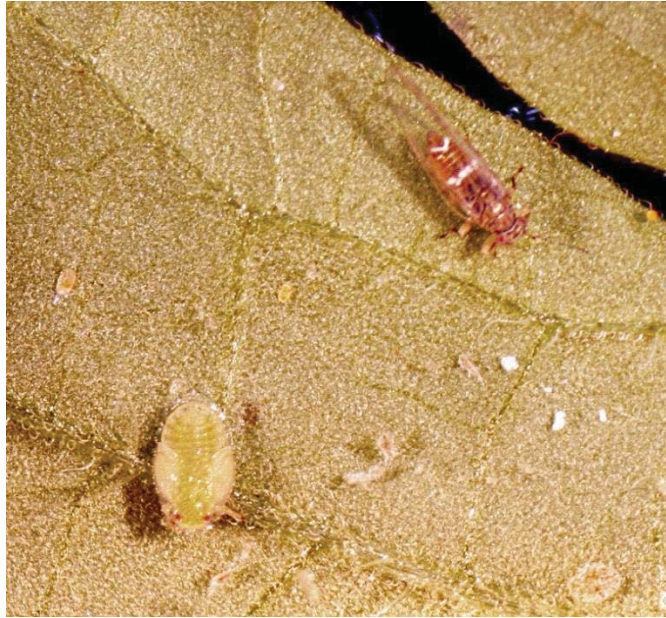
Figure 1.1. Potato psyllid, *Bactericera cockerelli* (Sulc) (Hemiptera: Triozidae), adult (upper right) and nymph (lower left, fifth instar) feeding on potato leaf (photo: Q. Xia)

Observations of potato psyllids in North America, using information from entomological and agricultural literature, field survey records, and museum records

The locations of occurrence of potato psyllids cover a wide area in North America, beginning early in the twentieth century. Potato psyllids were first collected in Boulder, Colorado, by Dr. Theodore Dru Alison Cockerell from the University of Colorado, Boulder, in 1909.It was identified as a previously unknown species and given its scientific name, *Trioza cockerelli*, by Dr. Karel Sulc (Sulc. 1909). It was classified being under the genus *Trioza* due to its three-vein venation on its upper wing, and given its species name *cockerelli* after its collector, Dr. Cockerell. Later the term *Trioza cockerelli* (Sulc) was changed to *Paratrioza cockerelli* (Sulc), and then to *Bactericera cockerelli* (Sulc).

The lists of states with recorded potato psyllid observations based on museum specimens or literature records are summarized in the tables below. The potato psyllid specimen records are mainly from the Essig museum online database, hosted by the Essig Museum of Entomology, located at the University of California, Berkeley, California, USA (https://essigdb.berkeley.edu/)

Table 1.3. Years and counts of distinct localities for each state of the Essig database specimen records

State	Year Range	Distinct count of locality
Arizona	1929-2007	37
California	1908-2011	193
Nevada	1930-1995	11
New Mexico	1929-2002	10
Texas	1985	1
Utah	1912-1964	14

Table 1.4. Years and counts of distinct localities for each state from the plant disease periodicals, the Plant Disease Reporter (1923-1979), the Plant Disease Bulletin (1917-1922), the Plant Disease Reporter Supplement (1919-1929), the Canadian Plant Disease Survey (1929-2016), and other literature articles

State	Distinct count of Locality	Min. Year	Max. Year
Alberta	5	1928	1939
Arizona	6	1943	1945
California	8	1915	1943
Colorado	28	1926	1944
Idaho	2	1927	1927
Montana	55	1927	1946
Nebraska	23	1938	1938
New Mexico	1	1943	1943
Saskatchewan	1	1938	1938
South Dakota	1	1944	1944
Texas	9	1936	1945
Utah	34	1927	1943
Wyoming	7	1932	1944

Weather has an influence on the potato psyllid distribution

Temperature, humidity, and rainfall are the main conditions that favor or restrict potato psyllid growth and development, primarily through influence on development and survival, but also via impacts on reproduction and movement. The 100 meridian was said to be the approximate zone of geographical humidity that hinders the potato psyllids from spreading east (Wallis 1955). Higher humidity in winter was noted to decrease the survival of the potato psyllid during hibernation (Pletsch 1947). However, under non-lethal temperatures, higher humidity increases survival rate of fasting potato psyllids (Pletsch 1947).

Insects are ectothermic, so ambient temperature has a direct influence on body temperature, and determines rates of development and survival, and other physiological and ecological processes. Based on the laboratory results in a study (Tran et al. 2012), potato psyllids start to grow and develop at or above a temperature of 7 ℃. The development rate increases with temperature until the temperature reaches an optimum threshold (27.6 ℃) for the insect and begins to decrease when the temperature passes a threshold of 27.6 ℃ . Development ceases when the insect body temperature rises above 33.6 ℃ (if possible, psyllids will move or choose microclimate positions to maintain a lower body temperature). The accumulated heat (degree-day) requirements of potato psyllid to finish one life cycle, from egg to propagation, is 358 degree-days above 7.1 ℃ when feeding on potato. Psyllid yellows severity were observed to be lower when the summer temperature was around 90 °F, equal to 32.2 ℃ (Wallis 1946; Pletsch 1947). This insect was described as a "low-temperature" species by List (List 1939), and higher summer temperatures were known to decrease psyllid yellows severity, whereas the cooler, higher mountain areas were described as more favorable for the disease outbreak in Colorado (List 1939). Potato psyllid nymphs can withstand subfreezing temperatures, and were observed feeding on field bindweed when exposed to repeated frost and under temperatures as low as 6 °F, -14.4 ℃ (Pletsch 1947). Potato psyllid nymphs were observed to survive -15 ℃ for 24 hours with 100 % survival. Potato psyllid adults were observed to have 40% survival rate after being exposed to -10 ℃ for 24 hours when feeding on potato (Henne et al. 2010). Potato psyllids' ability to tolerate winter cold varies regionally, and apparently by population source and genetic characteristics. Psyllids from a Nebraska colony were found to be more cold-tolerant than the Southern Texas colony (Whipple et al. 2013). Recently, potato psyllids were found overwintering

on bittersweet nightshade in the Pacific Northwest (Murphy et al. 2013). The population composition of the potato psyllid haplotypes varied regionally (Swisher et al. 2013). The northwestern haplotype is more cold-tolerant than the central and western haplotypes of potato psyllid (Swisher et al. 2013).

Historically, warm springs with below-average rainfall were recorded to be beneficial for psyllid yellows outbreaks, however low influence of rainfall on potato psyllid damage to crops was also suggested by Wallis (1955). Higher precipitation in July in Nebraska was described to benefit potato psyllid in field populations (Hill 1947). A study conducted in Montana from 1937 to 1941 showed that higher precipitation seemed to benefit and increase the psyllid infestation (Pletsch 1947).

Rapid increase in potato psyllid populations was observed in the field by Wallis (1955). Among the literature of the time, the opinion that potato psyllid could spread across different states seemed to be a consensus. As mentioned, the 100 meridian was used as a border line that defines the spread limitation due to humidity tolerance of the insect (Wallis 1955). Seasonal occurrence was observed in the North Platte River Valley of the western USA (Wallis 1946). Potato psyllid distribution pattern differs seasonally from south to north in the United States (Nelson et al. 2014).

Contribution of the thesis

Evidence of the influence of climate and weather on potato psyllid populations and distribution are apparent in the historical and recent literature. However, there is so far no large-scale quantitative study using historical records to analyze pattern of the combination of different climate conditions of the potato psyllid occurrence and, of its

nymphal feeding-induced potato disease psyllid yellows in the field environment. In order to investigate and explore the probable relationship between climate conditions, potato psyllid, and psyllid yellows, I used machine learning algorithms to analyze the relationship between the potato psyllid, psyllid yellows records, and climatic indices. The more influential the climatic conditions are on the potato psyllid, or psyllid yellows, in terms of predicting the probability of presence, the more important the indices will be indicated as significant by the model. The relationship between the potato psyllid, psyllid yellows, and the climatic conditions will be fitted by the model based on the historical, geographical and agricultural survey data. My study identifies candidate large-scale critical climate conditions and influences which can be used in smaller-scaled regional weather, climate studies of the species in the future.

A short description of the thesis structure

The thesis project aims to analyze the climate and potato psyllid apparent relationships and interactions, and climate and psyllid yellows interactions, and identify the critical variables indicated by historical potato psyllid observations, and psyllid yellows observations collected from the museum, periodical plant disease reports, and the literature, including reports and scientific articles. Standard bioclimate indices for species distribution modelling, BIOCLIM, and climate extremes indices, CLIMDEX were used to analyze the insect occurrence and psyllid yellows incidents, accompanied with the monthly, annual maximum, minimum, mean of the daily temperature data, precipitation data, and monthly and annual degree-days. The analysis of historical psyllid yellows records and climate indices using Random Forest is described in Chapter 2. The analysis

of potato psyllid's distribution in North America using Maxent is described in Chapter 3.

Chapter 4 is the conclusion of the thesis research.

Chapter 2: Historical analysis of psyllid yellows on potato using random forest model

Chapter overview:

In this chapter, an overview of the history of potato disease psyllid yellows in North America is given. This chapter uses a Random Forest ensemble learning method to select important climatic variables related to psyllid yellows incidents. Also, partial dependence plots were fitted by Random Forest to visualize the relationship between the psyllid yellows likelihood and climatic indices.

Overview of psyllid yellows research

Psyllid yellows was first noticed and documented in Utah by Richards (1933) in August,1927. This leaf-yellowing disease of potato crops was confirmed through controlled experiments to be caused by feeding of the nymph of the potato psyllid, *Bactericera cockerelli* (at that time known by the synonym, *Paratrioza cockerelli*), and not related to the feeding of adult potato psyllids. It was initially noted in field observations that feeding of potato psyllid nymphs is required for the development of psyllid yellows on potato. When the psyllid yellows condition is observed in the field, at least one generation of potato psyllid has started, and therefore immature stages with sucking mouthparts are present and feeding on tissues of leaves. Around 30 nymphs feeding on one plant for three days can develop psyllid yellows symptoms (Richards & Blood 1933). We confirmed this relationship at the University of Lethbridge, in which feeding by potato nymphs (especially immature instars 3-5) resulted in leaf yellowing

and purpling, with subsequent wilting, necrosis, and spotting, on potted potato plants grown under confined laboratory conditions (unpublished data, Fig. 2.1).

Figure 2.1. Potato leaves showing yellowing, and discoloration due to potato psyllid nymph feeding (photo: Dan Johnson)

Psyllid yellows is suggested to be caused by insect salivary toxin secreted by potato psyllid nymphs during feeding. However the toxin has not been identified (Sengoda et al. 2010). The psyllid yellows symptoms do not include necrosis of the potato tuber (Sengoda et al. 2010), and differs from the pathology and symptoms of zebra chip disease. Zebra chip disease, first noted in Texas, USA, in 2000, is caused by the Lso pathogen (*Candidatus* Liberibacter solanacearum, a gram-negative bacteria, of the family Rhizobiaceae) transmitted through potato psyllid feeding, and leads to necrotic potato tuber conditions that result in a striped or zebra pattern after being deep fried (Munyaneza 2012). Both of these conditions cause economic loss to potato production, although zebra chip is currently a much greater concern because of potential vectoring and rapid horizontal transmission by the insect. Prolonged nymphal feeding can create severe psyllid yellows symptoms which severely impedes photosynthesis and kills the potato plants, or in less severe cases results in small potato tubers (Richards and Blood

1933). Zebra chip can also lead to plant death, in most cases an economic impact is caused by unmarketable tubers that produce discoloured, often striped potato chips after slicing and deep frying. Less severely infested potato plants can recover from psyllid yellows with removal of the potato psyllid, whereas potato plants infected with zebra chip can not (Sengoda et al. 2010).

Weather, notably temperature, rainfall, humidity, wind, and related variables measured in standard meteorological stations, is well known as an environmental determinant for occurrence, growth, survival, and reproduction of living organisms, including insects and plants. Biometeorology and climatology are often the sources of key variables in the biogeographical analysis (Wellington et al. 1999). Initial explanations of the occurrence or psyllid yellows and expansion of psyllid populations considered the role of weather. Mild winter conditions, warmer spring weather with less rainfall, warm weather in August and September, and other weather patterns were noted as possibly related to psyllid yellows outbreaks in the past (Wallis 1955), but the relationship was unclear. Humidity seems to be unfavorable for psyllid population expansion. The 100 meridian was thought to be the geographical humidity limitation that hinders the eastward spread of potato psyllid in North America (Wallis 1955). Some authors claimed that more severe psyllid yellows developed during dry years (Daniels 1937; Wallis 1955). On the other hand, Wallis (1955) noted that in some cases precipitation does not appear to influence the severity of psyllid yellows damage. Historical data records of psyllid yellows provide insight into factors affecting the distribution and abundance of potato psyllids. In my study, to test the hypothesis that appearance of psyllid yellows outbreaks depends on key geographical environmental factors, and to understand whether standard methods of

classification method and comparison could yield predictive power, I compared the

weather and climate conditions of the locations of psyllid yellows observations to the

locations that did not have psyllid yellows observations.

Data sources

I collected historical observations of psyllid yellows from the Plant Disease Reporter, the

Plant Disease Bulletin, the Plant Disease Reporter Supplement, the Canadian Plant

Disease Survey, and literature articles.

Plant Disease Reporter is a nationwide collection of crop disease survey observations

throughout the United States. Plant Disease Reporter (1923 - 1979) was issued by the

Plant Disease Survey, published by the Bureau of Plant Industry, U.S. Department of

Agriculture, U.S. Bureau of Plant Industry, and the U.S. Agricultural Research Services.

It is a national government publication of the United States. This monthly periodical was

published and distributed for use by plant pathologists, and provided summaries on

agricultural plant diseases from across the states. It is a collection of facts and opinions

regarding plant disease control, epidemics, news, and locations, provided by qualified

observers, with different document types: reports, observations, notes, comments,

summery, maps, and data tables. It was a popular plant disease periodical during that

time. It has been used as a reliable source, and cited by different potato plant disease

researches (Krause et al. 1975; Mojtahedi et al. 1991; Harrison 1992; Butler & Trumble

2012). The previous title was the Plant Disease Bulletin (1917-1922), and it is currently

Plant Disease, published by the American Phytopathological Society. I went through

each issue of these three periodicals, to find the recorded observations and descriptions of

the historical psyllid yellows outbreaks, which serve as a proxy indicator of past potato

psyllid outbreaks. Volume 11 in 1927, to volume 31 in 1947 have records of the psyllid yellows outbreaks, so I extracted those observations, recorded the date and descriptions of the observations of the psyllid yellows incidents, and georeferenced the locations.

Canadian Plant Disease Survey is a national collection of crop disease survey observations throughout Canada. The Canadian Plant Disease Survey, published by the Canadian Phytopathological Society and, Agriculture and Agri-Food Canada, is a periodical government publication that reports information of plant disease epidemics, observations and related losses caused by plant diseases, and is based on reports provided by different plant pathology laboratories, plant pathologists, and botanists throughout Canada. The available records on the Canadian Phytopathological Society website are from 1927 to 2017. I went through volume 8 to volume 97 of this periodical literature to search for observations and reports that recorded information relevant to psyllid yellows. In this periodical, psyllid yellows was recorded between 1932 and 1944. It was mainly observed in Alberta and Saskatchewan, by J.W. Marritt, G.B. Sanford, and G.F. Manson.

Historical psyllid yellows studies, and potato psyllid ecological research in the past

Psyllid yellows disease caused significant economic losses to potato production in Canada and the United States in the 1930's and the 1940's, because of the damage and reduced yield production associated with the symptoms of leaf yellowing and purpling, stunting plants, smaller and misshapen tubers(Carter 1939; Munyaneza 2012). Only feeding by potato psyllid nymphs can induce psyllid yellows symptoms on potato plants. The feeding of potato psyllid adults can not produce the psyllid yellows symptoms on potato (Richards & Blood 1933; Schaal 1938; Carter 1939), and as noted, we have confirmed this experimentally at the University of Lethbridge (unpublished data). In a

study of Richard and Blood (1933), potato psyllid nymphs and adults of different numbers were placed separately on healthy potato plants in different cages. Healthy potato plants, free from potato psyllids within their own cages, were used as a control for comparison. The author found that psyllid yellows symptoms can develop on one potato plant with minimum 30 nymphs feeding for three days. Whereas, even up to 1000 potato psyllids adults feeding for 25 days, psyllid yellows symptoms failed to develop on the potato plants. Similar observations and descriptions were made by other studies during that time. Binkley recorded that feeding of potato psyllid nymphs was observed to cause psyllid yellows on tomato in Colorado (Binkley 1929). Psyllid yellows on potato and tomato had been noticed with certainty since 1927. Before 1929, a study by List stated that the psyllid yellows disease could be related to the feeding of potato psyllid nymphs (List 1925).

Overview of the historical psyllid yellows observations in North America

Psyllid yellows cases were recorded in Texas, Arizona, California, Colorado, Idaho, Montana, Nebraska, New Mexico, South Dakota, Utah, Wyoming in the United States, and in Alberta and Saskatchewan in Canada. The collected records cover observations from 1915 to 1947 in the US, and from 1928 to 1944 in Canada. In the US, the earliest psyllid yellows record was in 1915, in California. The latest record of psyllid yellows in this historical plant disease database found for the United States was in 1947, with general region description, "western States from North Dakota to Texas, California, and Idaho, especially in Colorado, Utah, and Wyoming" (Plant Disease Survey 1947). In Canada, the earliest found record of psyllid yellows with detailed locations was in Medicine Hat, Alberta, in 1928. A "local malady" of psyllid yellows epidemic was recorded in central Alberta as early as 1919 (Sanford 1934). The area coverage and the

severity of psyllid yellows are historically much lower in Canada than in the United States, which also suggests a role of weather and climate. The following section provides summaries of historical observations of psyllid yellows in each state and province.

Summary of the historical psyllid yellows observations in the United States

Montana

The earliest psyllid yellows record collected in Montana was in 1927, found in Bozeman (Richards and Blood 1933). Psyllid yellows in Montana in 1938 was widespread, covering most of the counties within Montana, and caused significant crop loss to the Montana potato production (Morris 1939). After the outbreak, a multiple-year survey by Pletsch (1947) was carried out from 1939 to 1946 to monitor the potato psyllid population in late June across Montana. During the months after June, news about psyllid yellows from different counties was reported and recorded by Pletsch (1947). Psyllid yellows cases were recorded in 1939, 1940, 1941, 1942, with one to three counties infested by psyllid yellows each year in Montana (Pletsch 1939; Pletsch 1947).

North Dakota

Psyllid yellows cases were recorded in North Dakota. However, no detailed locations and further information were available from publications and reports (Plant Disease Survey 1947). No other psyllid yellows records were found recorded in North Dakota.

South Dakota

Destructive potato psyllid outbreaks with severe crop loss were reported in the Black Hills region in South Dakota in 1944 (Plant Disease Survey 1944).

Wyoming

During years from 1932 to 1936, and 1944, psyllid yellows were observed at different

locations in Wyoming. During 1932 to 1936, the available outbreak records were in

southern Wyoming, whereas in 1944, available psyllids yellows records were from

northern regions in Wyoming. The psyllid yellows outbreaks from 1934 to 1936 were

extensive, from Riverton county to the Montana border, and caused severe potato crop

loss. A description was made for the outbreaks in Wyoming from 1934-1935 (Hartman

1937): "psyllids were reported in every potato-producing section of the state." During the

seasons from 1935 to 1936, all of the agricultural areas of Wyoming were said to be

influenced by the psyllid yellows epidemic (Hartman 1937).

Idaho

Based on the historical plant disease literature records, psyllid yellows were recorded in

Idaho in two regions during a short period. Psyllid yellows was observed in Twin Falls

and Idaho Falls in Idaho in 1927 (Wallis 1955). In 1928, psyllid yellows were found to

be absent from the southern potato growing region in Idaho (Plant Disease Survey 1929).

Nebraska

As in some other western states, 1938 was a widespread psyllid yellows disease year for

Nebraska, covering the southern and western agricultural areas of Nebraska. Severe

damage by psyllid yellows was observed in the irrigated and dry land area. Most of the

psyllid yellows infested area suffered moderate damage (Goss 1938).

Utah

Utah has psyllid yellows records from 1927 to 1943. Potato psyllids were observed

reproducing on matrimony vines, *Lycium barbarum* (synonym: *Lycium halimifolium*),

during the early spring in 1928, 1931, and 1933 (Plant Disease Survey 1929; Knowlton

1934; Cranshaw 1993). The potato psyllid population was observed to be abundant in the fields for multiple years, especially 1927 and 1928. Psyllid yellows was severe and destructive in 1927, with significant potato tuber yield reduction, and total failure of the potato crops in some locations (Richard et al. 1927; Richards & Blood 1933). In the upcoming 1928, early spring potato psyllid populations were observed on matrimony vine in Logan in Cache County, and Washington County. During the same year, extensive psyllid yellows took place in Utah (Richards 1928b, a; Plant Disease Survey 1929; Richards & Blood 1933). However, in 1939, psyllid yellows was not observed in Utah (Blood & Christiansen 1940). In 1947, psyllid yellows was observed in Utah, while no detailed location information available (Plant Disease Survey 1947). A survey conducted between 1932 to 1933 shows that, potato psyllid reproduced in Utah and the range covers the northern, central, and southwestern areas of Utah (Knowlton 1934).

Colorado

Colorado had its worst psyllid yellows outbreak in 1938, the same year that Wyoming, Montana, and Nebraska reported extensive psyllid yellows infestation and severe crop loss. The epidemic in Utah in 1938 cover the mountainous region of the state and the northeastern area. Colorado had psyllid yellows outbreak since 1911 (Daniels 1937; Daniels 1939). Among the collected records, psyllid yellows was last recorded in Colorado in 1944, with limited infested area and low damage on potato yield (Tervet 1944).

California

Severe and moderate damage was observed to be caused by potato psyllid feeding on potato and tomato crops in California from 1926 to 1943. Observations of potato psyllid

in California date back to 1915. Potato psyllids was said to overwinter in southern California (Jensen 1954). Potato psyllid was a "year-long brood" in Sacramento, California (Compere 1916). The collected potato psyllid museum specimen records show an extensive coverage of most of California.

Arizona

Slight psyllid yellows infestation were observed on potato in multiple places in Arizona in September, August 1943, and in March 1945 (Hoyman 1943b, c).

New Mexico

One psyllid yellows record with moderate severity from Cloudcroft, Otero County, New Mexico, was observed in September 1943 (Hoyman 1943a).

Texas

Severe psyllid yellows was observed in Texas, during July 1944 (Plant Disease Survey 1944). According to the collected records, Texas had potato psyllid population from 1936. However, locations differed from the area that had psyllid yellows in 1944.

Summary of the psyllid yellows observation in Canada

Alberta

Psyllid yellows epidemics were observed in a few locations in Alberta for multiple years from 1928 to 1944. The main locations were Calgary, Medicine Hat, Edmonton, and Lethbridge. Medicine Hat was the first city in Alberta with a reported potato psyllid population, which was found in a green house near Medicine Hat in 1928 on tomato plants (Strickland 1938). Before 1934, Epidemics with similarities to psyllid yellows were observed in central Alberta, however, the psyllid observed couldn't be determined as *Paratrioza (*currently *Bactericera) cockerelli* (Sanford, 1934).

Saskatchewan

Psyllid yellows was reported in Swift Current in 1938, the first record in Saskatchewan.

Potato psyllid was observed in Melford, Saskatchewan in 1939, with a lower population

than 1938, but no specimen exists (Marritt 1938, 1939).

**Historical observations of the relationship between potato psyllid populations,
psyllid yellows and weather conditions**

Temperature

Temperature and humidity are two main weather factors discussed in relation to potato

psyllid populations, or reports of psyllid yellows in the previous literature. The

descriptions of different weather conditions (usually qualitative: wetter, drier, warm, hot,

or cooler weather) potentially influencing potato psyllid populations or psyllid yellows

are highly regional and vary. Cooler, wetter July, and warmer August and September,

were described as benefitting the psyllid yellows outbreak in Nebraska (Hill 1947).

However, high summer temperature in Nebraska was observed to decrease the potato

psyllid populations in the field (Hill 1947). A study by Hartmen (1937) in Wyoming

from 1932 to 1936 concluded that precipitation was observed to have little effect on the

field potato psyllid populations. Cold winter conditions were believed to inhibit the

potato psyllid eggs from hatching (Hartmen 1937). In the same study, higher temperature

during the last 10 days of June and the first 10 days of July was thought to reduce the

potato psyllid population size and psyllid yellows severity (Hartmen 1937). Lower spring

temperature was found to hinder the potato psyllid population in the North Platte Valley

in Wyoming. Higher temperature in June was related to an increased potato psyllid

population. Higher temperature in July reduced the summer potato psyllid population

size. Early killing frost in September forced potato psyllids to move out of potato fields, because of loss of the food plant. Potato psyllid populations dropped in September when the temperature decreased, or the crop was harvested (Wallis 1946). In Montana, when the summer temperature was around 90 °F (32 °C), psyllid yellows severity was reported to be lower (Wallis 1946; Pletsch 1947). In Tran's (2012) life history study of potato psyllid on potato, development of potato psyllid paused when the temperature reaches 33.6 °C (Tran et al. 2012, Lactin et al. 1999). Conditions of lower latitude areas and cooler regions were noted as possibly more beneficial for the population development. Potato psyllid was described as a low-temperature species in Colorado, saying that higher summer temperature decreased the psyllid yellows severity (List 1939). The cooler, higher mountain areas were more favorable for the disease outbreak (List 1939).

