Does Climate Change Affect the Spread of Beech Bark Disease Across Ontario? A Predictive Forecast Model on the Spread of Beech Bark Disease.

Authors: Kaylee Bigelow, Cameron Guitard, Jonah Stegman
Date: April 16, 2021
Course: GEOG*4480
Abstract

Problem Context

Purpose of the Research

Research Objectives

Study Area

Methods

  Research Approach to find variables
  Climate data
  Preprocessing of Beech Bark Disease (BBD) and Beech Scale Data
  Logistic Regression
  Selection of new sites that will be used to predict BBD movement

Results and Discussion

  1. Identify variables related to the spread of beech bark disease (BBD)
  2. Develop a logistic regression model to map historical movement of beech bark disease with respect to climate factors
  3. Predict historical movement of beech bark disease based on climate factors using the developed model
  4. Apply different climate scenarios based on IPCC climate projections to the model to predict future climate change effects on beech bark disease movement
  5. Evaluate the strengths and weaknesses of the model

Conclusions

Acknowledgments

References
Abstract

Beech bark disease has spread across Ontario causing disastrous results for American beech tree stocks. With diseases like Armillaria root disease, which is predicted to get worse with climate change (Frankel, Juzwik, & Koch, 2012), it begs the question; is there also a link between climate change and beech bark disease? Using a spatial GIS-based analysis and logistic regression, this project looks to understand the links between the disease and climate change in Ontario, Canada. Four possible models were created: Model 1 (average yearly precipitation, and average yearly temperature), Model 2 (average yearly precipitation, average yearly temperature, number of infected trees with beech bark disease within 10km, and number of infected trees with beech scale), Model 3 (average yearly precipitation, average yearly temperature, and number of infected trees with beech scale within 10km), and Model 4 (average yearly precipitation, average yearly temperature, and number of infected trees with beech bark disease within 10km), with Model 4 being selected for the logistic regression. This model was selected due to it having the best AIC score. The model was then run on historical infected points and possible future infected points. The possible future infected points were run through 6 different climate scenarios, ranging from extreme highs for temperature and precipitation to extreme lows. The model produced a 65% success rate in predicting historically infected points. For possible future infected trees, the results showed that precipitation has little to no impact on the time a tree takes to be infected. In addition, increased temperature was shown to have a positive impact on time to infection.

Problem Context

Neonectria faginata, also known as beech bark disease, is a destructive combination of the invasive insect Cryptococcus fagisuga, also known as the beech scale, and an invasive phytopathogenic fungus Neonectria faginata (Ontario Ministry of Natural Resources and Forestry, 2020). An infected tree will experience a reduction in tree growth, deformation of the tree, and a decrease in its bark quality along with its mast production; all of which lead to the premature death of the infected tree (Ontario Ministry of Natural Resources and Forestry, 2020). These tree deaths can have a drastic effect on a forest’s ecosystem, since the beech tree nuts are a food source to birds and other animals (Lovett, Canham, Arthur, Weathers, & Fitzhugh, 2006).

Beech bark disease originates from Europe and was introduced to the Halifax, Nova Scotia region sometime around 1890 (Morin, Liebhold, Tobin, Gottschalk, & Luzader, 2007). It then first spread to Ontario in 1990 and it has continued to spread west, infecting beech trees as far as Lake Michigan (Morin et al, 2007). Determining what drives the spread of beech bark disease is paramount in controlling the spread of this disease and protecting the beech bark trees.
There are three sequential stages of beech bark disease spreading which consist of the advanced front, the killing front, and the aftermath stages (Cale, Garrison-Johnston, Teale, & Castello, 2017). The advanced stage consists of a big increase in beech scale colonizing unaffected beech trees and has a low tree mortality rate (Ontario Ministry of Natural Resources and Forestry, 2020). After the advanced stage comes the killing front stage, with the highest percentage of tree mortality. The killing front stage consists of a greater increase of beech scale infestation, a high *neocentric faginata* infection rate, and the severity of the disease is at its peak (Cale et al, 2017). In this stage, *neocentric faginata* start to grow on the trees through the cracks in the bark caused by beech scale infestations (Ontario Ministry of Natural Resources and Forestry, 2020). The *neocentric faginata* spread to other beech trees by wind and rain in this stage (Ontario Ministry of Natural Resources and Forestry, 2020). The aftermath stage consists of a low beech scale population and a dispersed but persistent population of *neonectria faginata* (Kasson & Livingston, 2011). From the different stages of the disease spread, the beech scale insect is the main driving force for the spread of beech bark disease.

