## Introduction:

Nature-based climate solutions (engagement with natural systems to address climatic challenges (Seddon *et al.*, 2020)) are of rapidly growing interest to help confront the perilous levels of carbon in our atmosphere (Masson-Delmotte *et al.*, 2021). The boreal forest, one of the largest biomes on Earth, has vast potential to act as a carbon sink (Pan *et al.*, 2011). Unfortunately, the boreal forest can also become a carbon source under anthropogenic influence (Leroux *et al.*, 2020, 2021). This is, in part, a result of past management of boreal forests which has modified the distribution and intensity of key disturbances such as spruce budworm (SBW) outbreaks, and ungulate herbivory. SBW outbreaks are now more widespread and severe (Morin *et al.*, 2021), and moose populations in eastern balsam fir dominant forests have become some of the densest in the world. It is now understood that these disturbances can have interacting, negative implications, decreasing the amount of carbon stored by boreal forests (Bergeron and Leduc, 1998; Dymond *et al.*, 2010; Leroux *et al.*, 2020). Fortunately, the management of these disturbances has potential to mitigate such negative impacts and, therefore, act as a nature-based climate solution.

The current management strategy for controlling SBW in Newfoundland relies on a biological pesticide, *Bacillus thuringiensis* (var. kurstaki). Spraying is based off observed SBW populations. However, this has been shown not to be economical (Fuentealba *et al.*, 2019), and efforts have been made to predict SBW defoliation based on environmental characteristics (Cooke and Regniere, n.d.). However, there are conflicting outcomes of these predictions between locations and environmental variables (Candau and Fleming, 2005; Li *et al.*, 2020). Therefore, local predictions for informing local SBW control are important.

We will describe the environmental conditions, and tree stand type, related to SBW defoliation across Newfoundland and, therefore, help inform best management practices (i.e. where to focus insecticide spraying). To identify the conditions related to SBW defoliation, we propose modelling the relationship of past SBW defoliation presence to environmental conditions and tree stand type. I will use generalized linear models fit to SBW defoliation presence across gradients of environmental conditions and between categories of tree stands types. SBW defoliation, environmental, and tree stand data was obtained from existing National and Provincial databases.

**Significance:** This project will contribute to a better understanding of the relationship between SBW and its environment. Furthermore, we will be able to establish employable recommendations to maximize efficacy of spraying for SBW.

## Methods:

### Data sources and type:

Data was originally sourced and formatted by Bo Zhang *et al.* 2021 (in review). This is an overview of these sources and how the data was subsequently converted to the data that was analysed.

#### Defoliation:

Data on defoliation across Newfoundland, from between 1972 and 1981, came from aerial surveys done by Canadian Forest Service (Sterner and Davidson, 1980). The original format was sketch-mapped polygons of defoliated areas which were each converted to a point (fishnet point layer, 2km x 2km).

#### Climate:

Climate data (MaxT0506:Sg\_gdd\_p2; Table 1) was obtained from Natural Resource Canada (McKenney *et al.*, 2011) and were exported as 2 km resolution rasters, for each month between 1972 – 1981. Each variable (MaxT0506:Sg\_gdd\_p2) shown in Table 1 was calculated and extracted as another 2 km resolution raster.

Table 1. Explanatory variable codes and descriptions.

|  |  |
| --- | --- |
| Variable code | Variable |
| MaxT0506 | Average maximum temperature (˚c) for two consecutive months (May/June, June/July, July/August) across years |
| MaxT0607 |
| MaxT0708 |
| MinT0506 | Average Minimum temperature (˚c) for two consecutive months (May/June, June/July, July/August) across years |
| MinT0607 |
| MinT0708 |
| Prcp0506 | Average Minimum precipitation (mm) for two consecutive months (May/June, June/July, July/August) across years |
| Prcp0607 |
| Prcp0708 |
| Sg\_date | “Julian date of the start of growing season and two variables of cumulative degree days above base temperature (5 °C) for two time periods: a) three months prior to the start of growing season, and b) first six weeks after the starting date of the growing season” -Bo |
| Sg\_gdd\_p1 |
| Sg\_gdd\_p2 |
|  |  |
| SpRasterNA | Dominant forest species polygons classified into 4 categories (BS – black spruce, 1; BF – balsam fir, 2; MW – mixed wood, 3; OT – other, 5;) |
| DistRi  DistRo | Distance (km) to nearest corridor (nearest river, nearest road) |