Overwintering

Potato psyllid nymphs were observed feeding on field bindweed after exposure to repeated frost and temperatures as low as 6 °F (-14.4 °C) (Pletsch 1947). Potato psyllid nymphs were observed to survive -15 °C for 24 hours with 100% success. Potato psyllid adults were observed to have 40% survival rate after being exposed to -10 °C for 24 hours when feeding on potato (Henne et al. 2010). Potato psyllid's ability to tolerate winter cold varies regionally. A Nebraska colony was found to be more cold-tolerant than a colony that originated in southern Texas (Whipple et al. 2013). Recently, overwintering on bittersweet nightshade was found in Pacific Northwest locations (Murphy et al. 2013), and composition of the potato psyllid haplotypes varies in the Pacific Northwest regions when the temperature gets colder (Swisher et al. 2013; Swisher

et al. 2014). The northwestern haplotype is more cold-tolerant than the central and western haplotypes of potato psyllid (Swisher et al. 2013).

Moisture

Moderately warm spring conditions with low rainfall and low humidity were said to be favorable for a psyllid yellows outbreak (Wallis 1955). Drier years were observed to generally have more psyllid yellows outbreaks (Wallis 1955). Humidity was thought to inhibit the potato psyllid from spreading eastwards, divided by the 100 meridian that separates the warm and dry area of the United States (Wallis, 1955). However, a study conducted in Montana from 1937 to 1941 showed that higher precipitation seems to benefit the psyllid infestation. High humidity in winter decreased the potato psyllid survival during hibernation (Pletsch 1947). When the temperature is not lethal, the survival rate of fasting potato psyllid individuals increases with the relative humidity. In the experiment reported by Pletsch (1947), under 4 ℃ and 11 ℃, the survival rate of potato psyllid increased with the relative humidity. The longest survival period of non-feeding potato psyllid adults was 92 days at 4 ℃ with 100% relative humidity (Pletsch 1947).

In order to compare how the weather conditions differ between the historical reported locations with psyllid yellows observations and locations with "none", I used two main types of weather indices to describe the weather, which are based on calculations involving annual, monthly total, mean, and extremes (maximum, minimum). The indices are mainly two standard types: BIOCLIM variables that are widely used in ecological studies and species distribution modelling, and CLIMDEX climatic extremes indices that used to describe climatic extremes in climate change study. Annual and monthly mean of

daily maximum, minimum, mean temperatures, annual and monthly total precipitation were calculated to describe the general values and ranges of daily weather variables. Additionally, degree-days accumulation for each month, for each year, from April to October were calculated for comparison based on the results of Tran et al. (2012), a model of potato psyllid developmental rate on potato, using single sine method (daily maximum and minimum temperatures are used to fit a sine function, and the integrated area is the number of degree-days).

Methodology

Collecting psyllid yellows observation data

As noted, psyllid yellows observation records were mainly collected from the periodical Plant Disease Bulletin (1917 - 1922), periodical Plant Disease Reporter (1923 - 1979), and periodical Plant Disease (1980 - 2017) the periodical Canadian Plant Disease Survey (1929 – 2015), described above, and other literature articles. Only the records with a location description (at least county name), clear statement of psyllid yellows observations, and the observation year, were recorded. The localities were georeferenced in Google Map and ArcGIS, using coordinates (town, city, agricultural location), and polygons (county). A total of 434 psyllid yellows records were extracted from these historical documents and literature articles for the northern United States and Canada.

Weather data collection

The Global Historical Climatological Network Daily Weather Dataset version 3.22 (GHCN-DAILY v. 3.22) (Menne et al. 2012; Menne et al. 2016) obtained from National Oceanic and Atmospheric Administration online database was used for calculating the weather for the analysis of the historical psyllid yellows records. Another set of weather

data used for weather indices calculation is the Livneh daily CONUS near-surface

gridded observed meteorological dataset (Livneh et al. 2013), retrieved from the Livneh

research group at National Oceanic and Atmospheric Administration (NOAA) Physical

Sciences Division's online database

(https://www.esrl.noaa.gov/psd/data/gridded/data.livneh.metvars.html).

GHCND-DAILY dataset downloaded files are in fixed-width, *.dly format. Each file

contains the weather data records of one weather station, with variables names, values,

and date. Weather data files were read and stored in HDF5 files using Python. Daily

precipitation (tenths of mm), daily maximum temperature (tenths of degree Celsius),

daily minimum temperature (tenths of degree Celsius), were extracted and divided by

ten. The time coverage of the dataset is dependnt on the available weather data from the

station.

Livneh interpolated daily gridded meteorological data were downloaded in plain text

format (ASCII). The spatial resolution of the dataset is 1/16 degree. The time coverage of

the data is from January 1, 1915, to December 31, 2011. Each plain-text file stores the

daily weather data for one spatial grid from 1915-01-02 to 2011-12-31. The weather data

were read and stored in HDF5 files.

Climatic indices calculation

For the purpose of comparing the climate to the historical psyllid yellows records, annual

and monthly climatic indices were calculated for each recorded observation of psyllid

yellows collected from the historical agricultural plant disease reports, and literature

articles. The weather indices of the point records (observations made in land location,

city or town, etc.) were computed using the weather data from the nearest weather station

within 7 km, or the nearest Livneh (Livneh et al. 2013) weather data point if the weather

station data were not available. The climatic indices for the county records were obtained

by averaging the climatic indices generated using the Livneh weather data of the point

locations that fall within the county polygon. The R package 'dismo'(Hijmans et al.

2017) was used to calculate the BIOCLIM variables. The R package 'climdex.pcic'

(Bronaugh 2015), a package provided by Pacific Climate Impacts Consortium (PCIC),

was used to calculate the extreme climate indices. The accuracy of the calculated indices

was checked and compared with published datasets. The whole process of data

importing, indices calculation, and data storage were programmed in Python (Rossum

2005). The python module, PypeR v. 1.1.2, was used to run R in python environment. It

is a python interface to R language through PIPE (Xia et al. 2010).

BIOCLIM indices

BIOCLIM indices are annual climatic indices generated using monthly maximum

temperature, monthly minimum temperature, and monthly total precipitation (Hijmans et

al. 2005). These indices describe the annual mean, extremes, and seasonality of the

climate (Busby 1991). BIOCLIM indices have been used to quantify the species

environmental niche in a wide range of studies since the 1980's. They are the commonly

used and relevant input variables of species distribution models (Booth et al. 2014;

Kriticos et al. 2014). In one of the earliest such studies, by Booth (1985), 12 BIOCLIM

indices were used to identify the homo-climatic area in Africa that has similar climate

condition with the climatic situation of the natural range of *Eucalyptus citriodora*

(lemon-scented gum tree) in Australia. Later, BIOCLIM indices developed into a set of

19 indices which are commonly used in species distribution modelling (Hijmans et al.

29

2005; Booth et al. 2014). I generated those 19 BIOCLIM indices using the R package, 'dismo', for the psyllid yellows observations (see table below), and checked their calculation accuracy with additional programming and statistical summaries. The calculation accuracy was tested by comparing with the calculation of the WorldClim dataset (Hijmans et al. 2005).

Table 2.1. The 19 BIOCLIM indices available from WorldClim database (Hijmans et al. 2005)

Index	Description	Unit
BIO1	Annual Mean Temperature	°C
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	X
BIO3	Isothermality (BIO2/BIO7) (* 100)	X
BIO4	Temperature Seasonality (standard deviation *100) *standard deviation of the monthly mean temperature	X
BIO5	Max Temperature of Warmest Month	°C
BIO6	Min Temperature of Coldest Month	°C
BIO7	Temperature Annual Range (BIO5-BIO6)	°C
BIO8	Mean Temperature of Wettest Quarter	°C
BIO9	Mean Temperature of Driest Quarter	°C
BIO10	Mean Temperature of Warmest Quarter	°C
BIO11	Mean Temperature of Coldest Quarter	°C
BIO12	Annual Precipitation	mm
BIO13	Precipitation of Wettest Month	mm
BIO14	Precipitation of Driest Month	mm
BIO15	Precipitation Seasonality (Coefficient of Variation)	X
BIO16	Precipitation of Wettest Quarter	mm
BIO17	Precipitation of Driest Quarter	mm
BIO18	Precipitation of Warmest Quarter	mm
BIO19	Precipitation of Coldest Quarter	mm

CLIMDEX indices

Climate change indices, CLIMDEX, describes annual and monthly temperature and precipitation extremes. They are used in climate change studies to analyze the changes in climatic extremes related to global warming (Alexander et al. 2006). It facilitates a better understanding of climatic extremes. BIOCLIM indices generally describe the seasonality, quarterly mean and annual mean and extreme of the climate computed by monthly temperature and precipitation data. CLIMDEX (http://www.climdex.org/) provides more detailed monthly and annual extremes measures and easier to understand indices, calculated from daily temperature and precipitation data. One of the CLIMDEX indices, the number of frost days, describes annual count of days when the daily minimum temperature is less than zero Celsius degree. Monthly maximum 5-day precipitation, another CLIMDEX index, describes the monthly maximum of the continuous five day precipitation total. Growing season length defines the growing season length based on daily minimum temperature. Moreover, monthly extremes are computed by taking the maximum or minimum values of the daily temperature data of the month (Karl et al. 1999). The computed CLIMDEX indices are listed in the table below.

Table 2.2. The CLIMDEX indices calculated for the Random Forest model fitting
(http://www.climdex.org/indices.html)

Variable	Description	Unit
FD	Number of frost days: total number of days when daily minimum temperature is smaller than 0 ℃ during the year	Day
SU	Number of summer days: total number of days when daily maximum temperature is bigger than 25 ℃ during the year	Day
ID	Number of icing days: total number of days when daily maximum temperature is smaller than 0 ℃ during the year	Day
TR	Number of tropical nights: total number of days when daily minimum temperature is bigger than 20 ℃ during the year	Day
GSL	Growing season length: the numbers of days between the first six consecutive days with mean temperature bigger than 5 ℃ and the first six consecutive days with mean temperature smaller than 5 ℃	Day
TXX	Monthly maximum value of daily maximum temperature	℃
TNX	Monthly maximum of daily minimum temperature	℃
TXN	Monthly minimum of daily maximum temperature	℃
TNN	Monthly minimum of daily minimum temperature	℃
RX1DAY	Monthly maximum 1-day precipitation	mm
RX5DAY	Monthly maximum 5-day total precipitation	mm
SDII	Simple precipitation intensity index, mean precipitation calculated by the monthly total precipitation divided by the number of days with precipitation	mm
R10MM	Annual count of days when daily precipitation is bigger than 10 mm	mm
R20MM	Annual count of days when daily precipitation is bigger than 20 mm	mm
R30MM	Annual count of days when daily precipitation is bigger than 30 mm	mm

CDD	Maximum length of dry spell, maximum number of consecutive days with precipitation smaller 1mm	Day
CWD	Maximum length of wet spell, maximum number of consecutive days with precipitation bigger than 1 mm	Day
PRCPTOT	Annual total precipitation in wet days	mm

Random Forest: psyllid yellows records of the northern United States and Canada

Random Forest model

Random Forest (RF) is a non-parametric modelling approach, an ensemble model of multiple de-correlated Classification and Regression Trees (CART) that provides a result by averaging the votes or predictions of the different classifiers, called "trees", within the "forest." Each tree within the Random Forest is fitted by a randomly sampled subset of the full dataset. The subsamples are sampled from the data records with replacement, which is called bootstrapping. In other words, it generates a certain number N(T) of trees by drawing a certain number n(boot) of samples with replacement from the subset of variables from the input dataset (Breiman 2001; Cutler et al. 2007; Plant 2012). Because of the randomness (repeated data subsets) of the Random Forest model fitting procedure, it addresses the challenges of population randomness, spatial autocorrelation among observations, the non-linear relationship, collinearity among predictors, that the parametric models fail to manage (Murphy et al. 2010).

Classification and Regression Tree is a machine learning algorithm that recursively partitions data into two homogeneous groups on the response variable. It has no assumption of the data probability density distribution. It can model complex relationship other than a linear relationship (Plant 2012).

Random forest is an ensemble of multiple Classification and Regression Trees. Fitted by randomly sampled data and randomly selected variables from the input dataset, each tree is a weak classifier and de-correlates with other classifiers. (Cutler et al. 2007). Random Forest outperforms the Classification and Regression Tree, by aggregating the results of different trees fitted by part of the training data, which avoid the issue of model overfitting the data (Cutler et al. 2007). Random Forest can have hundreds or thousands of trees, each based on a randomly selected subset of the data (Breiman 2001)

Advantages of using Random Forest model

Random Forest has advantages in modelling complex systems of different ecological relationships, compared to the linear regression models. RF is flexible and able to model complex, non-linear relationships between variables (Cutler et al. 2007; Murphy et al. 2010). Random Forest has no assumption of the data probability distribution, nor does it face restrictions of assumptions of normality and independence among the predictors, as required by the linear regression model. In contrast to the typical linear regression model, the Random Forest model is not sensitive to data collinearity, the degree to which predictors are correlated (Murphy et al. 2010). Collinearity among different climatic variables varies among different eco-zones (Dormann et al. 2013). Random Forest outperforms the linear models when dealing with collinear data (Cutler et al. 2007; Dormann et al. 2013). Random Forest model does not need to exclude collinear variables, which allows the algorithm to take in more information when training the model. However, a smaller number of variables in the model is an advantage because it means less computation time, which can be challenging for large datasets. Later in this chapter, a procedure to reduce the number of variables will be introduced, applied, and discussed.

Random Forest is less sensitive to data perturbations and irregularities than classification

and regression trees because the estimate of the response variable is the average of the

estimates of the trees generated by a different subset of data (Plant 2012).

Random Forest can be used to rank the importance of the variables by calculating the

percentage increase in MSE (mean squared error) over all trees using the out of the bag

cases, as a result of rearranging the values of the evaluated variable. Calculation of the

increase in node purity (or impurity) is another way to assess the importance of the

variable. An increase in node purity means the variable is more important. Node impurity

is a measure, usually given as the Gini index, of how the trees split the data. Accuracy

(correct classification) can also be used to assess the importance of variables. Percent

increase in MSE is considered as a more accurate indication of the variable importance

than the increase in node purity (Plant 2012).

Limitations of using Random Forest model

The Random Forest model is a kind of "black box" algorithm, an aggregation of multiple

weak classifiers. As it is a combination of different CART models, its interpretation is

less straightforward than linear model or CART alone. However, partial dependence

plots can visualize the relationship between the explanatory variable of interest and the

response variable (Cutler et al. 2007).

Application of Random Forest in ecological predictive modelling

The Random Forest can be used for predictive modelling in Ecology (Cutler et al. 2007).

Wang et al. (2016) used Random Forest model to analyze the relationship between

climate and the occurrences of three Chinese tree species, Chinese fir (*Cunninghamia*

lanceolate (Lamb.) Hook), Masson Pine (*Pinus massoniana*) and Chinese pine (*Pinus*

tabuliformis Carr.). The fitted Random Forest model was used to project distributions of those tree species based on future climate under different climate change scenarios (Wang et al. 2016).

Variable reduction with Random Forest

I generated 123 climatic variables, indices indicating aspects of weather and climate, using daily maximum temperature, daily minimum temperature, and daily total precipitation. The number of predictors and the combination of predictors will influence Random Forest model performance, so this list was reduced to the key variables through a variable reduction procedure. More predictors used in the model require more calculation time, and more model computation time. Random Forest is not sensitive to the irrelevant variables in the model (Svetnik et al. 2004; Heung et al. 2014). The reduction of irrelevant variables does not necessarily improve the model performance (Heung et al. 2014), but reduces the time required for model fitting. Increasing the number of predictors does not always improve the model performance (Scarpone et al. 2017). However, Random Forest model performance and the number of input variables can be balanced by a systematic variable reduction procedure developed by Diaz (Diaz-Uriarte & de Andrés 2005). This method was used to reduce the number of model input variables, for classification of the mountain land cover (Scarpone et al. 2017), to improve the pseudo-absence sampling of species distribution modelling (Senay et al. 2013), for the selection of critical disease factors for disease risk modelling (Zhou et al. 2012). The procedure is available in the R package, "varSelRF" (Diaz-Uriarte 2007).

The variable reduction is a backward elimination procedure that selects the smallest set of explanatory variables with the lowest out-of-bag (OOB) error rate (Svetnik et al.

2004). OOB error is the average prediction error of training samples. The more detailed

procedure used in this study is described below (Svetnik et al. 2004; Zhou et al. 2012;

Scarpone et al. 2017).

1. Fit the full dataset to the Random Forest model, and use the variable importance

   ranking based on the decrease of accuracy as the order for the upcoming variable

   reduction process.

2. Drop the least important variable based on the ranking, and fit the resulting

   dataset to a new Random Forest model, estimate the OOB error of the fitted

   Random Forest model.

3. Continue step 2, until only two variables remain in the fitted forest

4. The smallest set of variables with the OOB rate within one standard error of the

   OOB rate of the full-variable model will be selected.

I used this variable reduction procedure to determine the input predictors of the Random

Forest model predicting the relationship of psyllid yellows historical records to climatic

indices. The variable reduction procedure reduces the number of variables per set one by

one, compared to the out-of-bag error of the full model. The variable reduction procedure

determines the smallest set of 25 variables with the out-of-bag error within the one

standard error range. The selected variables were then used to fit the Random Forest

model. The selected variables are listed below.

Annual mean temperature (BIO1)

Mean temperature of coldest quarter (BIO11)

Mean diurnal range (mean of monthly (max temp - min temp)) (BIO2)

Monthly total degree days accumulation in April (DG_P_4)

Monthly total degree days accumulation in June (DG_P_6)

Monthly total degree days accumulation in September (DG_P_9)

Annual count of frost days (the day when minimum temperature is smaller than 0 Celsius degree) (FD)

Monthly maximum 1-day precipitation in May (RX1DAY_05)

Monthly maximum 5-day precipitation in May (RX5DAY_05)

Number of summer days: annual count of days when daily maximum temperature > 25 Celsius degree (SU)

Monthly mean of daily maximum temperature in April (TMAX_04)

Monthly mean of daily maximum temperature in June (TMAX_06)

Monthly mean of daily maximum temperature in September (TMAX_09)

Monthly mean of daily average temperature in June (TMEAN_06)

Monthly mean of daily average temperature in September (TMEAN_09)

Monthly mean of daily minimum temperature in June (tmin_06)

Monthly mean of daily minimum temperature in September (TMIN_09)

Monthly minimum of daily minimum temperature in September (TNN_09)

Monthly total precipitation in May (TOTAL_PREC_05)

Monthly total precipitation in June (TOTAL_PREC_06)

Monthly minimum daily maximum temperature in April (TXN_04)

Monthly minimum daily maximum temperature in May (TXN_05)

Monthly minimum daily maximum temperature in June (TXN_06)

Monthly minimum daily maximum temperature in September (TXN_09)
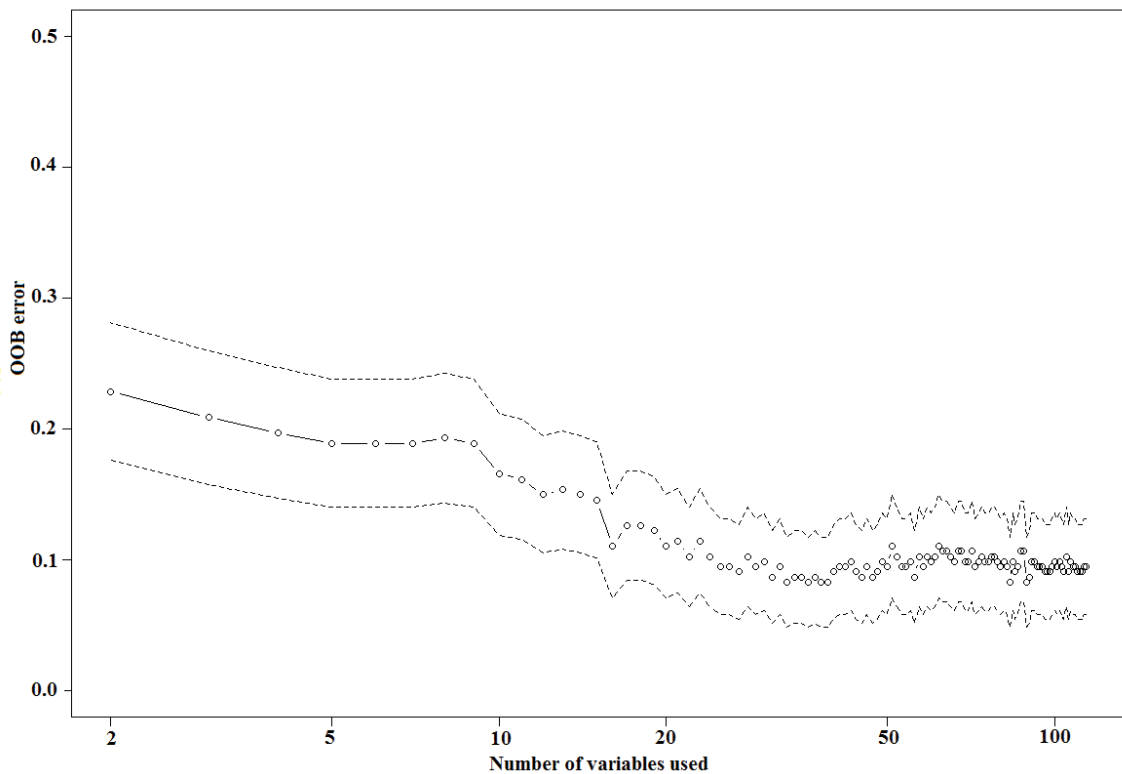
Figure 2.2 Backward variable reduction process, out-of-bag error (OOB) decreases with the increase of more variables used; the OOB error no long increase after 25 variables used to fit the Random Forest model.

Random Forest model parameter tuning using R 'Caret' package

The optimal setting of a number of trees and number of candidate variables of each node

for the queried dataset after variable reduction is decided by using the model tuning

40

procedure of R package, "Caret" (Kuhn 2016). The algorithm resamples training and test data for each tuning parameter combination for multiple times. Each set of the resampled training data will be used to fit a Random Forest model, and the model predictive accuracy will be estimated by the test dataset. For each combination of model parameters, the model accuracy is evaluated by 10-fold cross-validation, repeated three times. The overall resampled accuracy estimate is used to determine the optimal tuning parameters combination. The determined mtry is 13 (the number of variables randomly sampled as candidates for retention in the RF model), and the ntree is 1000 (the number of trees to grow in the RF). This selected set of parameters will be used in Random Forest fitting.
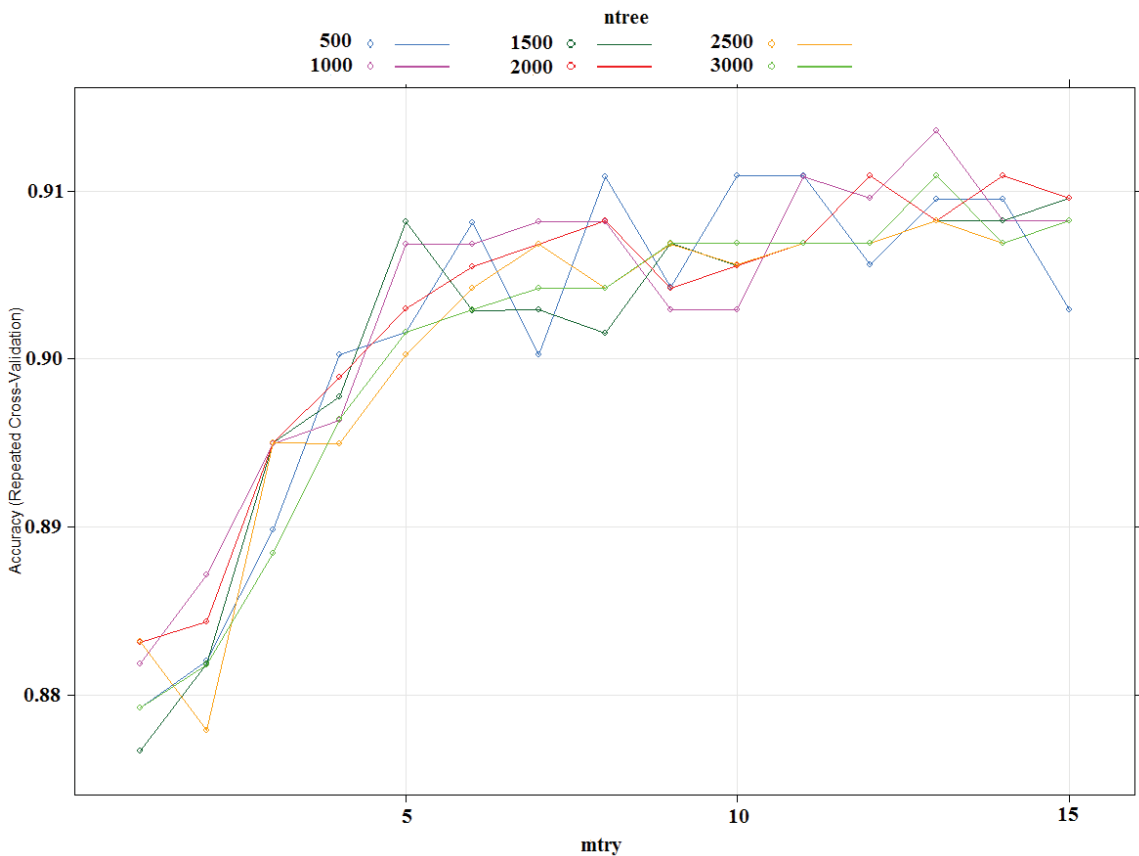
Figure 2.3. Model accuracy increases with different number of candidate variables at each tree node (mtry), and different number of trees in the Random Forest (ntree)

41

Random Forest model fitting and model validation

Different training sets generate different Random Forest models using the optimal model parameters from model tuning. To find the Random Forest model that fits the data the best, I repeatedly split the training and testing dataset 100 times (ratio: 70 train /30 test), and selected the Random Forest with the lowest out-of-bag error estimate and reasonable predictive accuracy, and AUC value. Boxplots were fitted with the full dataset. Partial dependence plots were fitted with the training dataset, based on the Random Forest model. It graphically describes the relationship between the logit (fraction of yes votes) and the predictor value (Cutler et al. 2007). The final model was be chosen based on the OOB value, AUC value, and the similarity between the boxplot and the partial dependence plot.