These stages are extremely well documented, however, there is little research that investigates if any of these stages are impacted by climate change. It is known that insects can be drastically affected by climate change and climate factors, like temperature and precipitation (Liebhold & Bentz, 2011). There is also a known connection between climate change and diseases (Frankel, Juzwik, & Koch, 2012). Knowing these two facts, it is likely that beech bark disease is no different and is directly impacted by the changing climate of Ontario. This connection can be checked by using a logistic regression model to predict future movement of the disease. A logistic regression model allows for future points to be given a likelihood of BBD spreading to them based on climate factors.

**Purpose of the Research**

Insect populations in Ontario are likely to increase with climate change, as both high temperature and low precipitation levels are known to help increase insect population (Liebhold & Bentz, 2011). Since the beech scale insect is one of the driving forces on the spread of beech bark disease, it begs the question; is the spread of beech bark disease affected by climate change? Does precipitation and temperature affect the disease rate of spread? Answering these questions to accurately estimate the spread of beech bark disease in current (1999-2019) and future (2020-2029) climates is the goal of this paper.
Research Objectives

1) Identify climate and spatial variables related to the spread of beech bark disease
2) Develop a logistic regression model to map historical movement of beech bark disease with respect to climate factors
3) Predict historical movement of beech bark disease based on climate factors using the developed model
4) Apply different climate scenarios based on Intergovernmental Panel on Climate Change’s (IPCC) climate projections to the model to predict future climate change effects on beech bark disease movement
5) Evaluate the strengths and weaknesses of the model

Study Area

The specific region of Ontario, Canada was chosen because of the historical research that exists on the topic from not only the NRCan (Natural Resources Canada) and Ontario Ministry of Natural Resources and Forestry (OMNRF), but also individual research teams. This factor combined with the team’s location and connections make Ontario the most reasonable area to research.

As beech trees mainly reside within Southern Ontario, with its biggest cluster near Algonquin park and a peak aboveground biomass percent of 22.36 (Beaudoin A et al, 2017), this area is the primary focus of this study as shown in Figure 1. Additionally, beech bark disease (BBD) currently resides in these southern regions and is making its way west (Morin et al, 2007). This southern region of the province also has a high volume of climate stations that track precipitation and temperature for the area. These stations allow for accurate measurements of climate information to be obtained for the model.
Methods

The methods are broken down into sections that reflect developmental stages of the project. These sections represent the different steps that were taken to complete the research. Each section will explain the different steps and tools that were used to obtain its result.

Research Approach to find variables

The research approach that was taken to find variables was split into two distinct research objectives. The first was to find what drives BBD to spread, and what climate factors would then affect that spread.

Through research, it was found that beech scale is a precursor to BBD (Ontario Ministry of Natural Resources and Forestry, 2020), the presence of beech scale in the surrounding area would then indicate that a point would become infested with BBD the next year. This leads to the idea that BBD in the surrounding area would lead to BBD in the next year (McCullough, Heyd & O’Brien, 2005). Since BBD has beech scale as a precursor (Ontario
Ministry of Natural Resources and Forestry, 2020), the presence of BBD in the surrounding area could indicate that a point would become infested with BBD the next year.

Through further research, papers were found regarding the effects of climate change on BBD. Some of these papers suggested that temperature and precipitation are major contributing factors to the rate at which BBD moves (Stephanson & Coe, 2017, Kasson & Livingston, 2011).

Climate data

The climate data for Ontario in 1990 - 2020 was retrieved from the Government of Canada’s monthly climate summaries. This was achieved by using a Python library called Selenium to navigate and download the monthly climate data from each weather station in Ontario for a given year. The data is then parsed and converted into one JSON file. From this JSON file, the data is converted into the GeoJSON (H. Butler et al, 2016) format with each weather station being a point feature. The average temperature and precipitation values for each year were calculated, keeping any no-data values as null to ensure the data was not biased towards 0, and then added as properties to the station for the related year attribute. One of the limitations of the data is that not every weather station has information for every month and every year of the study. Because of this, each year’s data has a different number of weather stations that contain relevant information. Finally, the climate data was interpolated to rasters. The method used to interpolate each year of temperature and precipitation data was inverse distance to a power of 2. Each of these steps, and how they relate to each other, can be seen below in Figure 2.

Figure 2: Flow chart showing the steps that took place to transform the climate data. This figure contains a loop on the step "Convert station to GeoJSON Point Feature" for each year, then moves on to export.
Preprocessing of Beech Bark Disease (BBD) and Beech Scale Data

The data that was used to show the historical spread of beech bark disease across Ontario comes from the OMNRF. The data is point data for the years 1999-2019, provided through eastings and northings. While this data is extensive, it only covers South Eastern Ontario, leaving half of the area in which beech trees are present in Ontario untested.