#### Forest stand type:

Information on dominant forest species (SpRasterNA; Table 1) was sourced from the Newfoundland Forest Inventory, Dept of Fisheries, Forestry, and Agriculture of Newfoundland and Labrador (Department of Fisheries Forestry and Agriculture of Newfoundalnd and Labrador, 2022). The data was originally polygons classified into 4 categories (BS – black spruce, 1; BF – balsam fir, 2; MW – mixed wood, 3; OT – other, 5;). Fishnet points were overlayed at 200 m x 200 m resolution and each point assigned tree category. This was converted to a 2 km resolution raster with each cell being classified as the most common tree type within the cell.

#### Distance to road and river:

The distance to corridors (i.e. rivers and roads; DistRi, DistRo; Table 1) was obtained as a 2 km resolution raster from CanVec Series - Hydrographic Features, Natural Resources Canada and National Road Network GeoBase Series from Stats Canada respectively.

### Data extraction: in QGIS

Environmental data was extracted in QGIS for relevant points. Raster (defoliation, climate, forest stand type, distance to corridors) were uploaded as separate layers. We also created a set of 3000 random points bound within a polygon of mainland Newfoundland. The value of each layer (i.e. value of each environmental variable), and the latitude and longitude, was extracted for each defoliation point and each random point into an excel sheet.

### Data analysis: in RStudio

Our data analysis consisted of checking for correlation between variables, fitting generalized linear models, checking assumptions of generalized linear models, fitting generalized additive models, checking the assumptions of generalized additive models, and comparing models developed *a priori*. All analysis was done in RStudio (R Core Team, 2020), with packages tidyverse, readxl, terra, stats, dismo, usdm, corrplot, PerformanceAnalytics, oddsratio, DHARMa, MASS, nlme, mgcv, mgcViz, and bibtex (Venables and Ripley, 2002; Wood, 2003, 2004, 2011, 2017; Naimi *et al.*, 2014; Wood *et al.*, 2016; Schratz, 2017; Fasiolo *et al.*, 2018; Wickham and Bryan, 2019; Wickham *et al.*, 2019; Francois, 2020; Peterson and Carl, 2020; Pinheiro *et al.*, 2020; R Core Team, 2020; Hijmans *et al.*, 2021; Wei and Simko, 2021; Hartig, 2022; Hijmans, 2022).

I initially randomly sampled 600 points from each set of points, random points and defoliation points. However, this gave inconsistent results depending on run when checking assumptions (see Results: Check assumptions) and so I increased the number of points to 2400 points from the set of random points and defoliation points, each.

## Results:

### Correlation structure:

Environmental variables can often be highly correlated. This is especially likely in our dataset given some environmental variables (precipitation, maximum temperature, and minimum temperature) were measured in moving windows across months (May – June, June – July, July – August). I assessed correlation between variables using three methods: correlation plot (Figure 1), clustered dendrogram of distance between variables (Figure 2), and variance inflation factor (Table 1).

Chart, bubble chart

Description automatically generated

Figure 1. Correlation between environmental variables

Chart

Description automatically generated

Figure 2. Clustered dendrogram based on distances between environmental variables.

Table 2. Variance Inflation Factor (VIF) for environmental variables. A VIF above 10 is considered problematic.

|  |  |
| --- | --- |
| Variables | VIF |
| X | 15.54 |
| Y | 9.92 |
| sg\_date | 505.56 |
| sg\_gdd\_p1 | 3.81 |
| sg\_gdd\_p2 | 142.58 |
| pcp0506 | 34.29 |
| pcp0607 | 130.53 |
| pcp0708 | 59.2 |
| MaxT0506 | 761.39 |
| MaxT0607 | 1716.92 |
| MaxT0708 | 771.59 |
| MinT0506 | 172.37 |
| MinT0607 | 358.33 |
| MinT0708 | 234.54 |
| DistanceRo | 1.32 |
| DistanceRi | 4.57 |

Based on Figure 1, Figure 2, and Table 1 I decided to remove MaxT0607, MaxT0708, sg\_date, sg\_\_gdd\_s2, minT0607, minT0708. I chose to keep MaxT0506 and pcp0607 despite high VIFs (Table 1), because they are thought to be controlling factors for SBW populations. I then reassessed correlation between the remaining variables using a correlation plot (Figure 3), a dendrogram of distances between variables (Figure 4), and variance inflation factors (Table 2).