OOB error is a standard value used to show how well the model fits the training data. The out-of-bag error is the misclassification rate of the out-of-bag subset left out when fitting the Random Forest model. OOB error is the model predictive accuracy estimate for Random Forest model (Friedman et al. 2001). The smaller the OOB error is, the better the model fits the training data.

The area under the ROC (receiver-operating characteristic) curve, AUC, is widely used as a species distribution model (SDM) performance measure. The ROC curve is a graph of false-positive error rate as its x-axis (1-Specificity) and true positive error rate as its y-axis (Sensitivity) (Franklin & Miller 2009). The true positive rate (TPR), is the actual detection of the presence of the measured variable, for example, disease. The false-positive error is the probability of concluding a difference or effect when it is not justified i.e., a false alarm. The area accumulated under the ROC curve is calculated as

the AUC value, which ranges from 0 – 1.0. When AUC is 0.5, the chance of the prediction having accurate detection is the same as the false alarm, no better than random. When the AUC is bigger than 0.5, the model prediction is better than random (Franklin & Miller 2009). The model performance with AUC ranging from 0.7 to 1.0 is considered to be acceptable (Franklin & Miller 2009).

Results and discussion

The final selected model fits the dataset well, with an AUC rate of 91.9%, and on the out-of-bag error of 9.04%. Among the BIOCLIM indices that describe the seasonality of the climate, annual mean temperature (BIO1), mean diurnal range (BIO2), and mean temperature of the coldest quarter (BIO11) were selected as part of the model input. In the preliminary analysis of the psyllid yellows observations dataset, the observations of psyllid yellows presence (observed and reported) have higher mean BIO1 value than that of psyllid yellows absence (observed to be absent). The cases with psyllid yellows reported tend to have higher annual mean temperature than the cases without psyllid yellows.

Psyllid yellows development on potato is related to potato psyllid nymph feeding on potato plants. The severity of symptoms is related to the duration of feeding, and the number of feeding individuals. The feeding of potato psyllid adults is believed to have no impact on psyllid yellows development (Richards & Blood 1933).

The warmer general annual climate possibly means the weather is less limiting for potato psyllid survival in the winter months and spread during warmer months, which potentially provides more climatic favorable conditions for potato psyllid propagation. The partial dependence plots visualize the relationship between the response and the

predictor for the Random Forest model. In this classification study, the y-axis is the logit of the fractions of votes for presence by the trees in the Random Forest. The x-axis is the predictor of interest. As noted, when the fraction of votes is bigger than 0.5, the final prediction of Random Forest model will be considered as a positive prediction of presence. When the fraction of yes votes is equal to 0.5, the logit will be equal to zero, which means the prediction is neither positive nor negative. When the fraction of yes votes is bigger than 0.5, the logit will be larger than zero; when the fraction of yes votes is smaller than 0.5, the logit will be smaller than zero.

The partial dependence plot for BIO1 is found in figure 2.4. To interpret the relationship between BIO1 and presence and absence of psyllid yellows, one must find the cut-off point where the logit is equal to zero and the BIO1 value where the logit starts to rise. In the boxplot of BIO1, the overlapping BIO1 values range from five to seven, between the positive cases (occurrence of psyllid yellows, py = 1) and the negative cases (py = 0). The logit is greater than zero with BIO1 ranging from two to four, and the partial plot shows a positive relationship. etween the logit and an annual mean. Between 4 ℃ and 6 ℃, and the logit is smaller than zero. When the value of BIO1 is close to 6 ℃, the logit of the yes vote fraction rises above zero and keeps increasing until BIO1 is greater than 10 ℃. When the BIO1 is greater than 10 ℃, the chance of having positive psyllid yellows prediction remains the same. The partial dependence plot and the boxplot of BIO1 suggest a positive relationship between annual mean temperature and psyllid yellows incidence.
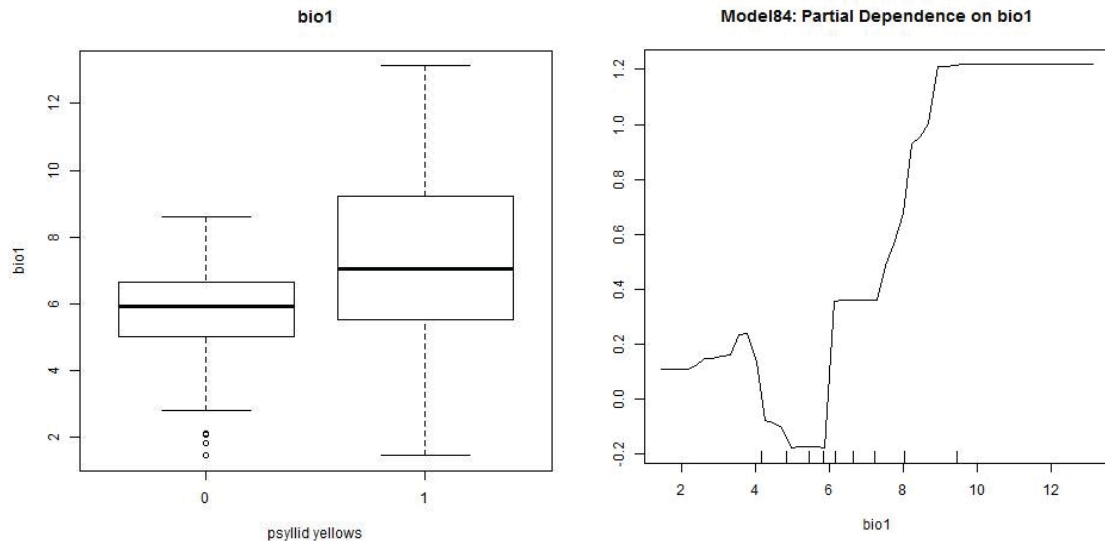
Figure 2.4. Boxplot and partial plot of BIO1, mean annual temperature (°C)

Similar to the pattern of annual mean temperature (BIO1), observations of psyllid

yellows outbreaks have a higher mean temperature of the coldest quarter (BIO11), than

the observations with no psyllid yellows. In the boxplot of BIO11, the psyllid yellows

incident has a higher mean BIO11 value than that of the no psyllid yellows incidents. In a

partial dependence plot of BIO11, the probability of predicting psyllid yellows as

positive starts to be bigger than 0.5 when BIO11 is around minus five °C. Warmer winter

months increase the winter survival of potato psyllids. Recent potato psyllid

overwintering studies show different potato psyllid haplotypes have different cold

tolerance. The winter survival of potato psyllid will decrease when the winter

temperature is lower (Swisher et al. 2013; Horton et al. 2014; Swisher et al. 2014; Horton

et al. 2015).
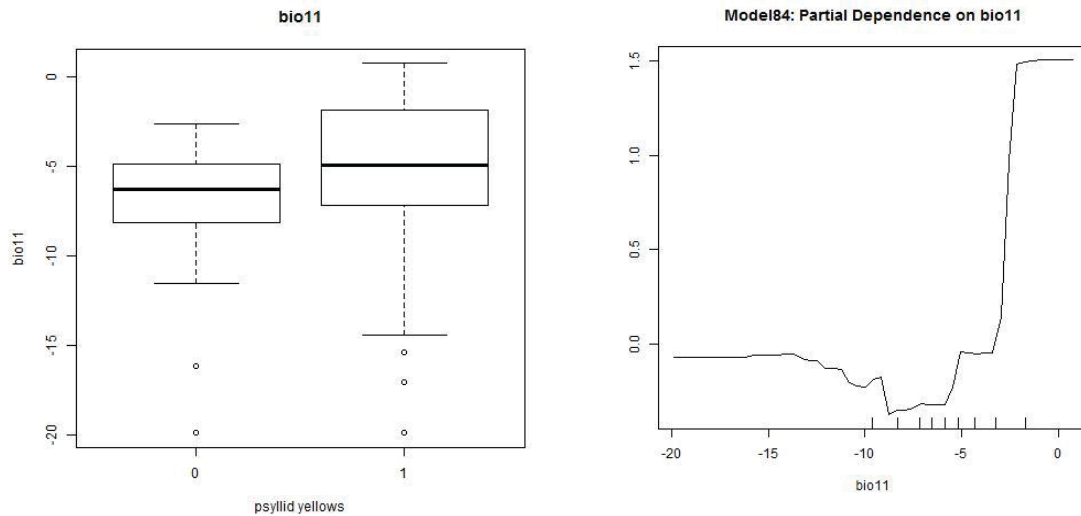
Figure 2.5. Boxplot and partial plot of BIO11, mean temperature of the coldest quarter (°C)

Mean diurnal range (mean of monthly (max temp – min temp)), BIO2 does not distinguish well between the psyllid yellows presence cases and psyllid yellow absence cases.

The first 8 significant variables indicated by the model are: monthly minimum of daily minimum temperature in September (TNN_09), minimum daily maximum temperature in June (TXN_06), mean temperature of coldest quarter (BIO11), minimum daily maximum temperature in April (TXN_04), minimum daily maximum temperature in May (TXN_05), maximum 5-day precipitation in May (RX5DAY_05), total precipitation in May (TOTAL_PREC_05), maximum one-day precipitation in May (RX1DAY_05). The variable importance selection shows that the precipitation in May, the temperature in April, May, June, September and mean temperature of the coldest quarter have higher influence on the model predictive accuracy of the psyllid yellows than annual mean temperature and number of summer days.
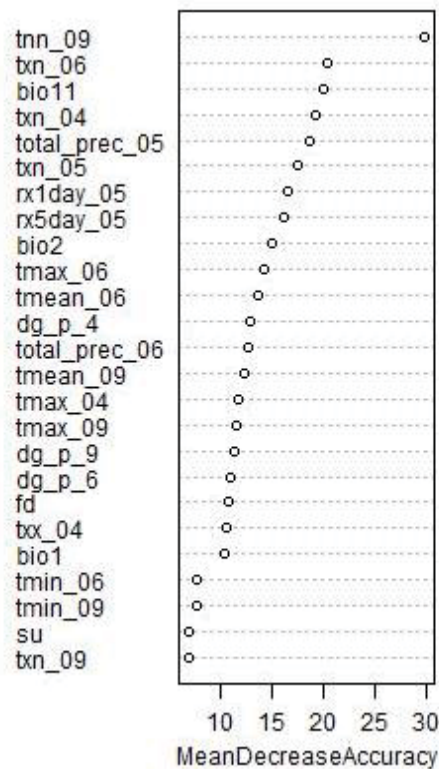
Figure 2.6. Variable importance plot of the fitted Random Forest model

The psyllid yellows incidents have higher precipitation in May and larger precipitation

extremes than the no psyllid yellows incidents. The total precipitation in May

(TOTAL_PREC_05), and other two precipitation indices in May (RX1DAY_05,

RX5DAY_05), are indicated by the Random Forest model as significant indicators of

climate associated with psyllid yellows. They describe the May precipitation

characteristics altogether. When monthly total precipitation is around 70 mm, the

probability of psyllid yellows is around 0.5. The likelihood of psyllid yellows increases

when the monthly precipitation in May increases. These relationships may not reflect

fundamental processes or requirements for the life of the insect, but could indicate

weather and climate conditions relevant to the plant location, timing, and growth. Indices

calculated based on daily precipitation data, RX1DAY_05 and RX5DAY_05 describe the

monthly precipitation extremes in May. RX1DAY_05 describes the monthly maximum

1-day precipitation in May, and RX5DAY_05 describes the monthly maximum 5-day

precipitation in May. The psyllid yellows positive cases have higher monthly

precipitation extremes than the psyllid yellows negative cases. The mean of the

maximum 1-day precipitation of the psyllid yellows positive cases is 22.34 mm, the

average of the maximum 1-day precipitation of the psyllid yellows negative cases is

13.34 mm. The mean of the maximum 5-day precipitation of the psyllid yellows positive

cases is 45.66 mm, the mean of the maximum 5-day precipitation of the psyllid yellows

negative cases is 31.16 mm. In the partial dependence plot of RX1DAY_05, where

rx1day_05 is around 15 mm, the logit of the probability is zero. The logit of probability

increases with RX1DAY_05 value. The logit remains stable above 1.0 when the

rx1day_05 is equal to 60 mm. Similarly, the logit of RX5DAY_05 (monthly maximum

5-day precipitation in May) is less than zero (so less than 50% of the votes of the RF)

where RX5DAY_05 is around 20 mm. The logit increases with increasing RX5DAY_05.
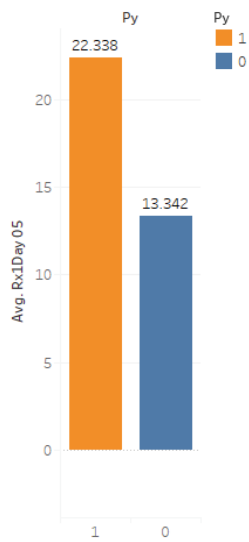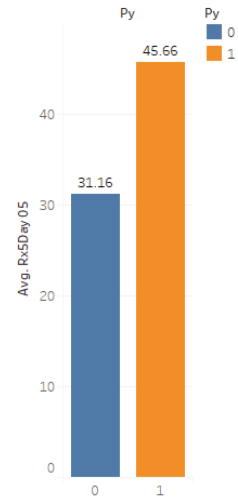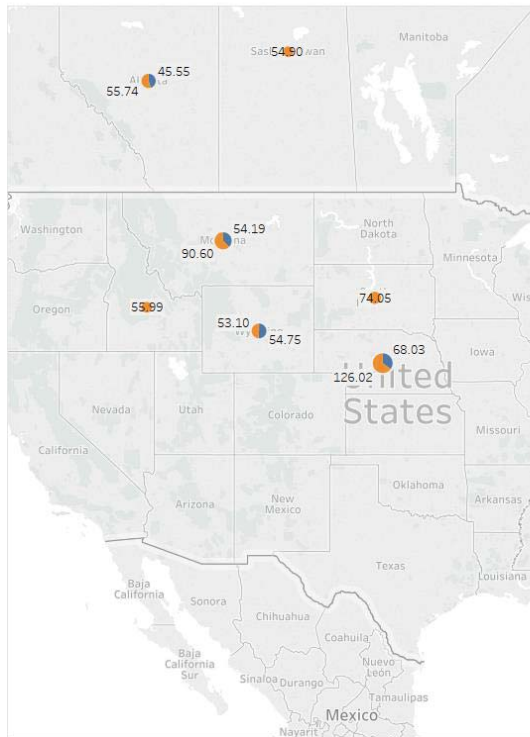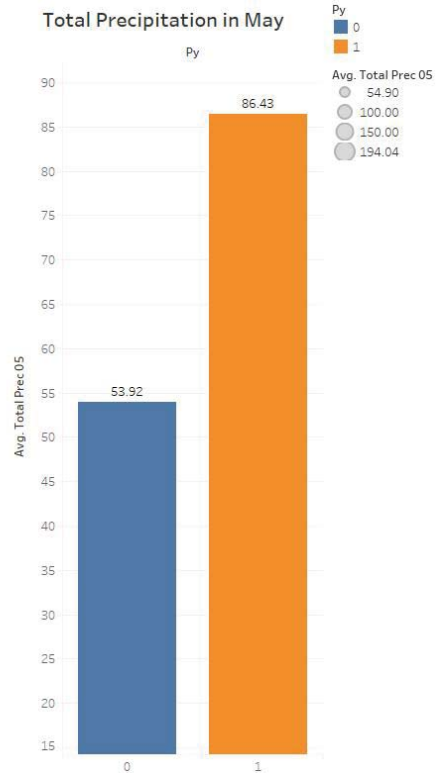
Figure 2.7. Comparing the means of the precipitation indices between the psyllid yellows incidents and no psyllid yellows observations (TOTAL_PREC_05, RX1DAY_05, RX5DAY_05)
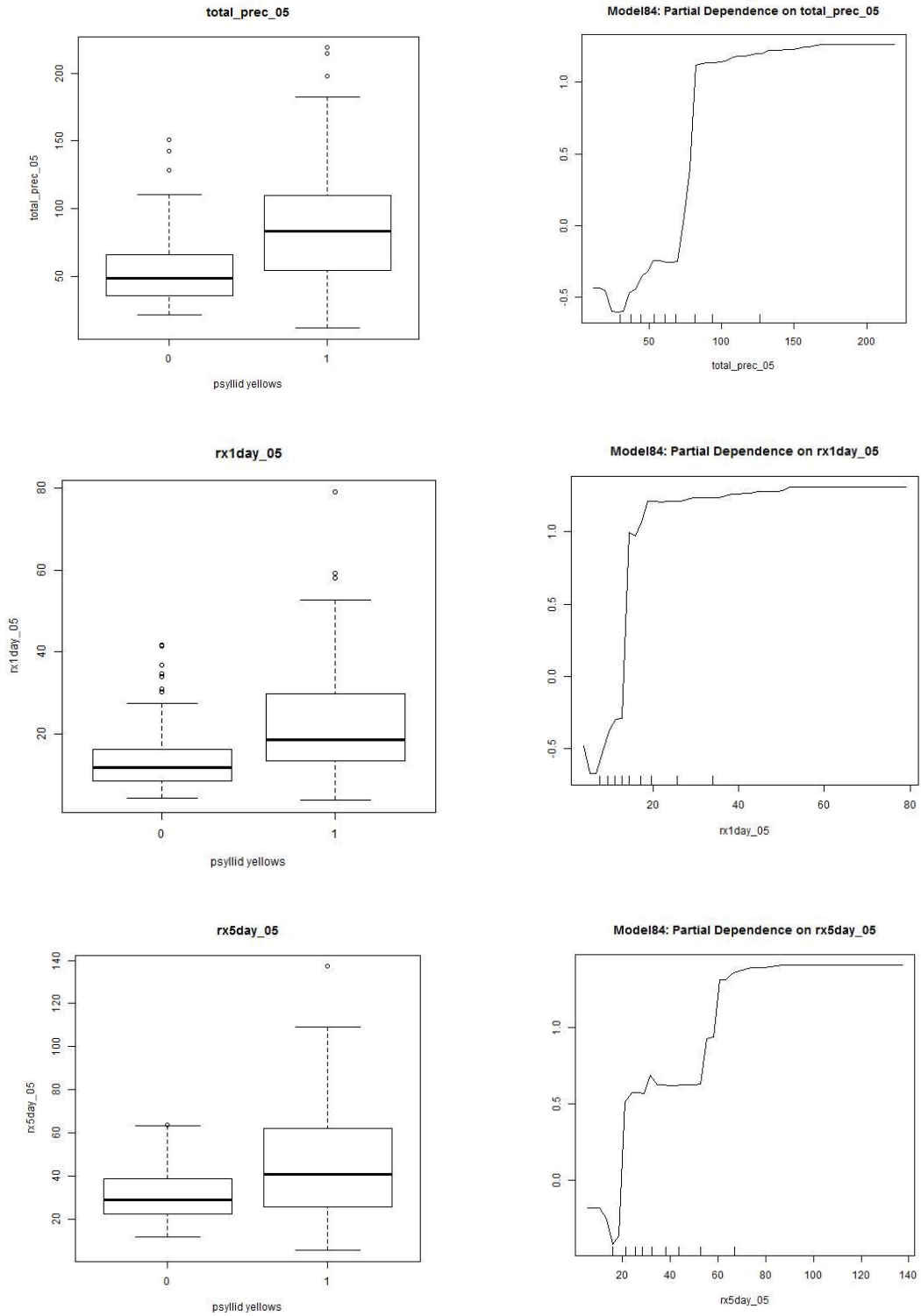
Figure 2.8. Comparing the means of the precipitation indices between the psyllid yellows incidents and no psyllid yellows observations using boxplots and partial plots (TOTAL_PREC_05, RX1DAY_05, RX5DAY_05)

Monthly temperature extremes, TXN_04, and TXX_04 vary between the positive and negative psyllid yellows cases. TXN_04, the monthly minimum of daily minimum temperature in April, is the fourth important variable determined by the model. TXN_04, the April minimum of daily maximum temperature, is higher when py equals to zero. The minimum TXN_04 (py = 1) is -11.1 °C, which is not lethal to the early spring potato psyllid population. Potato psyllid nymphs are more cold tolerant than the adults, and can survive -15 °C for 24 hours, whereas adult survival is less than half when exposed to -10 °C for 24 hours (Henne et al. 2010). TXX_04, the April maximum of daily maximum temperature is slightly higher when py =1, with means equal to 26.15 °C (py = 1), and 24.59 °C (py = 0).

Psyllid yellows positive cases have slightly higher degree days accumulation in April (DG_P_4) than no psyllid yellows incidents. In the partial dependence plot of DG_P_4, the logit is smaller than zero when DG_P_4 is lower than 90 DD. Between the value of 50 DD and 100 DD, the logit of disease probability increases with DG_P_4. When the value of DG_P_4 reaches to 100 DD, logit remains stable. The values of TMAX_04, April mean daily average temperature in April, has bigger value range when py equals to zero than that when py equals to one. In the partial dependence plot of TMAX_04, where TMAX_04 is within the range of 13 °C and 19 °C, the logit is bigger than zero. Moreover, the similar pattern shows in the boxplot of TMAX_04.