The steps to process the data, as seen in Figure 3, was first reading in the file to ArcGIS and creating a shape file. Then, to determine the distance that entails the surrounding area, 10km was chosen as described in various papers, such as that by McCullough, Heyd & O’Brien. These buffers were created around the points with BBD using ArcGIS’ buffer tool. Finally, the number of points with BBD and beech scale that were contained inside these buffers were recorded for each point.

Logistic Regression

Logistic regression was performed in R using the Generalized Linear Models tool (GLM). Four models were created and analyzed to determine which would yield proper results. The four models had a combination of the four predictors that were gathered; temperature, precipitation, number of points with BBD within 10km, and number of points with beech scale within 10km. The combinations of the predictors that were used were as follows:

1) precipitation and temperature
2) precipitation, temperature, number of points with BBD, and number of points with beech scale
3) precipitation, temperature, and number of points with beech scale
4) precipitation, temperature, and number of points with BBD

Once the logistic regression models were created following the outline in Figure 4, analysis of variance (ANOVA) tests and Akaike Information Criterion (AIC) tests were
performed. Both tests were used as they test different criteria of a model. ANOVA tests measure the variance in models using F-tests, whereas AIC estimates and compares the prediction power of the models with a penalizing factor for models with a greater number of variables. Both of these tests were performed in R, by entering the Logistic Regression models as input variables. To determine the results of an ANOVA test, the p-value is the major concern. A p-value under 0.05 tells the user that the additional variable is significant and thus the model with the variable is a better model. AIC tests all models against one another at the same time and returns AIC scores for each. The model with the lowest AIC score is then the best model in terms of prediction power. The results of these tests may not agree with each other, in which case, for this study, the best model from each of these tests will get further testing.

![Flow chart outlining the process used to create logistic regression models in R based on the input variables.](image)

Figure 4: Flow chart outlining the process used to create logistic regression models in R based on the input variables.

**Selection of new sites that will be used to predict BBD movement**

To predict the movement of BBD over the next 10 years, new sites with beech trees were selected. These sites were selected at random, provided they met the following criteria: they were close to another study point (historical or future), and the area contained beech trees. These criteria were chosen to ensure that each study point had the possibility of contracting BBD within the time frame. The data that was used to verify if a point has beech trees comes from a 2011 raster layer file with a scale of 250m (Beaudoin, Bernier, Vilemaire, Guindon, Guo, 2017).
Results and Discussion

1. Identify variables related to the spread of beech bark disease (BBD)

To identify variables related to the spread of BBD, research was completed, as described in Methods, around how BBD spreads and how the changing climate will affect that spread. From this research, 4 factors were identified: number of points infested with beech scale in the surrounding area, number of points infested with BBD in the surrounding area, total precipitation for the previous year, and average temperature for the previous year. As seen in Figure 5, rasters were created for each year of precipitation and temperature. The rasters allowed for estimated temperature and precipitation data for each point where BBD testing took place.

![Figure 5: A map of Ontario displaying precipitation average for the year 1999. B: A map of Ontario showing temperature average for the year 1999. C: A zoomed in map of Ontario displaying historical locations of BBD, and their 10km buffers that were made during preprocessing.](image)
2. Develop a logistic regression model to map historical movement of beech bark disease with respect to climate factors

Using the steps outlined in the Methods section, four logistic regression models were created to predict the presence of BBD based on the defined factors of average yearly temperature, average yearly precipitation, number of points infested with beech scale in the surrounding area, and number of points infested with BBD in the surrounding area.

Table 1: Displays the p-values associated with ANOVA tests run during model testing.

<table>
<thead>
<tr>
<th>Models being compared</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 3 vs. Model 2</td>
<td>2.75e-06</td>
</tr>
<tr>
<td>Model 4 vs. Model 2</td>
<td>0.0002468</td>
</tr>
<tr>
<td>Model 1 vs. Model 4</td>
<td>0.00299</td>
</tr>
<tr>
<td>Model 1 vs. Model 3</td>
<td>0.6057</td>
</tr>
</tbody>
</table>

After the models were created, they were compared using ANOVA and AIC tests following the procedure outlined in Methods under Logistic Regression. The results of the ANOVA test can be seen in Table 1, which outlines the resulting p-values from the tests. A p-value under a value of 0.05 shows that subtracting a parameter results in a stronger model. Subsequently, the results of the AIC test are found in Table 2. These results are read as the model with the smaller AIC score has the strongest predictive power. From these results, Model 4 and Model 2 were the models that went forward in testing.