txn_04

Model84: Partial Dependence on txn_04

dg_p_4

Model84: Partial Dependence on dg_p_4

tmax_04

Model84: Partial Dependence on tmax_04
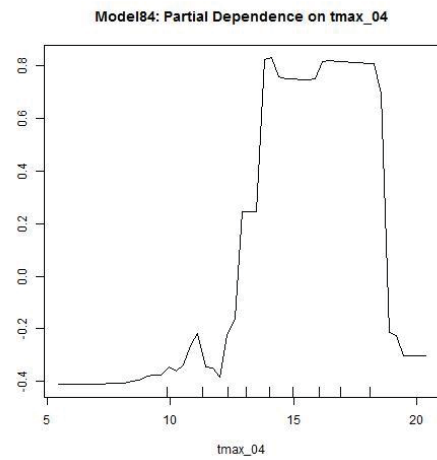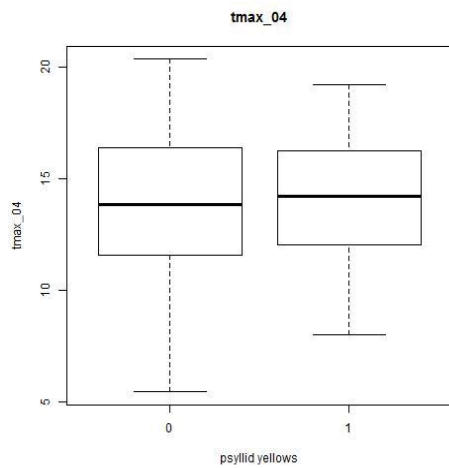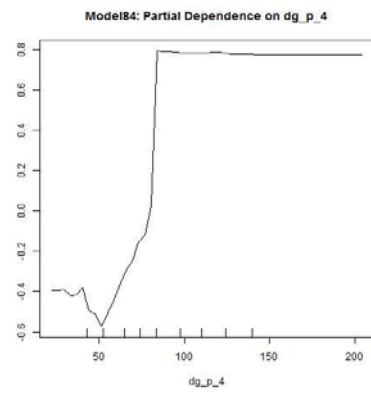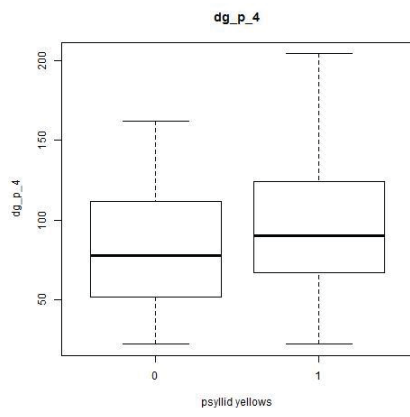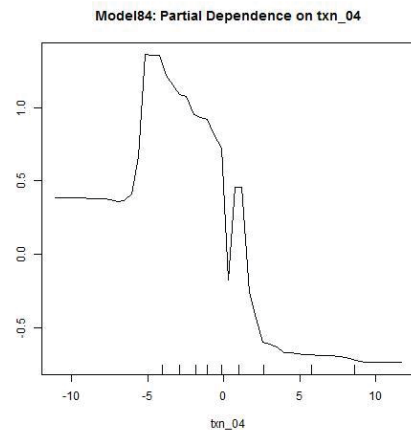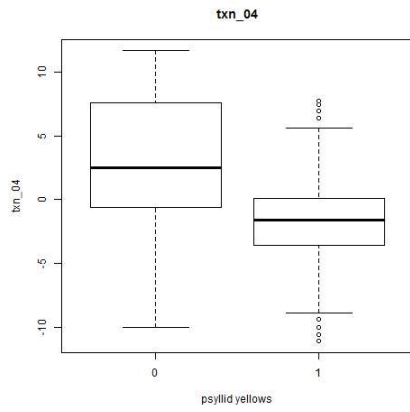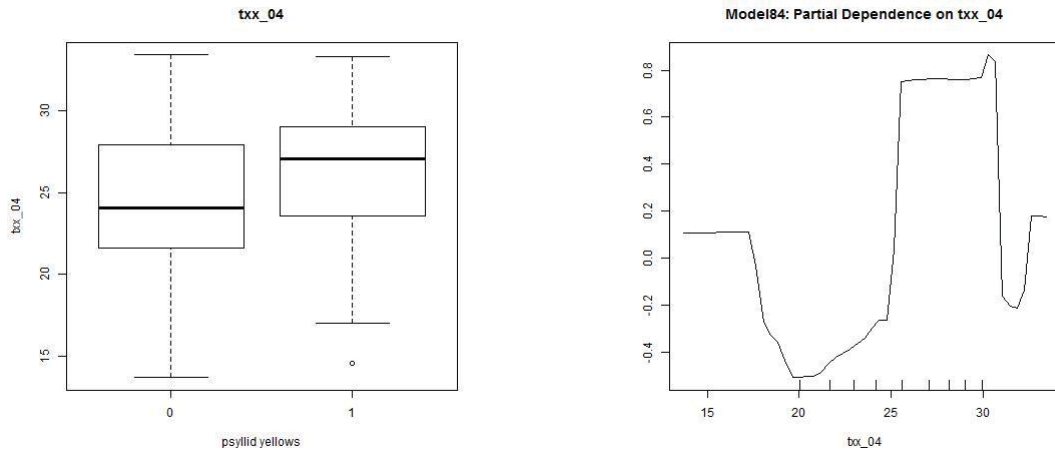
Figure 2.9. Comparing degree-days and temperature indices in April between the psyllid yellows presence and absence cases (DD)
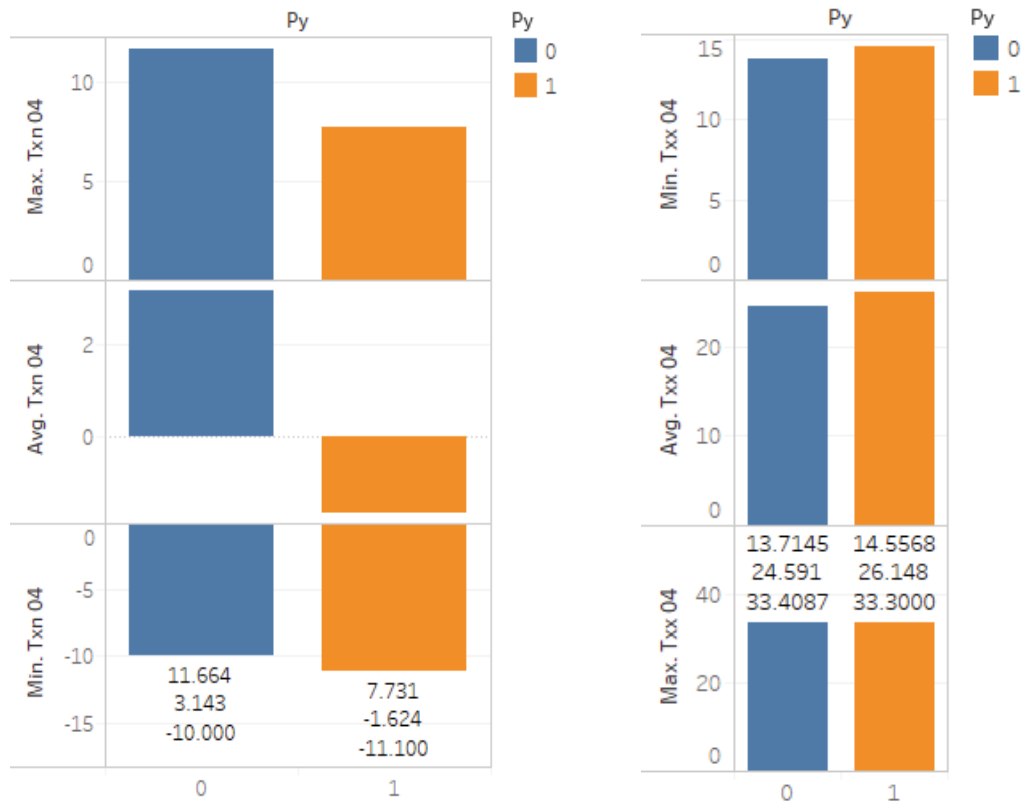
Figure 2.10. Comparing April temperature extremes, TXN_04, and TXX_04 (°C)

The temperature in June has a positive influence on the probability of psyllid yellows incidents. The related indices, TMIN_06, TXN_06, TMAX_06, TMEAN_06, all have higher values when py is equal to one. TMIN_06 is the June mean daily minimum temperature. TXN_06 is the June minimum of daily maximum temperature. TMAX_06 is the mean daily maximum temperature in June. TMEAN_06 is the June monthly mean temperature. Cases with psyllid yellows equal to one generally have a warmer June than the cases with no psyllid yellows equal to zero. Because insects are ectotherm, the ambient temperature has a direct influence on body temperature. Temperature has a significant influence on insect developmental rate (Tran et al. 2012). Developmental rates and degree-days are two measures that describe the relationship between the temperature and insect development. Within a certain temperature range, between the lower developmental threshold and the optimum developmental threshold, insect developmental rate increases with increasing temperature. Based on the laboratory results in Tran et al. (2012), potato psyllids start to grow and develop under the temperature of 7 ℃. The development rate increases with temperature until the temperature reaches the optimum threshold (27.6 ℃) of the insect and decreases when temperature passes the 27.6 ℃ threshold. The development ceases, when the temperature rises above 33.6 ℃. The degree days requirement of potato psyllid to finish one life cycle, from egg to propagation, is 358 degree-days when feeding on potato. In the data below, the mean June degree days accumulation for py = 1 and py = 0 are lower than that in the states in the southern USA. It shows lower degree days do not necessarily mean lower psyllid yellows risk. Based on the comments recorded in the Canadian Plant Disease Survey (1927 - recent), a widely acknowledged publication of plant disease records in information periodical across Canada, infestations of psyllid yellows in 1938 in Alberta

and Saskatchewan could be (?might have been) the northern expansion of the Montana

psyllid population, where psyllid yellows was widespread and caused a 25% loss of

potato crops during the same year. Potato psyllid can travel by hitchhiking the plant

materials (potato, tomato, eggplant, *etc*.) during transportation (Munyaneza 2015). The

rapid increase of the potato psyllid population has been observed in the field by Wallis

(1955). Among literature articles, the opinion that potato psyllid can spread across

different states seems to be a consensus. The 100 meridian was used as a border line that

defines the spread limitation due to humidity tolerance of the insect (Wallis 1955). In

Glick's study (1939), insects from different orders and families including potato psyllid

(dated scientific name*: Paratrioza cockerelli*) were collected during airplane flights in

Mexico (Glick 1939). Considering all those factors and records, long distance dispersal

can not be ignored as a candidate cause of the psyllid yellows incidents in Canada in

1938.
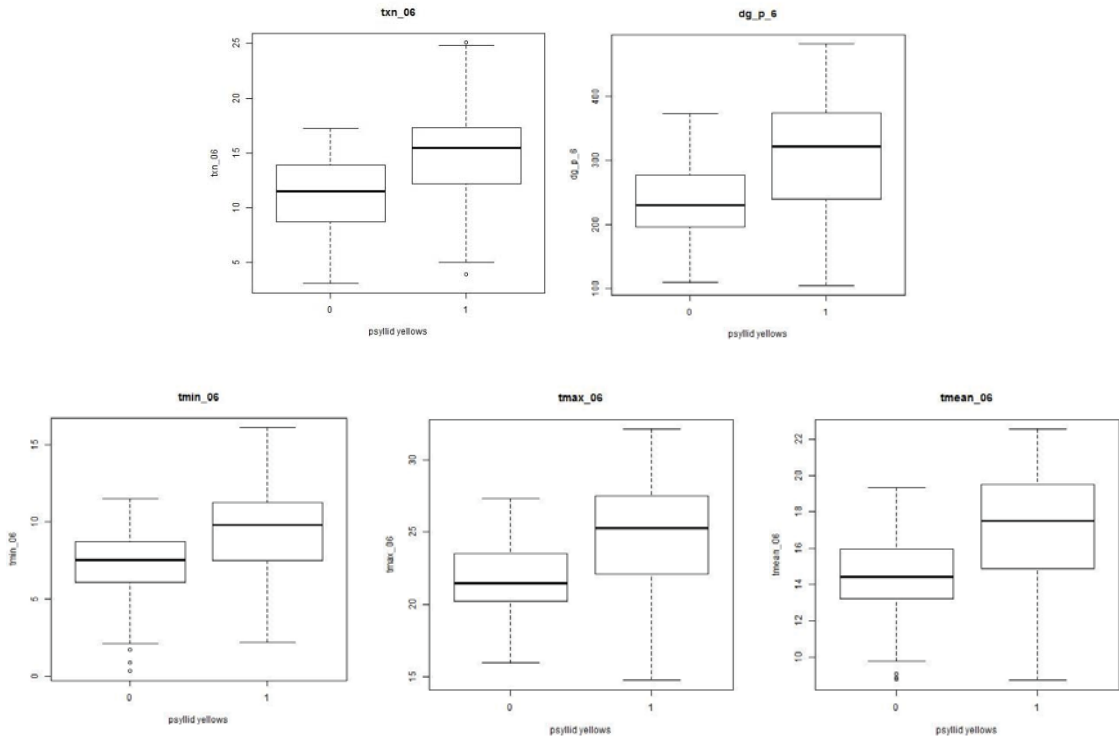
Figure 2.11. Temperature indices and degree days in June, compared between the psyllid yellows positive observations and the no psyllid yellows observations
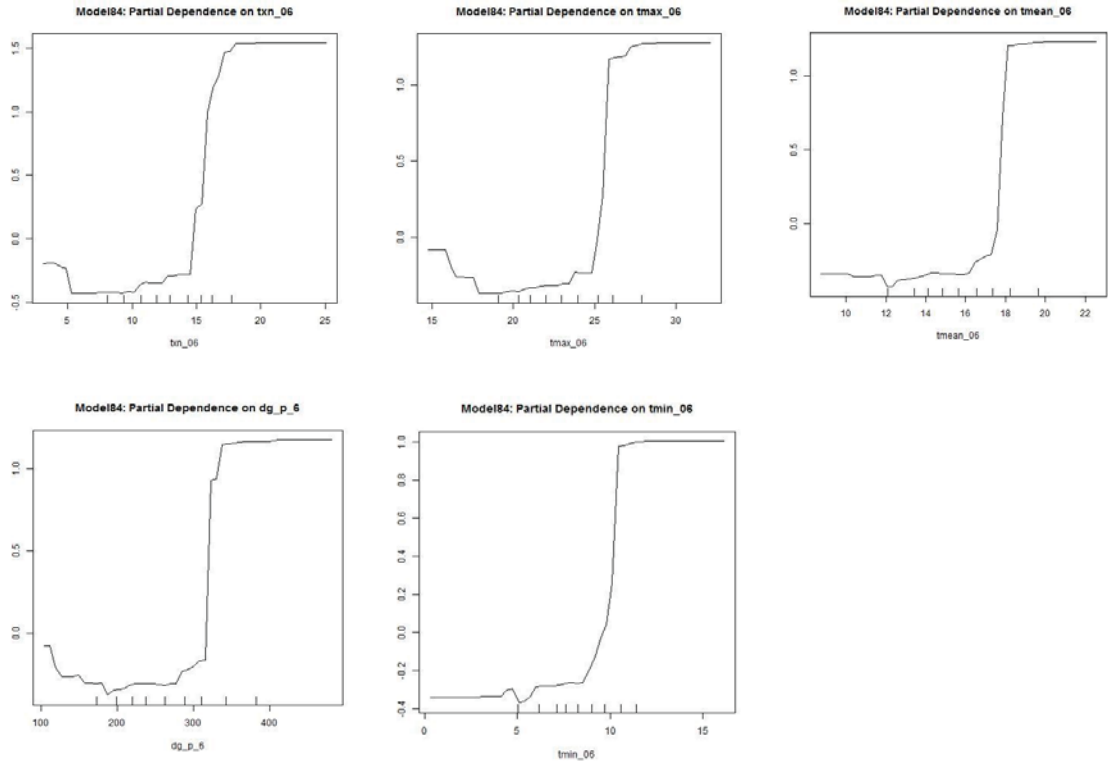
Figure 2.12. Partial dependence plots of the June degree days and temperature indices

Aside from the psyllid yellows observations in Canada, the psyllid yellows records in the states have higher June degree-days accumulation than the no psyllid yellows incidents. Higher June degree-days accumulation means more growth and development accomplished by the potato psyllid population in June. The mean of June degree-days accumulation is 314.36 degree-days, which is close to the heat accumulation requirement of one life cycle, 358 DD. Whereas the mean June degree-days accumulation of case py = 0 is 236.42 DD, which is a quarter of degree-days required less than that of the psyllid yellows positive cases. In the partial dependence plot of DG_P_6 (June degree-days accumulation when feeding on potato), the probability of psyllid yellows incidenceis positively related to the June degree-days accumulation. The logit of the likelihood is close to zero (where the likelihood of psyllid yellows is 0.5), where the degree-days

57

accumulation is around 310 DD. The probability rises with the increase of degree-days

accumulation until the value is close to 400 DD.
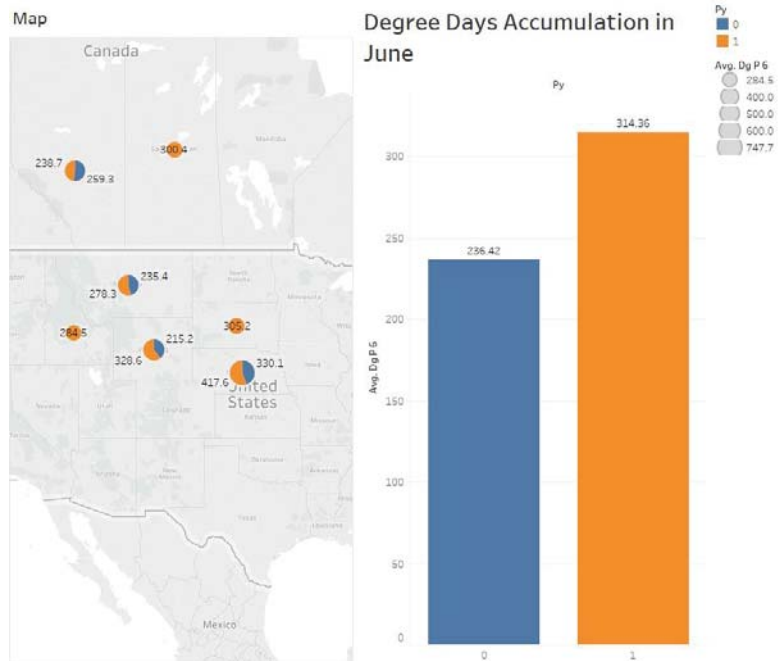
Figure 2.13. Geographical differences in the mean degree-days accummulation in
June between psyllid yellows positive and negative cases

The total precipitation in June is slightly lower when py = 1, with a mean of 77.23 mm, than when py = 0, with a mean of 91.71 mm. The partial dependence plot shows a similar pattern. There are more yes votes in the forest when June precipitation is smaller than 100 mm, and more no votes when June precipitation is larger than 100 mm.
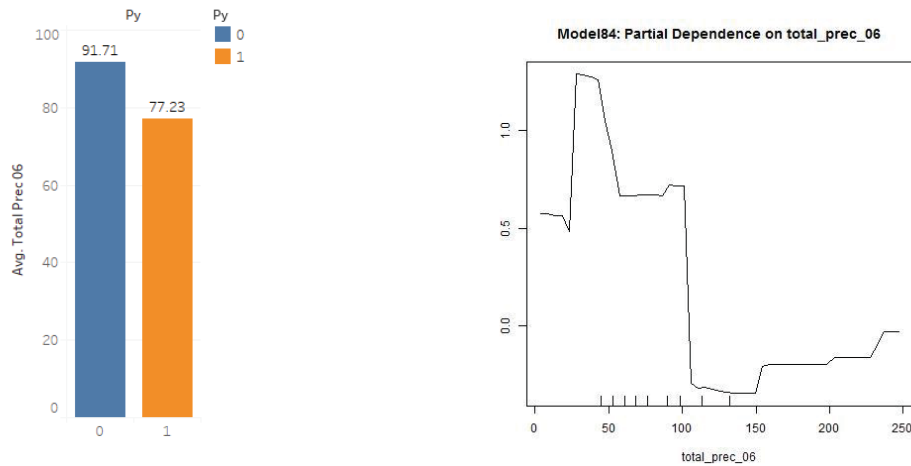
Figure 2.14. Total precipitation (mm) in June when psyllid yellows = 1, and psyllid yellows = 0

The indices related to September vary between the different psyllid yellows cases. The monthly mean, minimum of daily minimum temperature in September, TMIN_09, TNN_09, and monthly minimum of daily maximum temperature in September, TXN_09, all have higher values when py is equal to one. TNN_09, the monthly minimum of daily minimum temperature in September, has higher values (mean equals to 0.291 °C) when py is equaled to one, than when py is equal to zero (mean equals to -3.003 °C). The mean of TXN_09 of the psyllid yellows positive cases (13. 969 °C) is about 4 °C higher than that of the psyllid yellows negative cases (10.048 °C). TMIN_09, monthly mean of daily minimum temperature in September, has a higher mean value of 7. 346 °C when psyllid yellows equals to one than the mean of 4.589 °C when psyllid yellows is equal to zero. The monthly mean temperature in September (TMEAN_09), the monthly mean of daily

59

maximum temperature in September (TMAX_09), and degree-days accumulation in September (DG_P_9) show similar patterns. In the partial plots of the temperature-related indices in September, the logit of yes vote fraction increases with the indices. Altogether, those six temperature-related indices in September show that the positive psyllid yellows records have warmer September than the negative psyllid yellows records. Precipitation-related indices in September were not selected as model input variables determined by the variable reduction procedure.
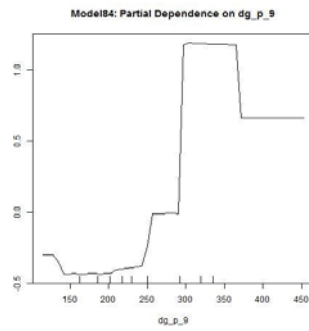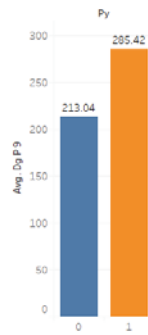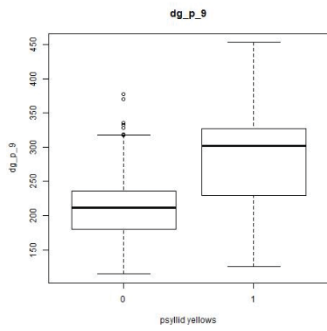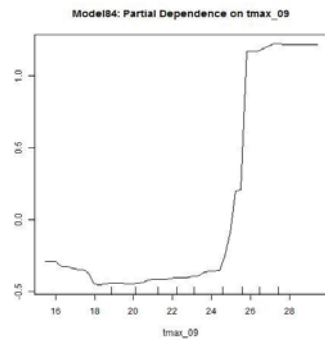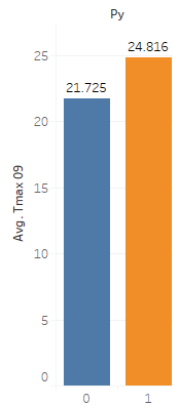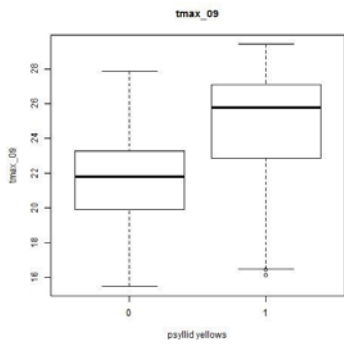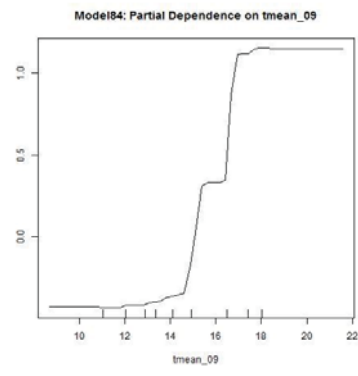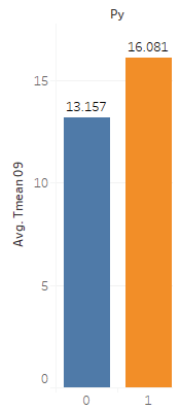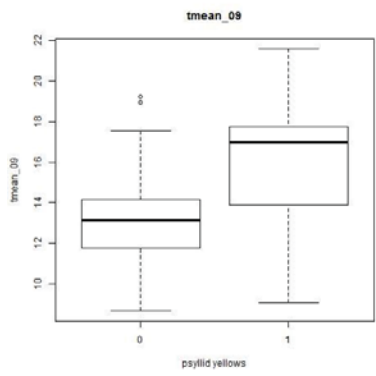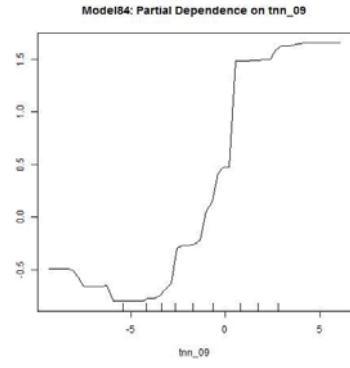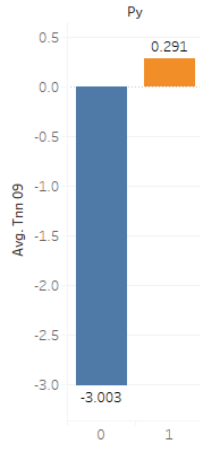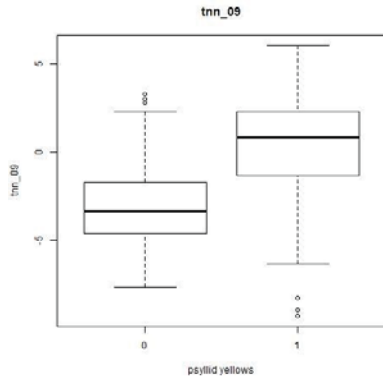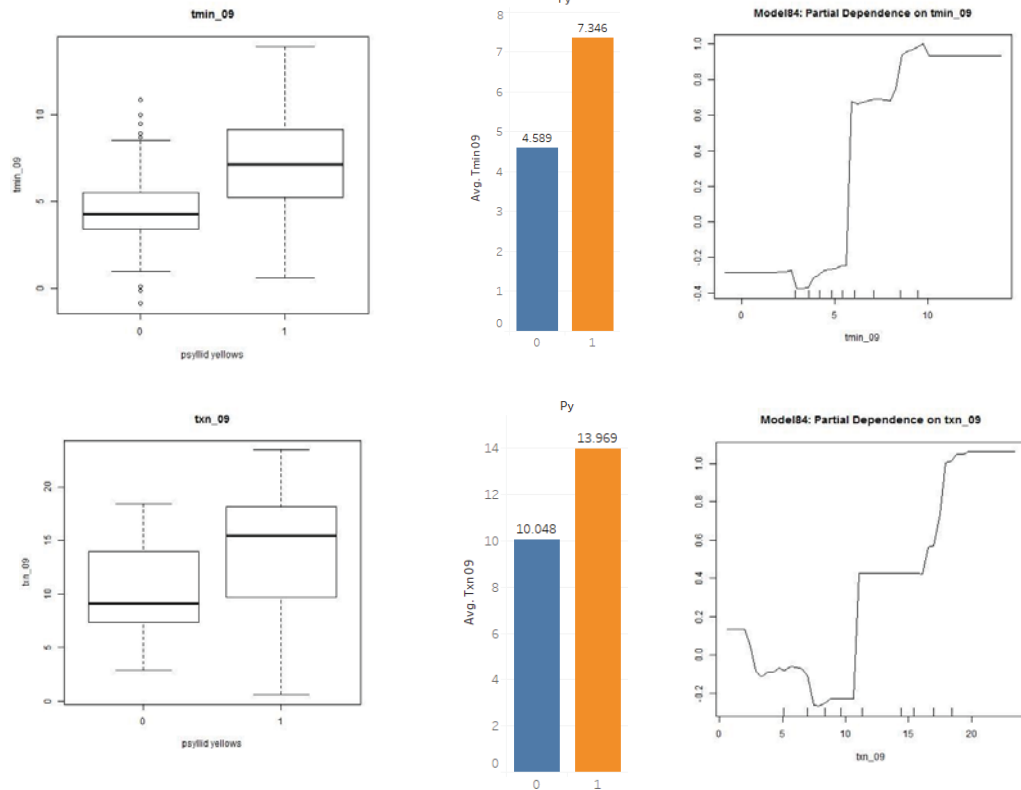
Figure 2.15. Comparing the September temperature and precipitation indices between psyllid yellows = 1, and psyllid yellows = 0. (TNN_09, TMEAN_09, TMAX_09, DG_P_9, TMIN_09, TXN_09)