Table 2: Displays the AIC score results from the AIC test that was run during model testing.

<table>
<thead>
<tr>
<th></th>
<th>AIC score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>127.5388</td>
</tr>
<tr>
<td>Model 2</td>
<td>109.2890</td>
</tr>
<tr>
<td>Model 3</td>
<td>129.2723</td>
</tr>
<tr>
<td>Model 4</td>
<td>120.7253</td>
</tr>
</tbody>
</table>

3. Predict historical movement of beech bark disease based on climate factors using the developed model

To further test the models, they were used to predict known values from the data. As seen in Table 3 and Figure 6, Model 2 and 4 were tested over 4 groupings of historical data. When viewing the accuracy of Model 2, it was realized that the variable number of beech scale in the surrounding area would have to be removed. While Model 2 showed an overall higher accuracy, the points it got wrong were only based on the number of beech scale that were surrounding it. Thus, Model 2 showed if a point had any beech scale in the surrounding area, the point was always >=90% more likely to get beech bark disease.
While this would make sense in most cases, as beech scale is a precursor to BBD, the model was removed as no other variables were influencing the outcome.

Model 4 showed promise, as it predicted the historical movement of BBD. As seen in Table 3, the accuracy of predicting all test sites was 65%. This accuracy was determined based on all points between 1999 and 2019, with each point needing to get a confidence score of 50% or above to be considered infected. The accuracy of the model did not handle a small set of points well, such as the year 2019. However, even in that case, it was still 50% accurate. Based on all the testing, Model 4 was chosen as the best model to move forward to future predictions.

Table 3: Displays the results of the model predicting historical locations of BBD. These results show that with smaller data sets the model has less confidence.

<table>
<thead>
<tr>
<th>Year(s)</th>
<th># of points</th>
<th>Model 4 Accuracy</th>
<th>Model 2 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014/2017</td>
<td>8</td>
<td>50%</td>
<td>66%</td>
</tr>
<tr>
<td>2010</td>
<td>18</td>
<td>72%</td>
<td>55%</td>
</tr>
<tr>
<td>2019</td>
<td>4</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>1999-2019</td>
<td>91</td>
<td>65%</td>
<td>78%</td>
</tr>
</tbody>
</table>
4. Apply different climate scenarios based on IPCC climate projections to the model to predict future climate change effects on beech bark disease movement

To predict future climate change effects, six scenarios were created based on IPCC climate projections. These scenarios can be seen in Table 4, with Scenario 1 having the highest likelihood of occurring according to the IPCC. The average value referred to in Table 4 was generated based on the steps outlined in Methods. The numbers used as increase and decrease per year were calculated from Canada’s Changing Climate Report, which is based on IPCC8.5. This report predicts temperature in Ontario will increase 2.3°C over 19 years, giving a 0.12°C increase per year. Similarly, it predicts precipitation will increase 6.6% over 19 years, resulting in 0.35% per year.

Table 4: Outlines the climate scenarios that were run through the model to predict future movement of BBD.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Temperature effect (°C)</th>
<th>Precipitation effect (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Increase temperature and increase precipitation</td>
<td>+0.12 per year</td>
<td>+0.35 per year</td>
</tr>
<tr>
<td>2 Increase temperature and unchanged precipitation</td>
<td>+0.12 per year</td>
<td>Average</td>
</tr>
<tr>
<td>3 Unchanged temperature and increase precipitation</td>
<td>Average</td>
<td>+0.35 per year</td>
</tr>
<tr>
<td>4 Unchanged temperature and unchanged precipitation</td>
<td>Average</td>
<td>Average</td>
</tr>
<tr>
<td>5 Decrease temperature and unchanged precipitation</td>
<td>-0.12 per year</td>
<td>Average</td>
</tr>
<tr>
<td>6 Unchanged temperature and decrease precipitation</td>
<td>Average</td>
<td>-0.35 per year</td>
</tr>
</tbody>
</table>

After the scenarios were generated, 50 new points of potential BBD were selected based on the process outlined in Methods. Average temperatures, average precipitation, and a number of points in the surrounding area with BBD were then gathered for each point. To run these points for each scenario, the temperature and precipitation were modified based on the values in Table 4, then all points were run through the model using Equation 1, based on the values in Table 5.
Equation 1: Prediction = $e^{(β_1 + β_2a + β_3b + β_4c)}/1 + e^{(β_1 + β_2a + β_3b + β_4c)}$

Where $a$ is the precipitation, $b$ is the temperature, $c$ is the number of BBD in the surrounding area, and the $β$ values are from Table 5, Model 4.