The number of frost days (FD, the annual count of days when their minimum temperature is smaller than 0 °C) is one of the important indicator suggested by the variable reduction process. The number of annual frost days of psyllid yellows presence cases (174.26 days) is lower than the annual frost days of the psyllid yellows absence cases (193.16 days). The probability of psyllid yellows positive cases suggested by the partial dependence plot decreases when the number of frost days increases. When the number of frost days is around 150, the probability of psyllid yellows remains positive and the same. When the number of frost days increases to around 175 days, the likelihood of psyllid yellows positive cases starts to drop, which becomes lower than 0.5 at the end. In the map of psyllid yellows cases related to the number of frost days, psyllid yellows incidents have a

lower annual count of frost days than the no psyllid yellows incidents among the records

of psyllid yellows from Alberta, Montana, Wyoming, and Wyoming.
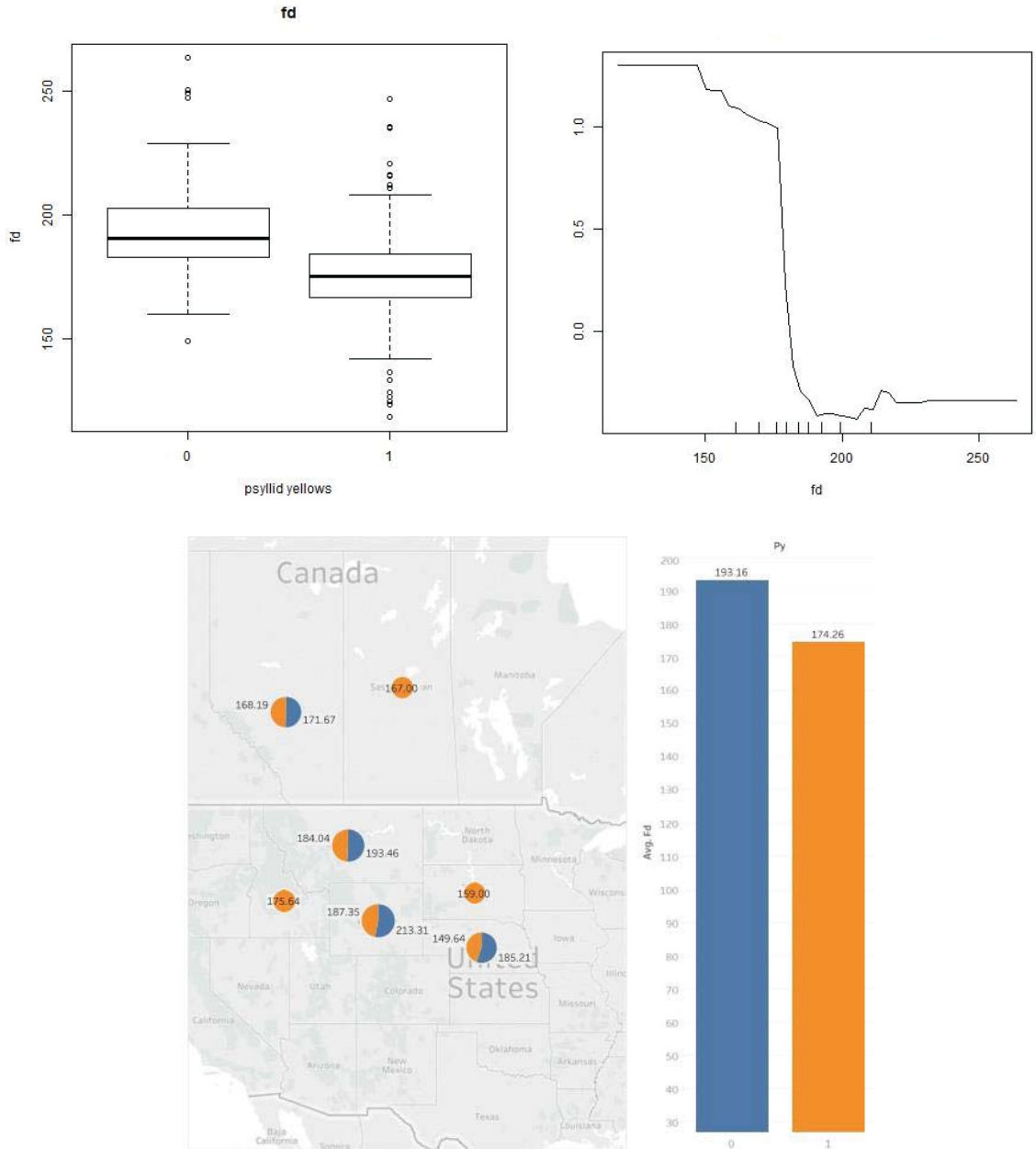
Figure 2.16. Comparing the FD, annual count of frost days between the psyllid yellow positive and psyllid yellows negative observations

The number of summer days, SU, the annual count of days when the daily maximum

temperature is larger than 25 °C, is another indicator determined as useful by the variable

reduction procedure. The number of summer nights is higher when the psyllid yellows

observation is positive (mean: 93.20 days) than that when the psyllid yellows observation

is negative (mean: 76.18 days).
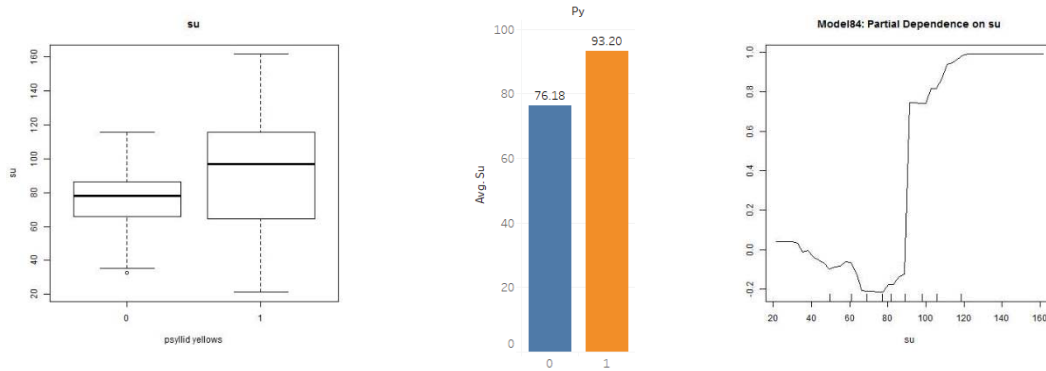
Figure 2.17. Comparing the number of summer days between the psyllid yellows
positive and psyllid yellows negative cases

Chapter 3: Modelling of potato psyllid distribution using Maxent

Chapter overview:

In this chapter, literature on the relationship between potato psyllids and environmental conditions is summarized. The Maxent model is described, and the methods of setting up the Maxent model for predicting potato psyllid occurrence probability are introduced and explained. The fitted Maxent model of potato psyllid occurrences was projected for different climate surfaces.

Background:

Tomato-potato psyllid, *Bactericera cockerelli* (Sulc) (Hemiptera: Triozidae) can transmit *Candidatus* Liberibacter solanacearum (Lso), a plant pathogen known to cause zebra chip disease (ZC) in potato, a condition in which starch and sugar are degraded in potato tubers of infected plants, and the inside of the potato is striped with brown (Munyaneza 2015). The main way to control the disease is to control the insect vector through spraying pesticides if required, and by adjusting the potato harvest time. This insect, also called the potato psyllid, appeared in field monitoring samples in Alberta, Canada, in 2015, but not in the field monitoring samples or collections in previous years. A program for field monitoring in Canada, managed at the University of Lethbridge, has been in place for four years to continue the search for the potato psyllid, which may be increasing in numbers, range expansion, or moving. Currently, no control measures are recommended in Canada, and the program is based on monitoring potato psyllids and their potential natural enemies, and DNA testing (in the Kawchuk lab at the Lethbridge Research Centre) of captured potato psyllids for the presence of Lso.

In addition to serving as a vector for the bacterial plant pathogen that causes zebra chip, potato psyllids have caused direct damage in the past. As described in the previous chapter, development of psyllid yellows symptoms is related to the feeding of potato psyllid nymphs (Richards & Blood 1933; Munyaneza 2012), leading to the development of psyllid yellows in the potato. The symptoms include leaf yellowing, curling, stunted plants and smaller potato tubers, and more severely plant death. Multiple psyllid yellows epidemics were observed and recorded during the 1930's and 1940's in the United States, Alberta, and Saskatchewan. The outbreaks of psyllid yellows were so severe that it caused significant losses in potato crop yields in multiple states during this period (summarized in Chapter 2 of my thesis). These historical data may have some value in understanding range and distribution, because they show where potato psyllid populations in the past were established, reproducing, and producing immature stages that fed on potato plant leaves. To test the hypothesis that standard species distribution modelling approaches could be used to predict the current and future distributions of this insect vector, a model based on the standard 19 BIOCLIM variables of SDM, and the historical distribution records of potato psyllid in the United States and Canada, was developed.

Other factors influence the impact of weather on potato psyllids, including genetic characteristics of the psyllid population at a site. The potato psyllid movements and populations in North America vary in source and may also vary in plant feeding attributes and responses to weather. There are four haplotypes of potato psyllids in the US, called Central, Western, Northwestern, and Southwestern. After the potato psyllid first became abundant in Texas in 2000, it expanded northward and produced zebra chip

outbreaks in Idaho, Oregon, and Washington in 2011 (Crosslin et al. 2012a; Crosslin et al. 2012b; Munyaneza 2012). At present, the Central haplotype is resident primarily in Texas, and the Northwestern haplotype appears year-round in the Pacific Northwest (PNW). The Western, Central, Northwestern haplotypes show up in the PNW area during the growing season, but in the winter the Northwestern haplotype is the only haplotype that is dominant (Swisher et al. 2013; Swisher et al. 2014). The Northwestern haplotype has been identified overwintering and living on its perennial host, bittersweet nightshade. Bittersweet nightshade significantly improves both the Northwestern's and the Central's tolerance to cold temperature as low as -12 ℃ (10 °F), but, does not improve cold hardiness of the Western haplotype (Horton et al. 2014). The distributions of the host plants are not under the scope of this study. Potato psyllids were found in the potato fields in southern and central Alberta during the 1930's and 1940's (Sanford 1952).

The potato psyllid is now well established in Canada, and increasing in density, although still rare. In July 2015, ten potato psyllids were found in Alberta, Canada, after hundreds of sticky cards were placed in fields at recorded locations and regularly examined (information from the Canadian potato Psyllid and Zebra Chip Monitoring Network), followed by over 200 collected in Alberta in 2016 (Johnson & participants 2016), and the first recorded collections of adult potato psyllids in Saskatchewan and Manitoba (Vikram et al. 2016). In 2017, the first psyllids of the year were found on sampling cards placed in fields June 13-23 (University of Lethbridge Potato Psyllid Monitoring lab data, July report to monitoring network, a national monitoring network currently sampling from New Brunswick to British Columbia). Being part of the monitoring program, I am working on a subproject studying the relationship between the potato psyllid and

weather, climate using species distribution modelling methods. So far, there have been no

previous studies of modelling the distribution of potato psyllids in Canada with Maxent,

related to weather and climate, using historical museum and literature data, and standard

BIOCLIM indices.

Biogeographical analysis: species distribution models

Species distribution models describe the species niche, and predict the suitable habitat for

the species, by comparing the relationship between a known species distribution (based

on records of presence) and environmental conditions at those locations (Franklin &

Miller 2009). An environmental niche of a species can be quantified as a composition of

multiple independent (but not necessarily uncorrelated) environmental variables, each

with a range of suitable conditions for the species (Vandermeer 1972). Average, seasonal

and extreme conditions influence species distribution due to adaptation, tolerance and

intolerance of the species to environmental fluctuations and extremes, often through

short-term impacts such as mortality, or medium-term factors such as failure of

development, impeded reproduction, or food limitations. Direct conditions, such as

nutrients, pH, temperature, or solar radiation, impact species distribution by influencing

the species' physiology (Franklin and Miller 2009). Temperature generally determines

rates of insect growth (Wellington et al. 1999). Insects are ectothermic animals, so their

body temperatures are based mainly on ambient temperature and solar radiation, and

influenced by wind speed and conduction from their surroundings. Other weather

conditions can influence insect distribution by influencing biology, ecology, behavior,

movement, diapause, fecundity, food plants, and seasonal polymorphism. Potato psyllid

growth has been modeled, based on rates of passage between developmental stages as a

function of temperature under lab conditions. Tran et al. (2012) determined that the full life cycle requires approximately 358 degree-days above a threshold of 7.1 °C, and that the nonlinear rates of development between stages and over the entire life cycle can be predicted from standard models. One of the models Tran et al. (2012) chose as giving a good fit to the development data was developed during an NSERC Strategic project at the University of Lethbridge, (Lactin et al. 1995). However, the dynamics of field populations are often different from the dynamics seen in under lab environments (Yang et al. 2010; Yang et al. 2013). Some of the highly variable and fluctuating effects of weather and climate may be integrated into simpler bioclimate indices or variables that include a range of effects that would be apparent in outcomes such as geographic occurrence and phenology. BIOCLIM variables have been used in different ecological studies defining the species distribution related to climate since 1986 (Nix 1986). Little study has been done on the relationships between bioclimate conditions and potato psyllid distribution, or applications of species distribution modelling methods for this purpose.

Presence-only data

The potato psyllid occurrence data are mostly presence observations collected from agricultural field surveys, plant disease periodicals and literature articles, and museum specimen online databases. Ideally, by comparing the climate conditions where the species was found and the climate conditions where the species was found to be absent, the climatic influence on species distribution can be analyzed. However, with only species presence data or highly disproportionate data with mostly presences, the lack of variation of the species occurrence data makes using regression method or classification

method to compare species occurrence change related to climate impossible. Presence-only data is commonly used in plant and animal studies. Different methods developed for using the species data (Elith et al. 2006).

Climate envelope model

Climate envelope model compares the area of interest with the climate profile, the environmental space of the species based on presence-only data, and the predicted distribution (Pearce & Boyce 2006). With a climate envelope model, such as "hyper-box classifier" (Franklin and Miller 2009), BIOCLIM, DOMAIN, prediction can be made based on only presence data. The objective of using a climate envelope model is to compare the climate of the areas of interest, to the climate profile of the species, and then evaluate its potential for occupation within the areas. The climate profile is a summary of the statistics of the environments related to the known presence of the organism of interest. It also represents a multidimensional environmental space describing suitable environmental conditions (ecological niches). For example, the BIOCLIM model generates its climate profile by calculating the minimum, five percentile, 95th percentile, and maximum, for the degree of presence density and environmental variables, and compares the climate values to the profile to see which percentile the values fall into (Busby 1991).

Presence-background model

A presence-background model is another way to use the species presence data. It uses random sampled background points as a contrast to the environmental covariates of the species known distribution (Elith et al. 2006). Contrasting the species covariates to the

background environment is a common method used in species distribution modelling (Elith et al. 2006). The Maxent model can be used as a presence-only model, by modelling species occurrences contrasting to the associated background environment.

Maxent was used to predict the potential distribution of a different potato psyllid species, *Russelliana solanicola* (Hemiptera: Psylloidae) in southern America, using the BIOCLIM variables including mean annual precipitation, mean annual temperature, precipitation seasonality, and temperature seasonality from the WorldClim database (http://www.worldclim.org), and other vegetation and geological indices (Syfert et al. 2017).

The Maxent model

Maximum entropy model can be used for mapping species distribution using presence-only data (Phillips et al. 2006; Elith et al. 2011).

The model structure of Maxent (Elith et al. 2011)

(x, y): coordinates, x is longitude, and y is latitude

L : landscape of interest

L_1 : subset of the L, landscape where the observations of species presence is positive (y =1)

E: a vector of environmental gradients

$P_1(E)$: probability density of E, when y = 1

$P_L(E)$: probability density of E in the landscape L

$P_0(E)$: probability density of E, when y = 0 (background)

R: prevalence, the proportion of the occupied sites by the species in the landscape L

$Pr(y = 1 \mid E)$: probability of absence, conditioned on E

71

$h(E)$: feature of Maxent used to fit the model, feature is a function of E

Maxent estimates the probability of species presence conditioned on specific

environment E, by estimating the ratio between the probability density of E (y = 1) and

the probability of E across the landscape, L. R is the prevalence, which is a constant.

$$\Pr(y = 1 \mid E) = \frac{P_1(E)R}{P_L(E)}$$

Multiple probability density functions can be fitted to the probability of E when y = 1.

Maxent algorithm selects the density function of $P_1(E)$ t close to $P_L(E)$, by minimizing

the relative entropy of $P_1(E)$ compared to the background $P_L(E)$. Minimizing the

relative entropy between the $P_1(E)$ and $P_L(E)$ is equivalent to maximize the entropy of

the probability of presence over a locality explained by Elithe et al. (2011). Gibbs

distribution, expressed by the following exponential model, is used to minimize the

relative entropy (Elith et al. 2011). $f(E)$ is estimated byfeature $h(E)$. Features are a set

of functions containing multiple covariates that Maxent uses to fit the data (Elith et al.

2011).

$$P_1(E) = P_L(E) \, e^{f(E)}$$

$$f(E) = \alpha + \beta \times h(E)$$

The probability of randomly selected background points $P_0(E)$, is used to construct the

$P_L(E)$ probability density of E of the landscape. It is determined by the background

points being sampled. Sampling background data is not intended to approximate the

probable absences of the focal species (Phillips et al. 2009). Maxent estimates the

probability of species presence under certain environmental conditions (E), by comparing

the probability density of the environmental covariates (E) where that species is present,

to the probability density of the environmental covariates (E) across the background

locations. The availability of certain environment conditions (E), across the landscape

determines species accessibility to those environmental conditions. By comparing to the

probability of covariates E in the background conditions, and knowing the probability of

E among the locations where a species is present, the probability of the species presence

under environment E can be estimated. Maxent minimizes the divergence (relative

entropy, dispersedness) between the probability density of E in the background, and the

probability density of E among the sites of occurrence (Elith et al. 2011).

Selecting background data to represent the background environment

I used equally weighted stratified random sampling to select the data points from

Livneh's dataset. Background data is used to represent the environments in the studied

area and studied period (Phillips et al. 2009). Random sampling background points

without excluding the known presences serves the purpose of characterizing the

background environment better, as species presence locations are part of the studied

region (Phillips et al. 2009). Environmentally dissimilar background locations produce

more accurate model predictions (Franklin and Miles 2009). Increasing the spatial extent

of the studied area might increase the possibility of selecting dissimilar background

points, and giving a greater range in which to distinguish the conditions associated with

presence. However, background data that are environmentally dissimilar from presences

might predict a larger suitable range for the studied species (Chefaoui & Lobo 2008). I

collected potato psyllid activity presence data from multiples years and multiple locations

across the United States and Canada. Similar to the method used for stratified random

sampling in a spatiotemporal space (Boschetti et al. 2016), in order to characterize the

climate of the background of multiple years, I divided the spatial temporal space into

10,000 * 24 equally sized strata, each covering the spatial temporal range of 100 km*100

km, 4 year period. For each stratum, I randomly sampled one background point with one

coordinate (longitude and latitude) and one year. Similar to random sampling in a two-

dimensional space used in most of the published species distribution modelling studies, I

randomly sampled background data from the three-dimensional space by adding time as

the third dimension. Background points randomly sampled fromthis three-dimensional

space were recorded by latitude, longitude and year. A total of 179,952 background

points were selected for analysis, approximately equal in numbers sampled, by year

interval. The distribution of the background points conditioned on over years is shown in

the graph below.
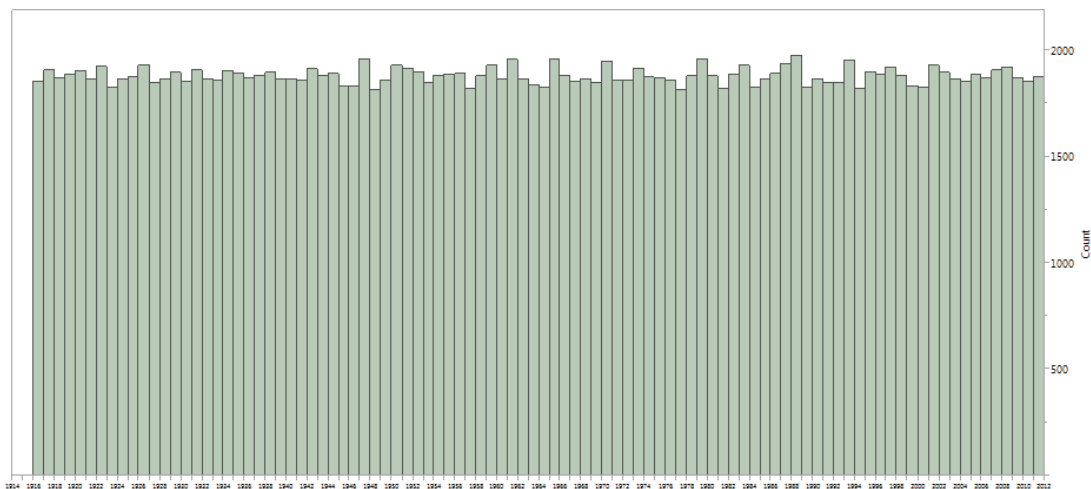
Figure 3.1. The distribution of background data count was approximately equal by year
interval, for 1916 to 2012

Potato psyllid observation data

Museum records

Species observations for the Maxent model were collected from the Global

Biodiversity Information Facility (GBIF; http://data.gbif.org/occurrences/), and the

potato psyllid records collected during the psyllid yellows records collecting process. The

GBIF data downloaded for potato psyllid collection of museum records. The main

provider of this dataset is the Essig Museum of Entomology, located at the University of

California, Berkeley, California, USA (https://essigdb.berkeley.edu/). A total of 393

museum records with unique location, date, and valid weather data for climatic indices

calculation were used as part of the input dataset.

Historical literature records

Historical literature records of potato psyllid observations were collected from plant

disease survey periodicals, the Plant Disease Reporter, the Plant Disease Bulletin, the

Plant Disease Reporter Supplement, the Canadian Plant Disease Survey, and entomology

and agriculture literature. A total of 271 potato psyllid presence observations with valid

date, location, clear statement of the presence of the species, and valid weather data for

climatic indices calculation were collected and selected as part of the input dataset.

Similarly, a total of 58 potato psyllid absence observation records were selected, and

included in the background dataset.

Use of the standard BIOCLIM indices

Originally developed as part of the bioclimate modelling system, BIOCLIM variables

measure the different aspects of the climate (Booth 1985; Nix 1986), and used in

modelling the distribution of different species (Franklin and Miller 2009). The standard

method of generating the BIOCLIM variables of the WorldClim dataset (Hijmans et al.

2005) was used for calculating the BIOCLIM variables for the species records and

background points. The method is available in the R package, 'dismo'(Hijmans et al.

2017) written by the same author of the WorldClim dataset. The WorldClim dataset is a

set of spatially gridded BIOCLIM indices calculated based on interpolated weather

surfaces (Hijmans et al. 2005). It is the main dataset used in the current species

distribution models and ecological niche models (Kriticos et al. 2014). The following is

the list of 19 BIOCLIM variables that I calculated and used in this test of the method's

potential value in understanding the distribution of the potato psyllid. These are

scenopoetic variables, which can not be consumed or influenced by the species (Peterson

et al. 2011).

Table 3.1. The 19 BIOCLIM variables of the WorldClim dataset (Hijmans 2005)

Index	Description	Unit
BIO1	Annual Mean Temperature	°C
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	°C
BIO3	Isothermality (BIO2/BIO7) (* 100)	X
BIO4	Temperature Seasonality (standard deviation *100) *standard deviation of the monthly mean temperature	X
BIO5	Max Temperature of Warmest Month	°C
BIO6	Min Temperature of Coldest Month	°C
BIO7	Temperature Annual Range (BIO5-BIO6)	°C
BIO8	Mean Temperature of Wettest Quarter	°C
BIO9	Mean Temperature of Driest Quarter	°C
BIO10	Mean Temperature of Warmest Quarter	°C
BIO11	Mean Temperature of Coldest Quarter	°C
BIO12	Annual Precipitation	mm
BIO13	Precipitation of Wettest Month	mm
BIO14	Precipitation of Driest Month	mm
BIO15	Precipitation Seasonality (Coefficient of Variation)	X
BIO16	Precipitation of Wettest Quarter	mm
BIO17	Precipitation of Driest Quarter	mm
BIO18	Precipitation of Warmest Quarter	mm
BIO19	Precipitation of Coldest Quarter	mm

Variable Selection

Highly correlated variables should not be used together to make predictions for the

Maxent model. Similar to the methods used for removing the highly collinear input

variables for the Maxent model (Bosso et al. 2017a; Bosso et al. 2017b), I calculated

Pearson's r correlation matrix for the input dataset, and removed the variables with an

absolute value of correlation strength bigger than 0.75. This concept is similar to

statistical model fitting such as multiple regression, in which highly correlated

independent variables (X) are sometimes reduced to a simpler set when used to predict a

dependent variable (Y). During the early stage of the model fitting, for the Maxent

model method, monthly total degree-days of the species and other background data (often

weather-based) were found to be highly collinear, that is, correlated with each other. As a

result, monthly degree-days are not included in this study, because they can in part be

derived from weather data used as the basis for these indices. BIO2, BIO3, BIO4, BIO7,

BIO15 were removed due to their correlation to the temperature or precipitation indices.

BIO1, BIO5, BIO6, BIO8, BIO9, BIO10, BIO11 are temperature indices. BIO12, BIO13,

BIO14, BIO16, BIO17, BIO18, and BIO19 are precipitation indices. Temperature indices

and precipitation indices have a correlation lower than 0.75 among each other. I selected

the temperature indices by excluding the ones with correlation strength $r > 0.75$ with

BIO1. Within the precipitation correlation matrix, I selected the indices BIO14, and

BIO18 due to their $r < 0.75$ correlation strength. The final set of the model input

environmental variables was BIO1, BIO8, BIO9, BIO12, BIO14, and BIO18. BIO1 is the

annual mean temperature. BIO12 is annual total precipitation.

Table 3.2. Correlation matrix of the BIOCLIM temperature indices

	BIO1	BIO5	BIO6	BIO8	BIO9	BIO10	BIO11
BIO1	1	0.82	0.86	0.59	0.51	0.91	0.93
BIO5	0.82	1	0.53	0.56	0.33	0.93	0.63
BIO6	0.86	0.53	1	0.30	0.66	0.63	0.96
BIO8	0.59	0.56	0.30	1	-0.17	0.70	0.39
BIO9	0.51	0.33	0.66	-0.17	1	0.29	0.64
BIO10	0.91	0.93	0.63	0.70	0.29	1	0.71
BIO11	0.93	0.63	0.96	0.39	0.64	0.71	1

Table 3.3. Correlation matrix of the BIOCLIM precipitation indices

	BIO12	BIO13	BIO14	BIO16	BIO17	BIO18	BIO19
BIO12	1	0.90	0.50	0.95	0.65	0.33	0.84
BIO13	0.90	1	0.31	0.96	0.42	0.26	0.82
BIO14	0.50	0.31	1	0.35	0.79	0.31	0.32
BIO16	0.95	0.96	0.35	1	0.46	0.28	0.86
BIO17	0.65	0.42	0.79	0.46	1	0.38	0.41
BIO18	0.33	0.26	0.31	0.28	0.38	1	-0.10
BIO19	0.84	0.82	0.32	0.86	0.41	-0.10	1

Table 3.4. The correlation matrix of the selected BIOCLIM temperature and precipitation indices

	BIO1	BIO5	BIO6	BIO8	BIO9	BIO10	BIO11	BIO12	BIO13	BIO14	BIO16	BIO17	BIO18	BIO19
BIO1	1	0.82	0.86	0.59	0.51	0.91	0.93	-0.10	-0.03	-0.19	-0.07	-0.13	0.03	-0.12
BIO5	0.82	1	0.53	0.56	0.33	0.93	0.63	-0.40	-0.32	-0.31	-0.37	-0.31	-0.06	-0.41
BIO6	0.86	0.53	1	0.30	0.66	0.63	0.96	0.13	0.17	-0.09	0.15	-0.02	-0.11	0.19
BIO8	0.59	0.56	0.30	1	-0.17	0.70	0.39	-0.20	-0.15	-0.18	-0.17	-0.17	0.45	-0.43
BIO9	0.51	0.33	0.66	-0.17	1	0.29	0.64	0.06	0.09	-0.07	0.07	0.00	-0.44	0.29
BIO10	0.91	0.93	0.63	0.70	0.29	1	0.71	-0.25	-0.19	-0.23	-0.23	-0.20	0.12	-0.35
BIO11	0.93	0.63	0.96	0.39	0.64	0.71	1	0.03	0.09	-0.15	0.06	-0.08	-0.10	0.08
BIO12	-0.10	-0.40	0.13	-0.20	0.06	-0.25	0.03	1	0.90	0.50	0.95	0.65	0.33	0.84
BIO13	-0.03	-0.32	0.17	-0.15	0.09	-0.19	0.09	0.90	1	0.31	0.96	0.42	0.26	0.82
BIO14	-0.19	-0.31	-0.09	-0.18	-0.07	-0.23	-0.15	0.50	0.31	1	0.35	0.79	0.31	0.32
BIO16	-0.07	-0.37	0.15	-0.17	0.07	-0.23	0.06	0.95	0.96	0.35	1	0.46	0.28	0.86
BIO17	-0.13	-0.31	-0.02	-0.17	0.00	-0.20	-0.08	0.65	0.42	0.79	0.46	1	0.38	0.41
BIO18	0.03	-0.06	-0.11	0.45	-0.44	0.12	-0.10	0.33	0.26	0.31	0.28	0.38	1	-0.10
BIO19	-0.12	-0.41	0.19	-0.43	0.29	-0.35	0.08	0.84	0.82	0.32	0.86	0.41	-0.10	1

BIOCLIM indices under climate change scenarios

The WorldClim database version 1.4 provides four sets of simulated BIOCLIM indices under four different Representative Concentration Pathways (RCPs), simulations of greenhouse gas concentrations, of 2050 (average for 2041-2060) and 2070 (average for 2061-2080). The database provides the IPPC5 GCM climate projections for the four RCPs. The GCM output is downscaled and provided to the user at various spatial resolutions (http://worldclim.org/). Different RCP has different radiative forcing levels. Radiative forcing is the difference between the incoming solar radiation to earth deducted from the amount of radiation reflected back to space. The concentration of greenhouse gases increases the level of radiative forcing. Different RCP pathway has a different level of radiative forcing. The RCP 2.6 pathway has the lowest increase of radiative forcing level, whereas RCp 8.5 pathway has the highest radiative forcing increase (Nazarenko et al. 2015). Temperature increases with different rates under different RCP pathways (Nazarenko et al. 2015). The order of RCP with a temperature increase during 2005 to 2100 from low to high is, RCP2.6, RCP4.5, RCP6.0, and RCP8.5 (Nazarenko et al. 2015).

Under the scenario RCP 2.6, global mean temperature will rise 0.4 to 1.6 °C from 2046 to 2065, and 0.3 to 1.7 °C from 2081 to 2100. Under the RCP 4.5 scenario, global warming can increase 0.9 to 2.0 °C during 2046 to 2065, and 1.1 to 2.6 °C during 2018 to 2100. For scenario 6.0, global temperature can add 0.8 to 1.8 °C from 2046 to 2065, and 1.4 to 3.1 °C during 2081 to 2100 (Stocker 2014). BIOCLIM layers under scenarios RCP2.6 and RCP8.5 were downloaded from the WorldClim database, and used for Maxent model

projection of the climate suitability for potato psyllid under climate change (Hijmans et al. 2005).

Table 3.5. Global temperature increased (°C) under different greenhouse gas concentration trajectories (Stocker 2014)

Scenario	2050 (°C)	2070 (°C)
RCP2.6	1.0	1.0
RCP4.5	1.4	1.8
RCP6.0	1.3	2.2
RCP8.5	2.0	3.7

Model fitting

The Maxent model fitting procedure was conducted using the R dismo package. The input variables are the selected BIOCLIM variables after removing the variables with an absolute correlation strength greater than 0.75. The randomly selected data were used as background data, and published potato psyllid records were used as presence data. No duplicate data records were included in the input datasets. For purposes of model fitting and probability calculation, the background data and species data were split into a 70% training dataset and 30% test dataset. The model was trained for 1000 iterations by the Maxent algorithm, drawing 70% each time and performing the probability computations. A selection of 10,000 background points for each iteration was the default setting of the Maxent model. The suggested and widely acknowledged number of pseudo-absences points typically used for fitting GLM is 10,000 points (Morgane et al. 2012), so this standard was utilized. Maxent can be interpreted in a manner similar to a logistic model (GLM) (Renner & Warton 2013). During the early stage of model fitting, a total of 20,000 background points returned a more stable model performance. As a result, 20,000

background data points were chosen randomly for each iteration, as the representation of the background environment.

Variable importance

Percentage contribution shows the contribution of the variable to the increase of regularized gain of the model by changing the coefficient of the variable within features. Permutation contribution of the variable show the contribution of the variable to the decrease of the training AUC of the full model, when the values of that variables are randomly permuted among the training points (Phillips et al. 2006). When the permutation importance is higher, the more the training AUC of the model decreases when the variable value being randomly permuted.

Response curves

Maxent outputs two types of response curves, marginal response curve, and the response curve of the Maxent model fitted on an isolated variable. The response curve I used to visualize the relationship between the cloglog and the variable is the second type, because of the certain level of correlation strengths in the dataset even though using the $r > 0.75$ threshold. Maxent outputs marginal response curves to show the change of the probability of species presences by changing only one specific environmental variable, while keeping the other predictors at their mean values. It shows the marginal effect of the variable variation. The y-axis, the cloglog output shows the estimates of species probability between 0 and 1. The correlation structure among the predictors makes evaluating the marginal effect of the variable unrealistic as the values of the correlated variables cannot be held at their means in reality.

Model evaluation

The receiver-operating characteristic (ROC) plot is the output of the Maxent algorithm showing the fit of the model to the training and the test data (Franklin & Miller 2009). The x-axis in the plot shows the false-positive rate (1- Specificity), the rate of falsely predicting the presence of the species. The y-axis shows the true positive rate (Sensitivity), the rate of correctly predicting actual species presences (Franklin and Miller 2009). The area under the curve (AUC) can be used to compare the model prediction versus random prediction, and the performance among different models. When the AUC value is equal to 0.5, it means the prediction is no better than random (similar to a plot of the logit). If the AUC value is bigger than 0.5, the model predictive performance is better than random. If the AUC value is smaller than 0.5, the model predictive performance is worse than random prediction. There is not a baseline value of AUC to suggest how "accurate" the model should be (Charles et al. 2013).

Results

I selected different combinations of BIOCLIM variables to examine the influence of the long-term temperature and precipitation extremes to the model prediction by comparing the 1960-1990 normal projection and the projections under different climate change scenarios. The purpose of this model test is to see if the BIOCLIM variables could provide explanatory or predictive value for an insect species in this case.

In total, 465 presence points and 180,010 background points were used for model training. A total of 198 presence points were used as test data to evaluate the fitted Maxent model. The iteration number was 1000 times, with 20000 background points

randomly selected during each iteration. The AUC value of the training data is 0.787, and

the AUC value of the test data is 0.767. It shows that the model's ability to make a

correct prediction is much better than random prediction.
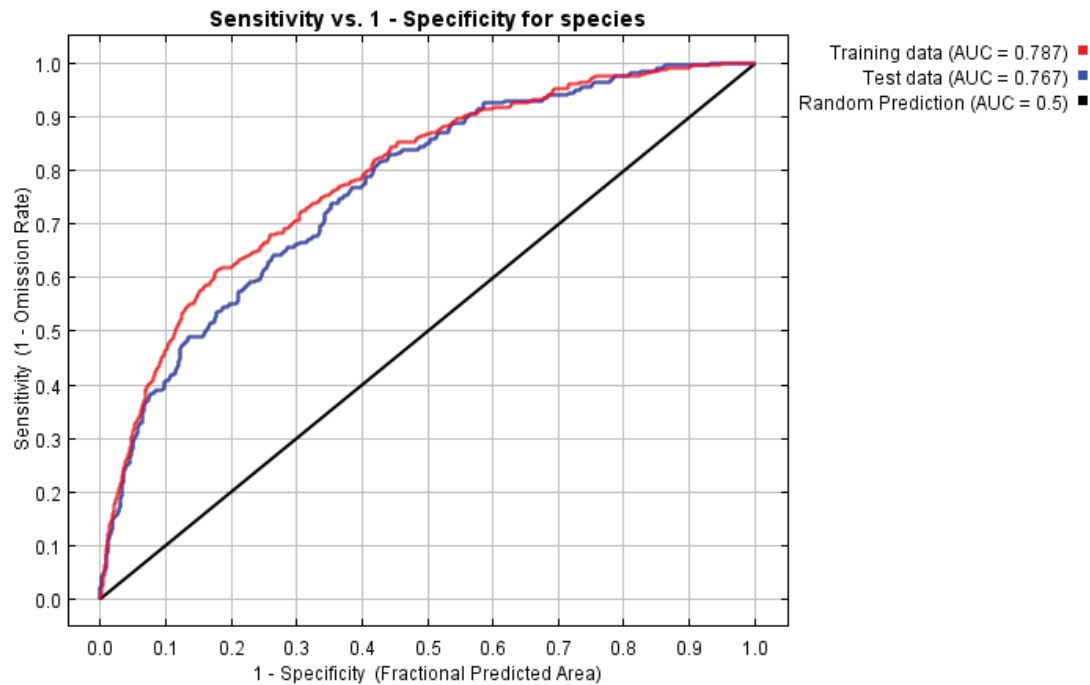
Figure 3.2. The receiver-operating characteristic (ROC) curve of the Maxent model

The model input environmental variables are BIO1 (annual mean temperature), BIO12

(annual precipitation), BIO14 (precipitation of driest month), BIO18 (precipitation of

warmest quarter), BIO8 (mean temperature of a wettest quarter), BIO9 (mean

temperature of a driest quarter). BIO12, annual precipitation is the most important

environmental variable of the model.

The boxplot of BIO1 (Fig. 3.4) shows that the annual temperature of the potato psyllid

occurrence data has higher 75th percentile, median, 25th percentile, and minimum than the

annual temperature of the background data. In the histogram of potato psyllid BIO1,

annual temperature of 7.5 °C has the highest total count among all the presence

observations. In the response curve of BIO1, the cloglog value, the estimate of the probability of presence from 0 to 1, increases when the annual temperature increases. The response curve is based on the Maxent model fitted solely on BIO1.

BIO8 (the mean temperature of the wettest quarter) has a lower 75th percentile, median, and higher 25th percentile values. There are more counts of presence observations when the BIO8 value is close to 10 °C (Fig. 3.4).

BIO9, the mean temperature of the driest month has more counts of observations when the BIO9 value is close to -5 °C, and 20 °C. In the boxplot for BIO9, it shows that the presences records have higher mean temperature of the driest month than the background environment (Fig. 3.4).

BIO12, is the annual temperature. The most important variable of the model. In the response curve of BIO12, the cloglog output decrease with BIO12. In the boxplot of BIO12, the BIO12 value of the potato psyllid data has a lower distribution than the background environment (Fig. 3.4).

BIO18, precipitation of the warmest quarter, shows a similar trend as the annual precipitation in its response curve. In the boxplot of BIO18, potato psyllid occurrence data has lower BIO18 values than the background data (Fig. 3.4).

Table 3.6. Percent contribution and permutation importance of each input variable of the Maxent model

Variable	Percent contribution (%)	Permutation importance (%)
BIO12	31.8	18.1
BIO14	25.8	1.8
BIO8	20.2	19.8
BIO18	16.1	28.3
BIO9	3.6	19.2
BIO1	2.5	12.8

Percentage contribution shows the contribution of the variable to the increase in regularized gain of the model by changing the coefficient of the variable within features. Permutation contribution of the variable shows the contribution of the variable to the decrease of the training AUC when the values of that variables are randomly permuted among the training points (Phillips 2006). When the permutation importance is higher, it means when the variable value being randomly permuted, the more the training AUC of the model decreases. BIO12, annual precipitation appears to be the most important variable by the variable importance measures. It has a percent contribution of 31.8 % and permutation importance of 18.1 % (Table 3.6).

Jackknife test (Fig. 3.3) shows the model performance gain when fitting the model with a specific variable, and without that variable. In the Jackknife graph below, it shows the model AUC gain on the test data. It shows that when the model is fitted only with BIO12 the model has the highest performance gain compared to other variables used in isolation.

It also shows that BIO12 is the variable decreases the model AUC value the most when it is omitted when fitting the model.
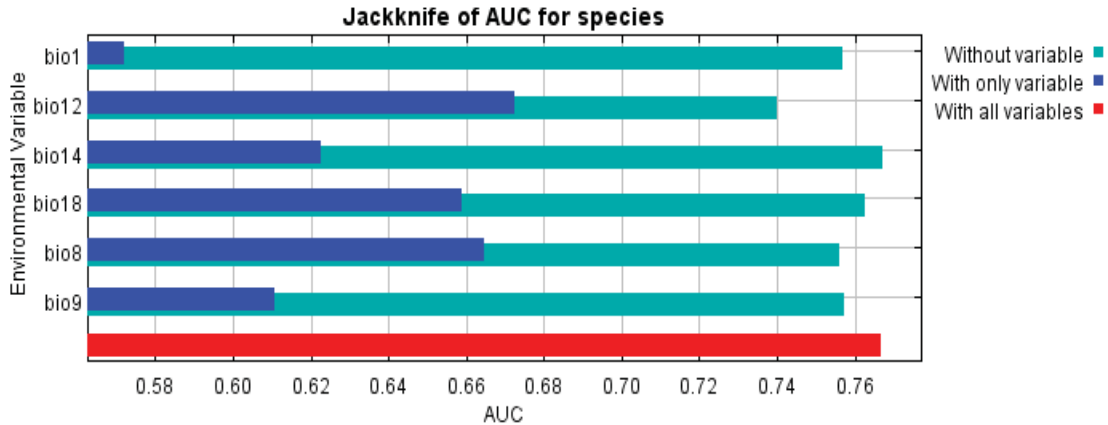

Figure 3.3. Jackknife test of AUC for species of the input BIOCLIM variables

Maxent projections of potato psyllid, *Bactericera cockerelli* (Sulc)

Six main BIOCLIM variables were used to determine the potato psyllid's potential distribution. The projection of potato psyllid presence probability under the 1960-1990 Climate Normals shows that California, central United States, Alberta, and British Columbia have higher presence probability than other areas. Under the climate change scenario projected through the RCP2.6 pathway, the probability of presence decreases in the central United States, however, increases in southwestern Canada. In the projected probability map of RCP8.5, the decrease in the presence probability in the central United States is more dominant. In the 1960-1990 climatic normal map, the distribution of high presence probability area is similar to the areas with actual recorded occurrence of potato psyllid. Although not used in fitting the model, Baja California, and Baja California Sur in Mexico, areas with potato psyllid museum specimen records, are given a high probability value by the model result.

California has 345 specimen records at different locations during different times. High probability estimates are assigned to the area within California. Yellowstone County in Montana had multiple outbreaks of psyllid yellows during the late 1930's to 1940's.

The predictive map under the 1960-1990 Climate Normals (Fig. 3.5), covers the area with potato psyllid occurrence records, regardless of the differences in time. The states and provinces with potato psyllid records before 1960 are Alberta, Saskatchewan, Montana, Idaho, Wyoming, Colorado, Utah, Arizona, Texas, and New Mexico.

The areas that have high probability values are: southern Alberta, central British Columbia, Southern Montana, eastern Idaho, northern Colorado, Utah, Nevada, central and southern California, Baja California, and Baja California Sur.

Under the RCP2.6 (2050) scenario (Fig. 3.6), the probability of presence in the southern, central United States decreases. Whereas the area of presence probability in southern Alberta and British Columbia expands.

Under the RCP2.6 (2070) scenario (Fig. 3.7), the suitable area in the central Alberta decreases. Only less than 10 presence records from Alberta in 1928 and the 1930's were used as part of the species data.

Under the RCP8.5 (2050) scenario (Fig. 3.8), the suitable area in central and northern US decreases, and the area close to Texas and Mexican boarders becomes more suitable. The Lower Rio Grande Valley in Texas shows a higher probability than the 1960-1990 normal projection. The LRGV area in Texas is the locations with recent potato psyllid population and zebra chip outbreaks. The first diagnosed ZC was found in Texas in 2000. Almost the whole northen BC, and Alberta are predicted with moderate and high

presence probability. (Note that this result concerns only suitable climate, and does not

includ predictions based on whether suitable food plants would be present.)

Under the scenario of RCP8.5 (2070) (Fig. 3.9), areas with high or moderate probability

in Alberta, BC, and the central and northern US decrease. However, the area of

probability 0.5 increases in northern Canada. It means in that area, the probability that

potato psyllid is present is random.

	Boxplot	Response curve
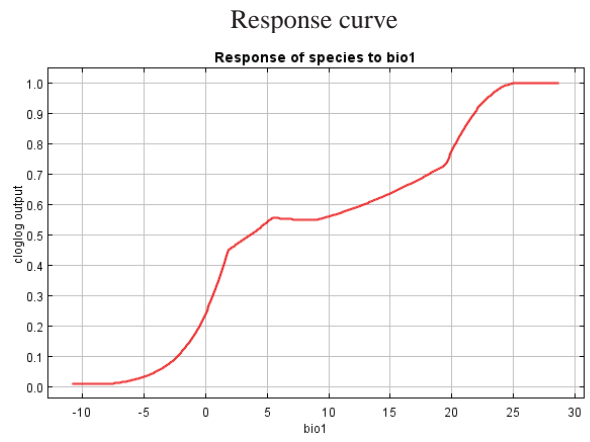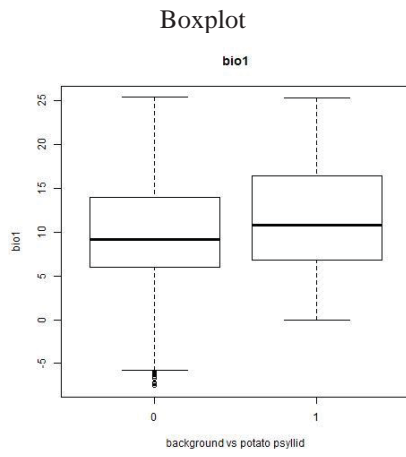
BIO1