Table 5: Displays the estimators created by the logistic regression models, used in Equation 1.

<table>
<thead>
<tr>
<th></th>
<th>Intercept ($β_1$)</th>
<th>Precipitation ($β_2$)</th>
<th>Temperature ($β_3$)</th>
<th># Surrounding BBD ($β_4$)</th>
<th># Surrounding Beech Scale ($β_5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2</td>
<td>-0.45026</td>
<td>-0.02941</td>
<td>0.13768</td>
<td>1.30005</td>
<td>-1.11075</td>
</tr>
<tr>
<td>Model 4</td>
<td>-0.32997</td>
<td>-0.03831</td>
<td>0.26865</td>
<td>0.52407</td>
<td></td>
</tr>
</tbody>
</table>

A point was deemed to have BBD that year if the prediction was over 50%. Points that got BBD were excluded from the next year, and the remaining points had their temperatures and precipitations changed as required. Each scenario was run over a time span of 10 years. These results are displayed in Figure 7.

Temperature appears to have a positive impact on infection, while precipitation appears to have no impact, as more locations contract BBD in the scenarios with a higher temperature, and scenarios with a changed precipitation the same number of locations contract BBD as an average precipitation. However, neither temperature nor precipitation have a large enough impact to actively change the rate of spread for BBD. The results, as displayed in Figure 7, show that only extreme high temperatures from Scenario 2 cause more trees to be infected than under normal or low conditions.
Figure 7: Figure 6: A: Displaying the results of Scenario 1. Where 10 points get BBD in 2020, 2 points get BBD in 2021, and 1 point gets BBD in 2025. B: Displaying the results of Scenario 2. Where 10 points get BBD in 2020, 2 points get BBD in 2021, 1 point gets 1 point gets BBD in 2025, and 1 addition point gets BBD in 2029. C: Displaying the results of Scenarios 3-6. Where 10 points get BBD in 2020, and 2 points get BBD in 2021.

5. Evaluate the strengths and weaknesses of the model
Through the above testing, Model 4 was evaluated against Model 1-3 to determine that it was the strongest predictor. The model is based upon known locations of BBD, allowing the model to have more accuracy concerning where the disease can spread. However, the temperature and precipitation data are limited based on the number of weather
stations in Ontario. As well, the model may have a bias in the logistic regression, as the training data is spatially correlated, leading to oversimplification. The model was the best based on the ANOVA tests, however, the model was still only 65% accurate when predicting historical BBD movement. This leads to the question of what can be improved.

Conclusions

The results of the study show that, under even the most extreme climate scenarios, precipitation does not have an impact on BBD. Additionally, the results show that temperature does have a positive impact on the spread of the disease. This shows that if climate change continues in the trajectory that it is on, BBD should spread at an accelerated pace in Ontario. However, with the model only testing 65% accurate, it is possible that temperature and precipitation have a larger impact on the spread of BBD than shown. The next steps for this study would be to gather more known data points to strengthen the model that is used to predict future points. This could also be done over more years to see how extreme weather from climate change could affect BBD in 20-30 years. Another angle that could be pursued in the future would be change in land use. If diseased beech trees were cut down and moved, it is likely to cause the disease to spread farther than the 10km buffer that this study used. It is likely that, because of the small amount of data and historical infected points, the factors affecting the spread of BBD are under or overplayed. More analysis will need to be done in the future to validate the results found in this study. If the results are validated, it could help explain how our forest ecosystems might change due to climate change. Additionally, it could help explain what types of weather and climate in which diseases may thrive or fail.

Acknowledgments

A special thanks to the OMNRF for the use of their datasets, which made this study possible. Additional thanks go to Ben DeVries and Zohreh Aljani for their expertise and guidance throughout the study.
References

Beaudoin A, Bernier PY, Villemaire P, Guindon L, Guo XJ. 2017. Species composition, forest properties and land cover types across Canada’s forests at 250m resolution for 2001 and 2011. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Quebec, Canada. https://doi.org/10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990


Lawrimore, Jay H.; Ray, Ron; Applequist, Scott; Korzeniewski, Bryant; Menne, Matthew J. (2016): Global Summary of the Year (GSOY), Version 1. [indicate subset used]. NOAA National Centers for Environmental Information. https://doi.org/10.7289/JWPF-Y430