BIO8

BIO9
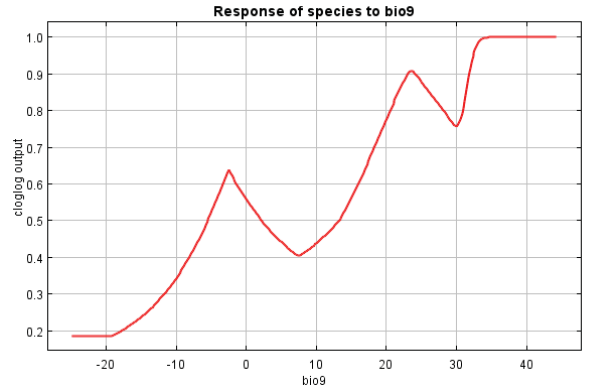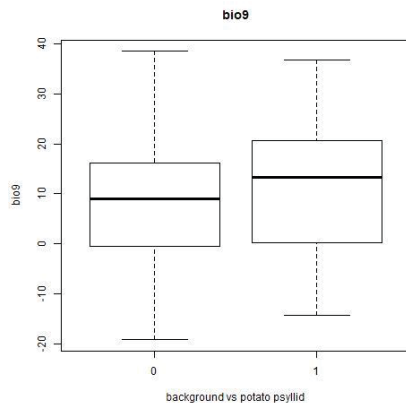
BIO12
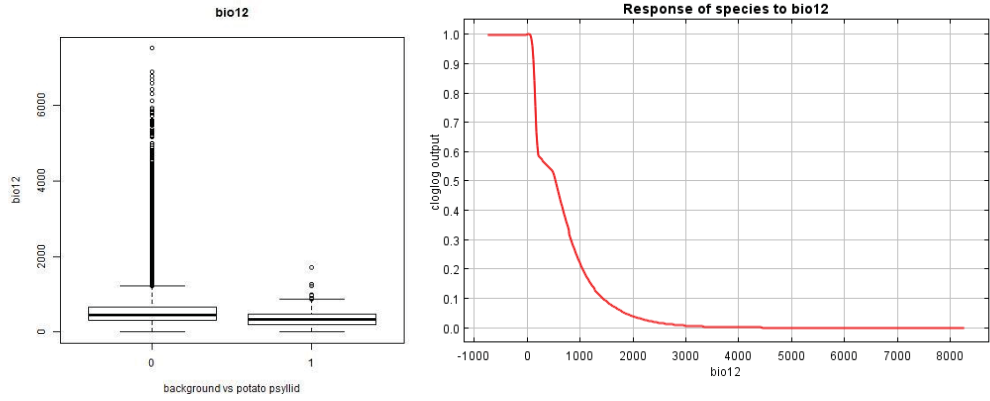
BIO14
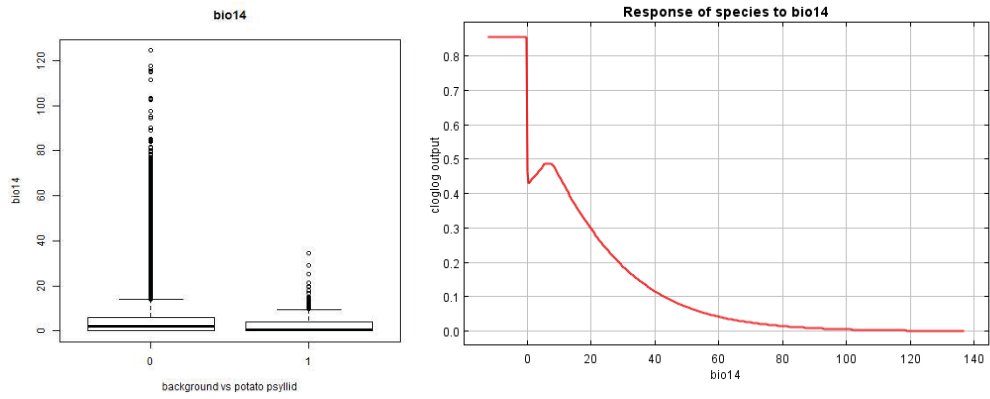
BIO18
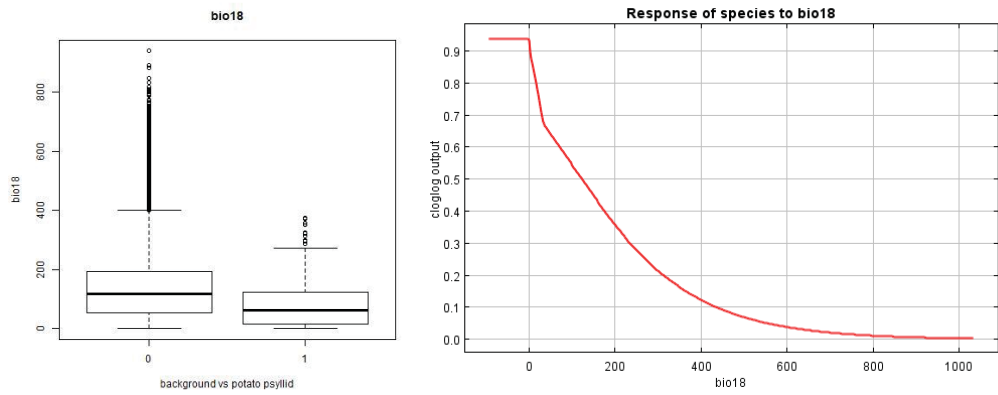
Figure 3.4. Boxplots and response curves for the input environmental variables of the fitted Maxent model
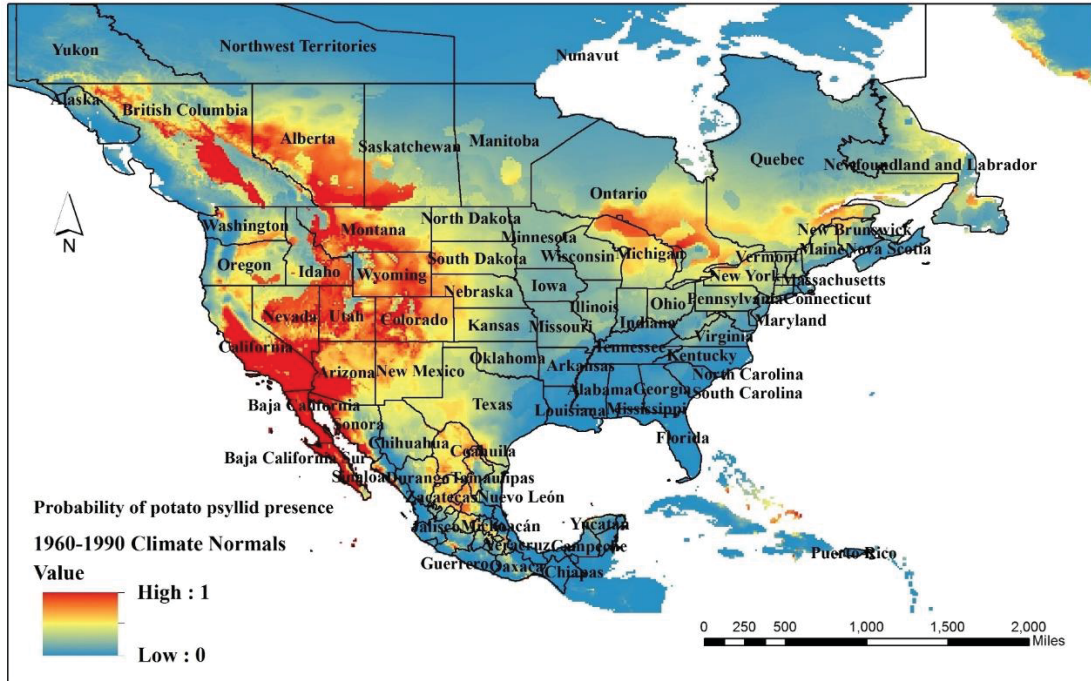
Figure 3.5. Geographical distribution of potato psyllid presence probability in North America, projected to the BIOCLIM layers generated using the 1960-1990 Climate Normal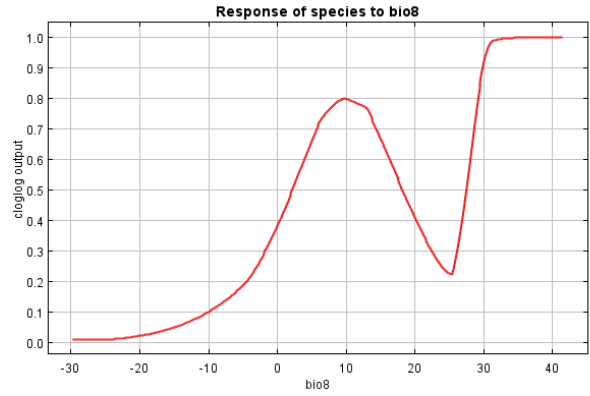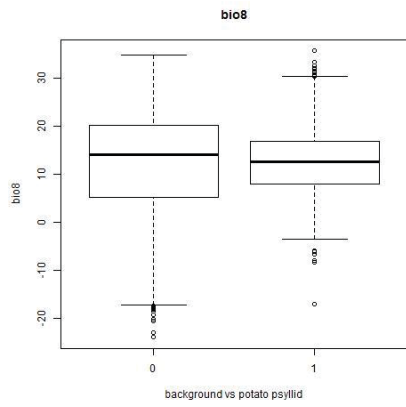s data (These results concern only suitable climate, and do not include predictions based on whether suitable food plants would be present.)
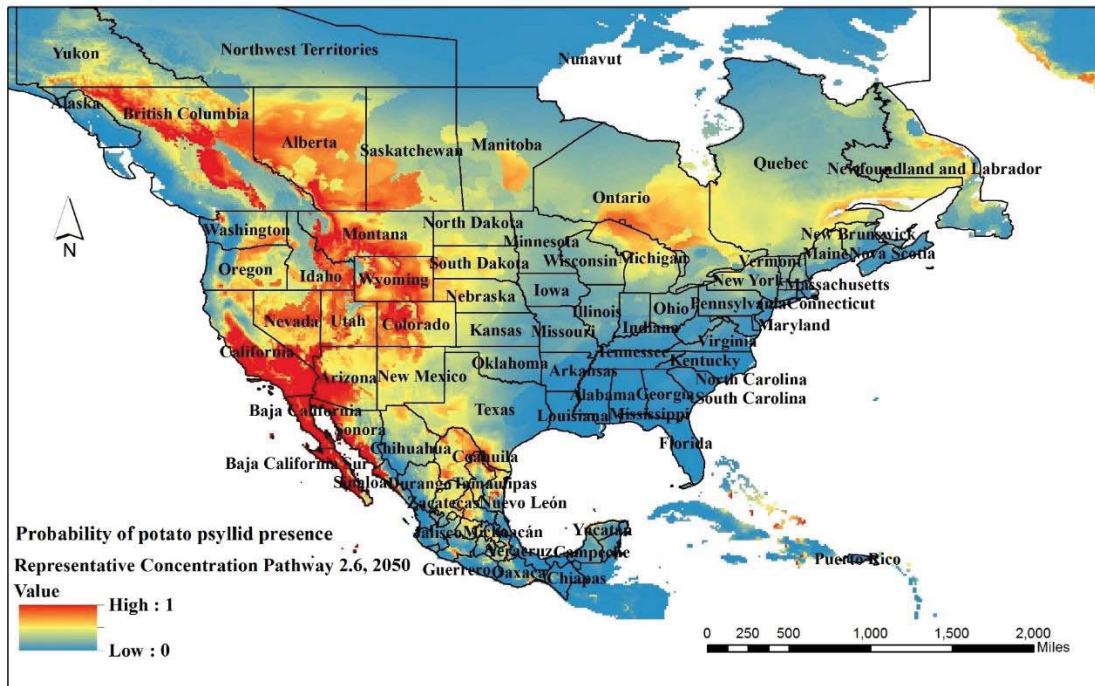
Figure 3.6. Geographical distribution of potato psyllid presence probability in North America, projected to the BIOCLIM layers generated using the RCP2.6, 2050 climate data

93
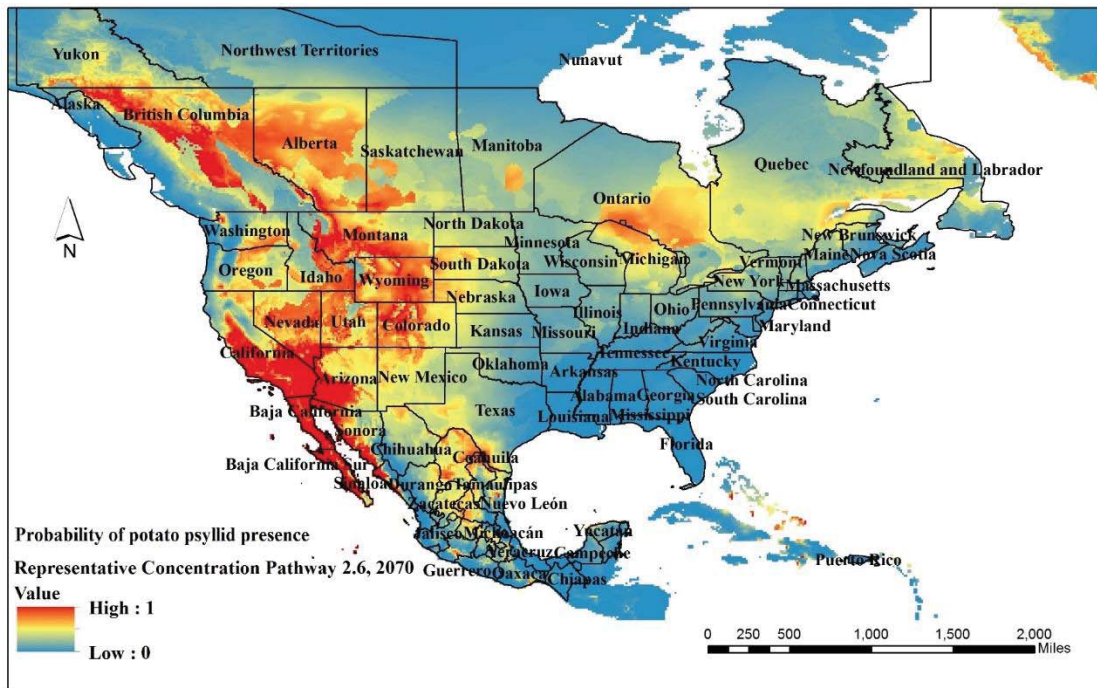
Figure 3.7. Geographical distribution of potato psyllid presence probability in North America, projected to the BIOCLIM layers generated using the RCP2.6, 2070 climate data
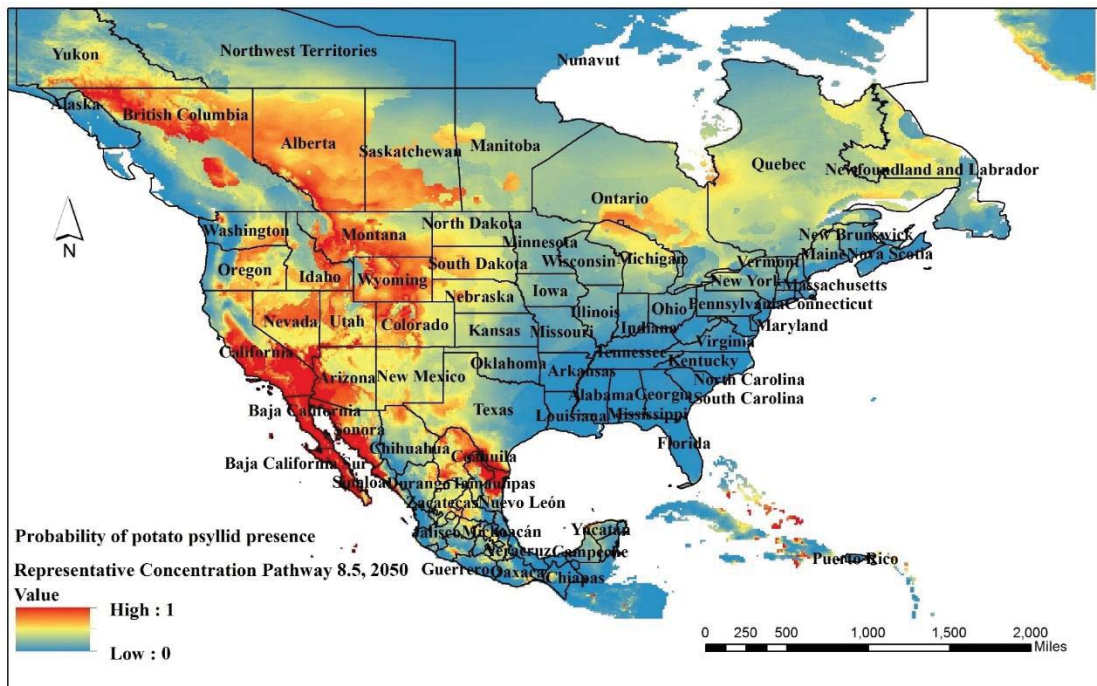
Figure 3.8. Geographical distribution of potato psyllid presence probability in North America, projected to the BIOCLIM layers generated using the RCP8.5, 2050 climate data
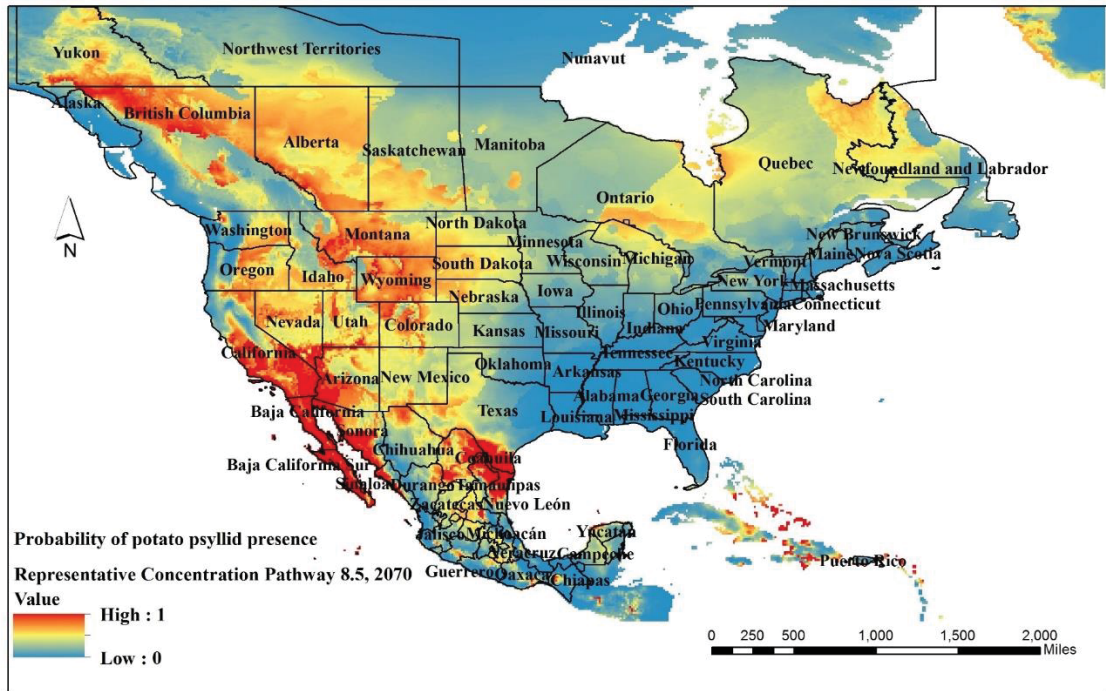
Figure 3.9. Geographical distribution of potato psyllid presence probability in North America, projected to the BIOCLIM layers generated using the RCP8.5, 2070 climate data

Chapter 4: Conclusions

Potato psyllids are naturally occurring herbivores that have been native to North America since it was first identified in 1901. Infestations of potato crops can induce two economically significant potato diseases, psyllid yellows, through feeding by immature stages, and zebra chip, through transmission of the bacterial plant pathogen *Candidatus* Liberibacter solanacearum. My thesis project analyzes the climate pattern of historical (and some current) locations of the potato psyllid population and psyllid yellows, patterns and relationships which are not well understood at continental or regional scales, although many observations of infestations and the probable effects of weather at the scale of states and provinces exist.

Outbreaks of psyllid yellows are caused by activity of potato psyllid nymphs. Historical records serve as evidence of actively reproducing and growing populations. I explored the data pattern of climatic indices related to psyllid yellows. I used weather data to calculate indices for the species and disease records by using standard BIOCLIM, and CLIMDEX indices to describe different aspects of the climate. I applied the Random Forest ensemble learning method to identify important climate indicators, and visualize the relationship between the probability of psyllid yellows disease incidents conditioned on the climate variables of interest using the fitted Random Forest model. This was essentially a test of whether these standard climate indices, and this methodology, could be applied to this pest biogeography problem to improve our understanding and prediction of possible future range and intensity of the impact on agriculture.

The results of the Random Forest model show that the climate indices related to winter temperatures, heat accumulation, and precipitation, can be used to distinguish between

the psyllid yellows observations and "no psyllid yellows" observations. Compared to "no psyllid yellows" observations, psyllid yellows outbreak data shows that a warmer annual temperature and warmer winter conditions than the "no psyllid yellow" records. The psyllid yellows records have a higher May precipitation amount and extremes than the no psyllid yellows records, whereas in June, the psyllid yellows record locations have lower precipitation. The means of the total precipitation in May were 53.9 mm for the no psyllid yellows observations, and 86.4 mm for the psyllid yellows observations, whereas mean of the total precipitation in June are 91.7 mm for the no psyllid yellows observations, and 77.2 mm for the psyllid yellows observations. Degree-day accumulations and temperature indices in June show that psyllid yellows cases were found more often in sites that had experienced higher heat accumulation. A pattern of higher heat accumulation in September and a higher temperature in September at the potato psyllid outbreak locations are apparent in the data. Psyllid yellows records also have fewer frost days and more warm summer days than the no psyllid yellows records. This data pattern is supported by models and understanding of potato psyllid population biology, since heat accumulation is positively related to more rapid and successful potato psyllid development, and low temperature negatively influence insect survival. In conclusion, psyllid yellows cases have a warmer annual temperature, a warmer winter, warmer June, warmer September, and moderate precipitation in May and June.

Weather conditions are known to influence the seasonality and movement of potato psyllid field populations. However, influence of environments on its probability of expansion and establishment of a wider distribution in Canada and the United States have not been studied broadly. By contrasting the species presence environment and the

97

background environments at multiple locations and years, the probability of potato psyllid presence can be modeled using Maxent. I tested use of the Maxent model to estimate the probability of potato psyllid presence, by comparing species presence data to background environment. The Maxent model showed that annual precipitation has the highest contribution to distinguish species presence and the background environment. Potato psyllid presence has a lower annual total precipitation than the background environment, and potato psyllid presence has a lower precipitation in the warmest quarter compared to the background environment. Although annual mean temperature has the lowest relative importance among significant variables in the Maxent model, the minimum, 25th percentile, median, and 75th percentile of annual mean temperature where the potato psyllid is present are higher than the background environment. Under different climate change projections, I studied with this model, suitability of the species to central United States decreases, however, the unsuitable area shrinks in northern Canada. These results reflect climate only, and other factors, such as presence of host plants (wild or crop) are necessary in making predictions. However, a decrease in the area in Canada unsuitable for potato psyllid is predicted, and an expansion into some area.

Differences in climatic indices for psyllid yellows cases and no psyllid yellows cases, and differences in climatic indices between potato psyllid presence and background data, indicate the environmental restrictions on the insect population and distribution. These climate indices integrate the effects of environment to allow some prediction of overall trends and expected changes in suitable habitat and conditions. If climate conditions do not restrict or favor growth of potato psyllid populations, indices related to those conditions will show no difference between the positive and negative cases, except for

the situation with insufficient information from the data. If the positive and negative cases show a distinct pattern of specific climate index, it means the index possibly related to an environmental condition that restricts or favors the potato psyllid development, survival, and distribution. Further confirmation on those data pattern, can be done through lab experiments studying the insect biology or more detailed field sampling and observations. Specific forecasts require detailed and local population models and weather data, which are a valuable tool to be used in addition to the broader analysis based on climate indices.

With the knowledge gained from the thesis results, a model system of rating potato psyllid infestation risk using standard climatic indices, based on the knowledge of the species distribution history, biology and ecology, seems possible to develop and potentially useful for large-scale understanding of the problem. More study needs to be done to examine whether the models fitted by the psyllid yellows and potato psyllid data in this study can make consistent prediction by testing it with new potato psyllid presence and outbreak data.

References

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M. & Vazquez-Aguirre, J.L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111, 1-22.

Bextine, B., Arp, A., Flores, E., Aguilar, E., Lastrea, L., Gomez, F.S., Powell, C. & Rueda, A. (2013). First report of zebra chip and '*Candidatus* Liberibacter solanacearum' on potatoes in Nicaragua. *Plant Disease*, 97, 1109.

Binkley, A.M. (1929). Transmission studies with the new psyllid-yellows disease of solanaceous plants. *Science*, 70, 615.

Blood, H.L. & Christiansen, R.M. (1940). The 1939 tomato disease situation in Utah. *Plant Disease Reporter*, 24, 292.

Booth, T.H. (1985). A new method for assisting species selection. *The Commonwealth Forestry Review*, 64, 241-250.

Booth, T.H., Nix, H.A., Busby, J.R. & Hutchinson, M.F. (2014). Bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, 20, 1-9.

Boschetti, L., Stehman, S.V. & Roy, D.P. (2016). A stratified random sampling design in space and time for regional to global scale burned area product validation. *Remote Sensing of Environment*, 186, 465-478.

Bosso, L., De Conno, C. & Russo, D. (2017a). Modelling the risk posed by the zebra mussel *Dreissena polymorpha*: Italy as a case study. *Environmental Management*, 60, 304-313.

Bosso, L., Luchi, N., Maresi, G., Cristinzio, G., Smeraldo, S. & Russo, D. (2017b). Predicting current and future disease outbreaks of *Diplodia sapinea* shoot blight in Italy: species distribution models as a tool for forest management planning. *Forest Ecology and Management*, 400, 655-664.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.

Bronaugh D. (2015) R package climdex.pcic: PCIC implementation of Climdex routines. Victoria, British Columbia, Canada: Pacific Climate Impact Consortium.

Brown, J.K., Rehman, M., Rogan, D., Martin, R.R. & Idris, A.M. (2010). First Report of "Candidatus Liberibacter psyllaurous"(synonym "Ca. L. solanacearum") Associated with'Tomato Vein-Greening'and'Tomato Psyllid Yellows' Diseases in Commercial Greenhouses in Arizona. *Plant Disease*, 94, 376-376.

Busby, J.R. (1991). BIOCLIM - a bioclimate analysis and prediction system. *Plant Protection Quarterly*, 6, 8-9.

Butler, C.D. & Trumble, J.T. (2012). The potato psyllid, *Bactericera cockerelli* (Sulc) (Hemiptera: Triozidae): life history, relationship to plant diseases, and management strategies. *Terrestrial Arthropod Reviews*, 5, 87-111.

Cameron, P.J., Surrey, M.R., Wigley, P.J., Anderson, J.A.D., Hartnett, D.E. & Wallace, A.R. (2009). Seasonality of *Bactericera cockerelli* in potatoes (*Solanum tuberosum*) in South Auckland, New Zealand. *New Zealand Journal of Crop and Horticultural Science*, 37, 295-301.

Carter, W. (1939). Injuries to plants caused by insect toxins. *The Botanical Review*, 5, 273.

Charles, B.Y., Richard, C., Elise, F.Z., Royle, J.A., James, D.N., Evan, H.C.G. & Sophie, V. (2013). Presence‐only modelling using Maxent: when can we trust the inferences? *Methods in Ecology and Evolution*, 4, 236-243.

Chefaoui, R.M. & Lobo, J.M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210, 478-486.

Compere, H. (1916). Notes on the tomato psylla. *Monthly Bulletin. California Commission of Horticulture*, 5, 189-191.

Cranshaw, W.S. (1993). An annotated bibliography of Potato/Tomato Psyllid, *Paratrioza cockerelli* (Sulc) (Homoptera; Psyllidae). *C TB93-5*.

Crosslin, J.M. & Bester, G. (2009). First report of '*Candidatus* Liberibacter psyllaurous' in zebra chip symptomatic potatoes from California. *Plant Disease*, 93, 551.

Crosslin, J.M., Hamm, P.B., Eggers, J.E., Rondon, S.I., Sengoda, V.G. & Munyaneza, J.E. (2012a). First report of zebra chip disease and "*Candidatus* Liberibacter solanacearum" on potatoes in Oregon and Washington State. *Plant Disease*, 96, 452-453.

Crosslin, J.M. & Munyaneza, J.E. (2009). Evidence that the zebra chip disease and the putative causal agent can be maintained in potatoes by grafting and *in vitro*. *American Journal of Potato Research*, 86, 183-187.

Crosslin, J.M., Olsen, N. & Nolte, P. (2012b). First report of zebra chip disease and "*Candidatus* Liberibacter solanacearum" on potatoes in Idaho. *Plant Disease*, 96, 453-453.

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007). Random Forests for classification in ecology. *Ecology*, 88, 2783-2792.

Daniels, L.B. (1937). Controlling Colorado potato pests. *Bulletin*, 437, 35.

Daniels, L.B. (1939). One of the worst outbreaks of psyllid yellows known in Colorado. *Plant Disease Reporter*, 23, 3-4.

Diaz-Uriarte, R. (2007). GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. *BMC Bioinformatics*, 8.

Diaz-Uriarte, R. & de Andrés, S.A. (2005). Variable selection from random forests: application to gene expression data. *arXiv preprint q-bio/0503025*.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27-46.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A.

Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.

Elith, J., Phillips, S.J., Hastie, T. & Dudik, M. (2011). A statistical explanation of Maxent for ecologists. *Diversity & Distributions*, 17, 43-57.

Franklin, J. & Miller, J.A. (2009). *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, New York, USA.

Friedman, J., Hastie, T. & Tibshirani, R. (2001). *The elements of statistical learning*. Springer Series in Statistics Springer, Berlin.

Glick, P.A. (1939). The Distribution of Insects, Spiders, and Mites in the Air. In: *USDA Technical Bulletin*, p. 150.

Goss, R.W. (1938). Psyllid yellows in central and eastern Nebraska. *Plant Disease Reporter*, 22, 327-328.

Greenway, G. (2014). Economic impact of zebra chip control costs on grower returns in seven US states. *American Journal of Potato Research*, 91, 714-719.

Guenthner, J. & Greenway, G. (2010). Zebra chip economics. *2010 Zebra Chip Reporting Session*, 93.

Harrison, J. (1992). Effects of the aerial environment on late blight of potato foliage–a review. *Plant Pathology*, 41, 384-416.

Hartman, G. (1937). A study of potato psyllid yellows in Wyoming. (ed. University of Wyoming, AES) Laramie, Wyoming, USA.

Henne, D., Paetzold, L., Workneh, F. & Rush, C. (2010). Evaluation of potato psyllid cold tolerance, overwintering survival, sticky trap sampling, and effects of liberibacter on potato psyllid alternate host plants. *Proc. 10th Annual Zebra Chip Rep. Session. Dallas, TX*, 149-153.

Heung, B., Bulmer, C.E. & Schmidt, M.G. (2014). Predictive soil parent material mapping at a regional-scale: a Random Forest approach. *Geoderma*, 214, 141-154.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. (2017). R package 'dismo', species distribution modeling. Available online at: http://cran.r project.org/web/packages/dismo/index.html.

Hill, R.E. (1947). An unusual weather sequence accompanying the severe potato psyllid outbreak of 1938 in Nebraska. *Journal of the Kansas Entomological Society*, 20, 88-92.

Horton, D.R., Cooper, W.R., Munyaneza, J.E., Swisher, K.D., Echegaray, E.R., Murphy, A.F., Rondon, S.I., Wohleb, C.H., Waters, T.D. & Jensen, A.S. (2015). A New problem and old questions: potato psyllid in the Pacific Northwest. *American Entomologist*, 61, 234-244.

Horton, D.R., Munyaneza, J.E., Swisher, K.D., E., E., Muller, A.F., Rondon, S.I., Sengoda, V.G., Neven, L.G. & Jensen, A.S. (2014). What is the source of potato psyllids colonizing Washington, Oregon, and Idaho potato fields? *Potato Progress*, 14, 1-6.

Hoyman, W.G. (1943a). Reports on other potato diseases. *Plant Disease Reporter*, 27, 423-430.

Hoyman, W.G. (1943b). Reports on potato diseases. *Plant Disease Reporter*, 27, 542-550.

Hoyman, W.G. (1943c). Summary of plant diseases observed on Arizona and New Mexico. *Plant Disease Reporter*, 27, 353-359.

Jensen, D.D. (1954). Notes on the potato psyllid, *Paratrioza cockerelli* (Sulc.). *Pan-Pacific Entomologist*, 30, 161-165.

Johnson, D.L. & participants, n. (2016). Newsletter of the Canadian potato psyllid and zebra chip monitoring network. In: *Can Psyll Net*. University of Lethbridge, pp. 1-12.

Karl, T.R., Nicholls, N. & Ghazi, A. (1999). Clivar/GCOS/WMO workshop on indices and indicators for climate extremes workshop summary. In: *Weather and Climate Extremes*. Springer, pp. 3-7.

Knowlton, G.F. (1934). Potato psyllid investigations. *Proceedings of the Utah Academy of Sciences*, 11, 261-265.

Krause, R., Massie, L. & Hyre, R. (1975). Blitecast: a computerized forecast of potato late blight. *Plant Disease Reporter*, 59, 95-98.

Kriticos, D.J., Jarošik, V. & Ota, N. (2014). Extending the suite of bioclim variables: a proposed registry system and case study using principal components analysis. *Methods in Ecology and Evolution*, 5, 956-960.

Kuhn, M. (2016). caret: Classification and regression training. R package version 6.0-71. Available online at: https://cran.r-project.org/web/packages/caret, p. 05003.

Lactin, D.J., Holliday, N., Johnson, D. & Craigen, R. (1995). Improved rate model of temperature-dependent development by arthropods. *Environmental Entomology*, 24, 68-75.

List, G.M. (1925). The tomato psyllid, *Paratrioza cockerelli* (Sulc.). *Colorado State Entomologist Circular*, 47, 16.

List, G.M. (1939). The effect of temperature upon egg deposition, egg hatch and nymphal development of *Paratrioza cockerelli* (Sulc.). *Journal of Economic Entomology*, 32, 30-36.

Livneh, B., Rosenberg, E.A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K.M., Maurer, E.P. & Lettenmaier, D.P. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. *Journal of Climate*, 26, 9384-9392.

Manson, G.F. (1944). Diseases of vegetable and field crops: Psyllid yellows (*Paratrioza cockerelli*). *Canadian Plant Disease Survey*, 24, 45-85.

Marritt, J.W. (1935). Diseases of vegetable and field crops: Psyllid yellows (virus). *Canadian Plant Disease Survey*, 15, 38.

Marritt, J.W. (1936). Diseases of vegetable and field crops: Psyllid yellows. *Canadian Plant Disease Survey*, 16, 37.

Marritt, J.W. (1937). Diseases of vegetable and field crops: Psyllid yellows (virus). *Canadian Plant Disease Survey*, 17, 38.

Marritt, J.W. (1938). Diseases of vegetable and field crops: Psyllid yellows. *Canadian Plant Disease Survey*, 18, 50.

Marritt, J.W. (1939). Diseases of vegetable and field crops: Psyllid yellows. *Canadian Plant Disease Survey*, 19, 55.

Marritt, J.W. (1940). Diseases of vegetable and field crops: Psyllid yellows (*Paratrioza cockerelli*). *Canadian Plant Disease Survey*, 20, 45.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. & Houston, T.G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29, 897-910.

Menne, M.J., Durre, I.B., Korzeniewski, S., McNeal, K., Thomas, X., Yin, S., Anthony, R., Ray, R.S., Vose, B., Gleason, E. & Houston, T.G.H. (2016). Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.22. NOAA National Climatic Data Center.

Miles, G.P., Samuel, M.A., Chen, J., Civerolo, E.L. & Munyaneza, J.E. (2010). Evidence that cell death is associated with zebra chip disease in potato tubers. *American Journal of Potato Research*, 87, 337-349.

Mojtahedi, H., Santo, G., Hang, A. & Wilson, J. (1991). Suppression of root-knot nematode populations with selected rapeseed cultivars as green manure. *Journal of Nematology*, 23, 170.

Morgane, B.M., Frédéric, J., Cécile Hélène, A. & Wilfried, T. (2012). Selecting pseudo‐absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3, 327-338.

Morris, H.E. (1939). Psyllid yellows in Montana in 1938. *Plant Disease Reporter*, 23, 18.

Munyaneza, J.E. (2012). Zebra chip disease of potato: biology, epidemiology, and management. *American Journal of Potato Research*, 89, 329-350.

Munyaneza, J.E. (2015). Zebra Chip Disease, *Candidatus* Liberibacter, and potato psyllid: a global threat to the potato industry. *American Journal of Potato Research*, 92, 230-235.

Munyaneza, J.E., Sengoda, V.G., Crosslin, J.M., De la Rosa-Lozano, G. & Sanchez, A. (2009). First Report of '*Candidatus* Liberibacter psyllaurous' in potato tubers with zebra chip disease in Mexico. *Plant Disease*, 93, 552.

Murphy, A., Rondon, S. & Jensen, A. (2013). First report of potato psyllids, *Bactericera cockerelli*, overwintering in the Pacific Northwest. *American Journal of Potato Research*, 90, 294-296.

Murphy, M.A., Evans, J.S. & Storfer, A. (2010). Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology*, 91, 252-261.

Nazarenko, L., Schmidt, G.A., Miller, R.L., Tausnev, N., Kelley, M., Ruedy, R., Russell, G.L., Aleinov, I., Bauer, M., Bauer, S., Bleck, R., Canuto, V., Cheng, Y., Clune, T.L., Del Genio, A.D., Faluvegi, G., Hansen, J.E., Healy, R.J., Kiang, N.Y., Koch, D., Lacis, A.A., LeGrande, A.N., Lerner, J., Lo, K.K., Menon, S., Oinas, V., Perlwitz, J., Puma, M.J., Rind, D., Romanou, A., Sato, M., Shindell, D.T., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.S. & Zhang, J. (2015). Future climate change under RCP emission scenarios with GISS ModelE2. *Journal of Advances in Modeling Earth Systems*, 7, 244-267.

Nelson, W.R., Warrick, R.N., Kylie, D.S., James, M.C. & Joseph, E.M. (2014). Seasonal dispersal of the potato psyllid, *Bactericera cockerelli*, into potato crops. *The Southwestern Entomologist*, 39, 177-186.

Nix, H.A. (1986). A biogeographic analysis of Australian elapid snakes. In: *Atlas of elapid snakes of Australia: Australian flora and fauna series 7* (ed. Longmore, R). Bureau of Flora and Fauna Canberra, pp. 4-15.

Pearce, J.L. & Boyce, M.S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43, 405-412.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011). *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press, Princeton, New Jersey, USA.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19, 181-197.

Pitman, A.R., Drayton, G.M., Kraberger, S.J., Genet, R.A. & Scott, I.A.W. (2011). Tuber transmission of '*Candidatus* Liberibacter solanacearum' and its association with zebra chip on potato in New Zealand. *European Journal of Plant Pathology*, 129, 389-398.

Plant, R.E. (2012). *Spatial data analysis in ecology and agriculture using R*. CRC Press, Boca Raton, Florida, USA.

Plant Disease Survey. (1929). Psyllid yellows (cause undetermined). *Plant Disease Reporter Suppliement*, 12, 43.

Plant Disease Survey. (1944). Reports on diseases of potato. *Plant Disease Reporter*, 28, 819.

Plant Disease Survey. (1947). Nonparasitic and miscellaneous. *Plant Disease Reporter*, 31, 353.

Pletsch, D.J. (1939). Potato psyllids and psyllid yellows in Montana. Plant Disease Reporter, 24, 173.

Pletsch, D.J. (1947). The potato psyllid, *Paratrioza cockerelli* (Sulc), its biology and control. *Montana Agricultural Experiment Station Bulletin*, 446, 95.

Renner, I.W. & Warton, D.I. (2013). Equivalence of Maxent and poisson point process models for species distribution modeling in ecology. *Biometrics*, 69, 274-281.

Richard, B.L., Blood, H.L. & Linford, M.B. (1927). Destructive outbreak of unknown potato disease in Utah. *Plant Disease Reporter*, 11, 93-94.

Richards, B.L. (1928a). Disease again threatens Utah potato crops. *Plant Disease Reporter*, 12, 21.

Richards, B.L. (1928b). Potato psyllid yellows in Utah. *Plant Disease Reporter*, 12, 43.

Richards, B.L. & Blood, H.L. (1933). Psyllid yellows of the potato. *Journal of Agricultural Research (Washington, D.C.)*, 46, 189.

Rossum, G.v. (2005). Python tutorial. In: *Technical Report CS-R9526* Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Sanford, G.B. (1934). A malady of the potato in Alberta similar to psyllid yellows. *Scientific Agriculture*, 15, 46-48.

Sanford, G.B. (1941). Diseases of vegetable and field crops: Psyllid yellows (*Paratrioza cockerelli*). *Canadian Plant Disease Survey*, 21, 44.

Sanford, G.B. (1952). Phloem Necrosis of Potato Tubers Associated with Infestation of Vines by *Paratrioza cockerelli* Sulc. *Scientific Agriculture*, 32, 433-439.

Scarpone, C., Schmidt, M.G., Bulmer, C.E. & Knudby, A. (2017). Semi-automated classification of exposed bedrock cover in British Columbia's southern mountains using a Random Forest approach. *Geomorphology*, 285, 214-224.

Schaal, L.A. (1938). Some factors affecting the symptoms of the psyllid yellows disease of potatoes. *American Potato Journal*, 15, 193-206.

Senay, S.D., Worner, S.P. & Ikeda, T. (2013). Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PLOS ONE*, 8, e71218.

Sengoda, V.G., Munyaneza, J.E., Crosslin, J.M., Buchman, J.L. & Pappu, H.R. (2010). Phenotypic and etiological differences between psyllid yellows and zebra chip diseases of potato. *American Journal of Potato Research*, 87, 41-49.

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Strickland, E. (1953). An annotated list of the Hemiptera (SL) of Alberta. *The Canadian Entomologist*, 85, 193-214.

Strickland, E.H. (1938). The Chermidae, (Homoptera), of Alberta. *The Canadian Entomologist*, 70, 200-206.

Strickland, E.H. (1939). Further notes on Psyllidae taken in Alberta. *The Canadian Entomologist*, 71, 212-215.

Sulc., K. (1909). *Trioza cockerelli* n. sp., a novelty from North America, being also of economic importance. *Acta Societatis Entomologicae Bohemiae*, 6, 102-108.

Svetnik, V., Liaw, A., Tong, C. & Wang, T. (2004). Application of Breiman's Random forest to modeling structure-activity relationships of pharmaceutical molecules. In: *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004. Proceedings* (eds. Roli, F, Kittler, J & Windeatt, T). Springer Berlin Heidelberg Berlin, Heidelberg, pp. 334-343.

Swisher, K.D., Munyaneza, J.E. & Crosslin, J.M. (2013). Temporal and spatial analysis of potato psyllid haplotypes in the United States. *Environmental Entomology*, 42, 381-393.

Swisher, K.D., Sengoda, V.G., Dixon, J., Munyaneza, J.E., Murphy, A.F., Rondon, S.I., Thompson, B., Karasev, A.V., Wenninger, E.J., Olsen, N. & Crosslin, J.M. (2014). Assessing potato psyllid haplotypes in potato crops in the Pacific Northwestern United States. *American Journal of Potato Research*, 91, 485-491.

Syfert, M.M., Serbina, L., Burckhardt, D., Knapp, S. & Percy, D.M. (2017). Emerging New Crop Pests: Ecological Modelling and Analysis of the South American Potato Psyllid *Russelliana solanicola* (Hemiptera: Psylloidea) and Its Wild Relatives. *PloS one*, 12, e0167764.

Tervet, I.V. (1944). Other reports on potato diseases. *Plant Disease Reporter*, 28, 819.

Texas A&M AgriLife Research & Extension. (2016). SCRI Zebra Chip. Integrated Research & Management Program.

Thinakaran, J., Cooper, W.R., Munyaneza, J.E. & Horton, D. (2015a). Matrimony vine/Goji berries-potential overwintering host for potato psyllids in the Pacific Northwest. In: *Proceedings of the Washington-Oregon Potato Conference*, pp. 27-29.

Thinakaran, J., Pierson, E., Kunta, M., Munyaneza, J.E., Rush, C.M. & Henne, D.C. (2015b). Silverleaf nightshade (*Solanum elaeagnifolium*), a reservoir host for '*Candidatus* Liberibacter solanacearum', the putative causal agent of zebra chip disease of potato. *Plant Disease*, 99, 910-915.

Tran, L.T., Worner, S.P., Hale, R.J. & Teulon, D.A. (2012). Estimating development rate and thermal requirements of *Bactericera cockerelli* (Hemiptera: Triozidae) reared on potato and tomato by using linear and nonlinear models. *Environmental Entomology*, 41, 1190-1198.

Vandermeer, J.H. (1972). Niche theory. *Annual Review of Ecology and Systematics*, 3, 107-132.

Vikram, B., Johnson, D.L., Kawchuk, L. & Meers, S. (2016). Occurrence of potato psyllids in Manitoba. In: *PMR Report: Section C-Insect Pests*

Wallis, R.L. (1946). Seasonal Occurrence of the Potato Psyllid in the North Platte Valley. *Journal of Economic Entomology*, 39, 689-694.

Wallis, R.L. (1955). Ecological studies on the potato psyllids as a pest of potatoes. *USDA Technical Bulletin*, 1107, 1-25.

Wang, T., Wang, G., Innes, J., Nitschke, C. & Kang, H. (2016). Climatic niche models and their consensus projections for future climates for four major forest tree species in the Asia–Pacific region. *Forest Ecology and Management*, 360, 357-366.

Wellington, W., Johnson, D. & Lactin, D. (1999). Weather and insects. *Ecological entomology, 2nd edn. Wiley, New York*, 313-353.

Whipple, S., Bradshaw, J. & Harveson, R. (2013). Regional variation in potato psyllid cold tolerance. *2013 Zebra Chip Reporting Session*, 16.

Xia, X.-Q., McClelland, M. & Wang, Y. (2010). PypeR, A Python Package for Using R in Python. *Journal of Statistical Software; Vol 1, Code Snippet 2*, 1.

Yang, X.-B., Zhang, Y.-M., Henne, D.C. & Liu, T.-X. (2013). Life Tables of *Bactericera cockerelli* (Hemiptera: Triozidae) on tomato under laboratory and field conditions in Southern Texas. *Florida Entomologist*, 96, 904-913.

Yang, X.B., Zhang, Y.M., Hua, L., Peng, L.N. & Munyaneza, J.E. (2010). Repellency of selected biorational insecticides to potato psyllid, *Bactericera cockerelli* (Hemiptera: Psyllidae). *Crop Protection*, 29, 1320-1324.

Zhou, X.H., Li, S.L., Tian, F., Cai, B.J., Xie, Y.M., Pei, Y., Kang, S., Fan, M. & Li, J.P. (2012). Building a disease risk model of osteoporosis based on traditional Chinese medicine symptoms and western medicine risk factors. *Statistics in Medicine*, 31, 643-652.
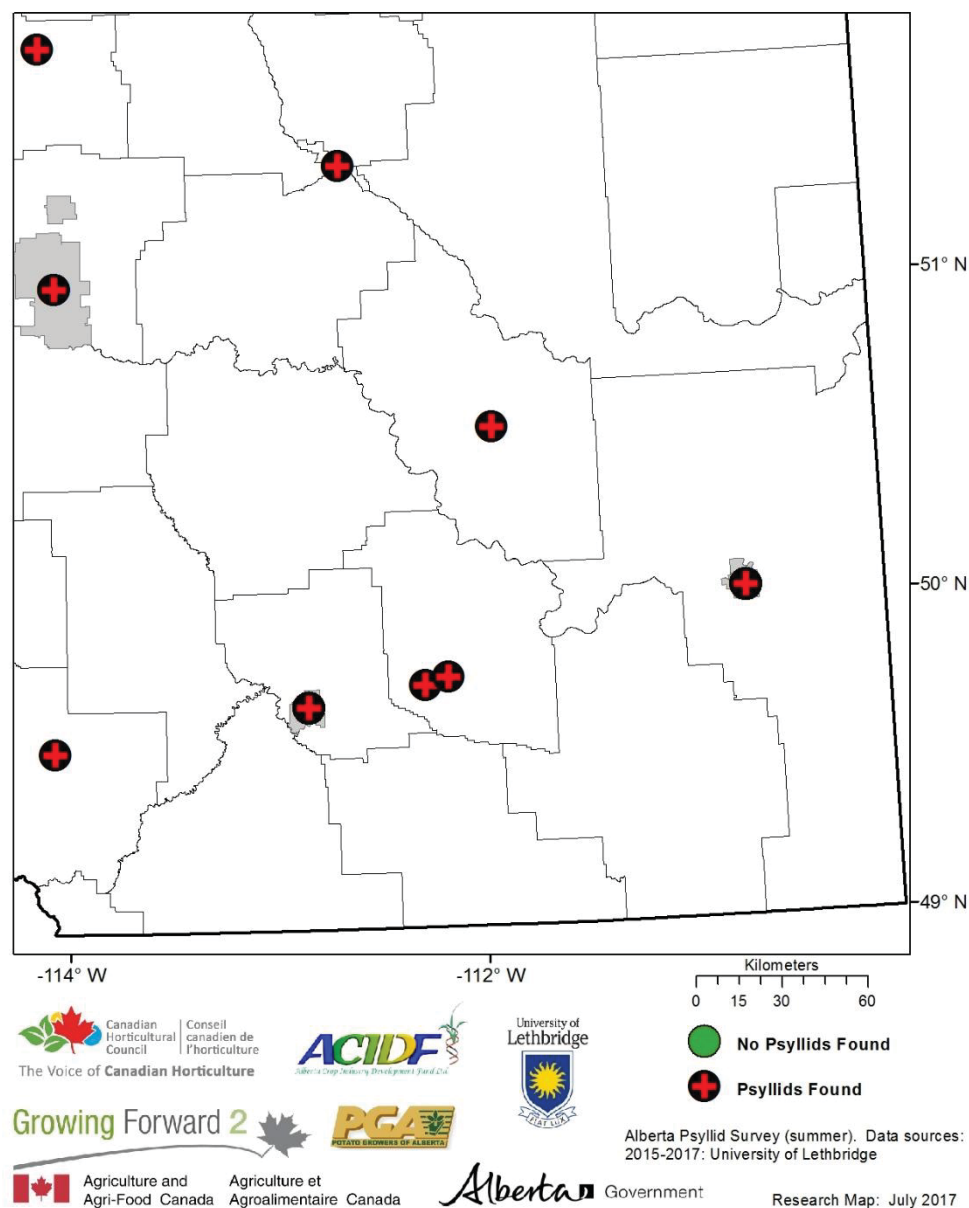
Figure A1. Potato psyllid observations from the Canadian Plant Disease Survey, entomology and agriculture literature articles (Marritt 1935, 1936, 1937, 1938; Strickland 1938; Marritt 1939; Strickland 1939; Marritt 1940; Sanford 1941; Manson 1944; Sanford 1952) (data provided by: Qing Xia, map made by: Celeste Barnes)
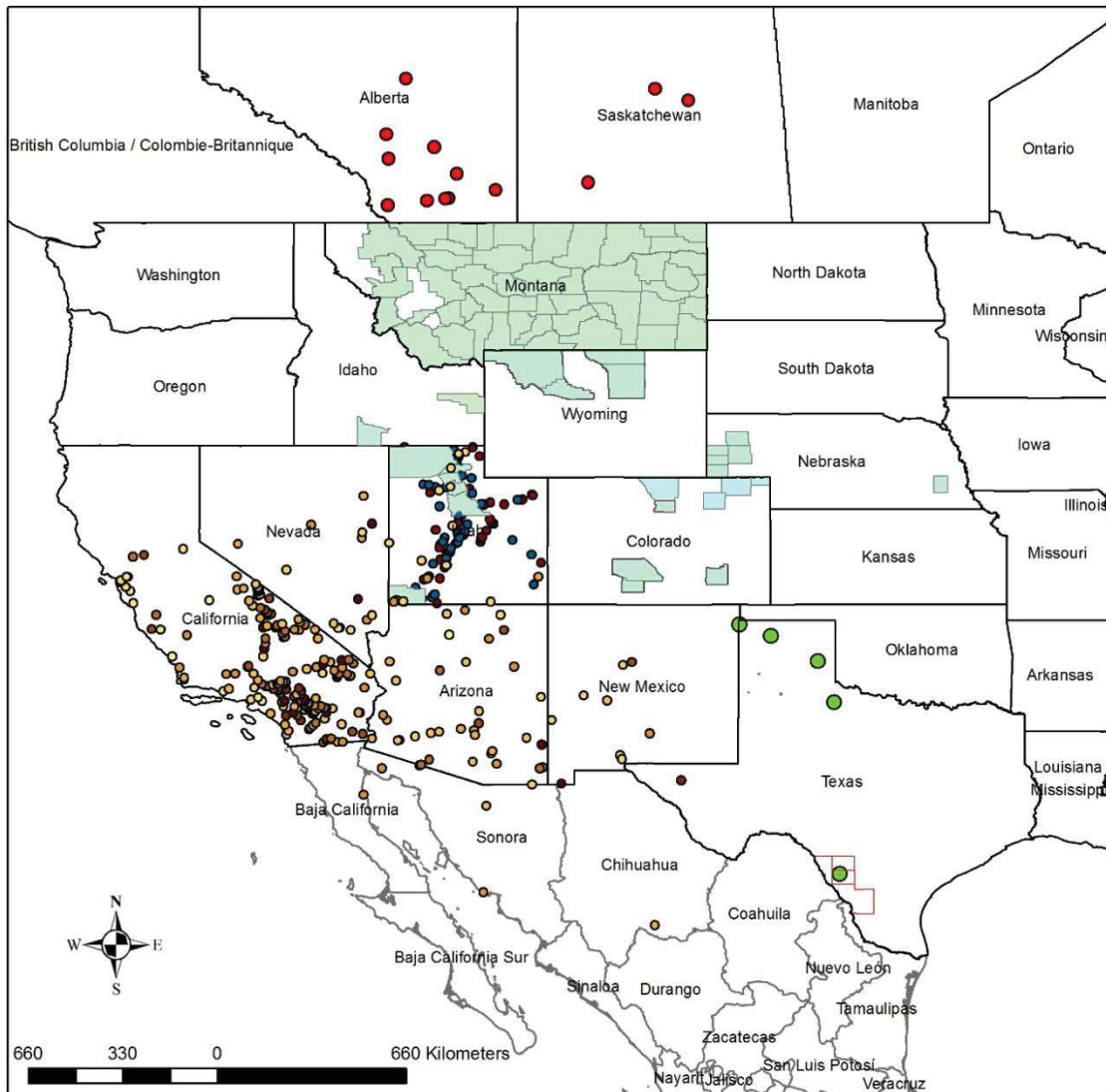
Figure A2. Georeferenced museum records and literature records of potato psyllid (town, city, and county) (data collected, and mapped by: Qing Xia)