Chart, bubble chart

Description automatically generated

Figure 3. Correlation plot of environmental variables after removing highly correlated variables.

Chart, box and whisker chart

Description automatically generated

Figure 4. Clustered dendrogram based on distances between environmental variables after removing highly correlated variables.

Table 3. Variance Inflation Factor (VIF) for environmental variables after removing highly correlated variables. A VIF above 10 is considered problematic.

|  |  |
| --- | --- |
| Variables | VIF |
| X | 8.67 |
| Y | 8.39 |
| sg\_gdd\_p1 | 2.71 |
| pcp0506 | 30.38 |
| pcp0607 | 98.15 |
| pcp0708 | 39.92 |
| MaxT0506 | 5.34 |
| MinT0506 | 2.68 |
| DistanceRo | 1.27 |
| DistanceRi | 3.24 |

There was still high correlation between precipitation variables (Figure 3, Figure 4, Table 2) so I removed the two precipitations not previously thought to be controlling of SBW populations (pcp0506, pcp0708). I then reassessed correlation between the remaining variables using a correlation plot (Figure 5), a dendrogram of distances between variables (Figure 6), and variance inflation factors (Table 3). After that, correlations, distances, and VIFs appeared low (Figure 5, Figure 6, Table 3).

Chart, bubble chart

Description automatically generated

Figure 5. Correlation plot for environmental variables after removing highly correlated variables a second time.

Chart

Description automatically generated

Figure 6. Clustered dendrogram of environmental variables after removing highly correlated variables a second time.

Table 4. Variance Inflation Factor (VIF) for environmental variables after removing highly correlated variables a second time. A VIF over 10 is considered problematic.

|  |  |
| --- | --- |
| Variables | VIF |
| X | 4.89 |
| Y | 6.71 |
| sg\_gdd\_p1 | 2.58 |
| pcp0607 | 7.57 |
| MaxT0506 | 5.28 |
| MinT0506 | 2.63 |
| DistanceRo | 1.26 |
| DistanceRi | 2.57 |

### Generalized linear model (GzLM)

I fit a model containing the variables remaining after correlation analysis and tree species data as fixed effect explanatory variables for the binomial response variable of presence or absence of SBW defoliation *( 1 )*. I insured the explanatory variables were not tested sequentially by rearranging variables in the model and comparing the resulting estimated coefficients.

(Eq. 1)

### Check assumptions for GzLM

I checked whether errors were homogenous, errors were normal, and whether there was spatial autocorrelation. To check these assumptions, I used the package DHARMa (). DHARMa creates scaled residuals using a simulation approach, so that the normal plot and the plot of residuals against fitted values are interpretable in the same manner as GLM residual plots. When using 600 points from each set of points (random and defoliated) the results were inconsistent between runs of my code. I, therefore, increased the number of points to 2400 points from each set. I found the data to be normal (Figure 7), and homogenous (Figure 8). The statistical test for homogeneity was significant (red lines in Figure 8), however, I believe this to be a result of my large number of samples and continued analysis under the assumption of homogeneity. I also checked the residuals against each rank transformed explanatory variable. The same situation arose, where the line appeared straight and homogenous but because of my large sample size the test for homogeneity was significant (Appendix A; Figure.). There was significant spatial autocorrelation (p < 2.2e-16).

Chart, line chart

Description automatically generated

Figure 7. QQ plot of DHARMa simulated residuals for full GzLM.

**Chart, line chart

Description automatically generated**

Figure 8. DHARMa residuals against rank transformed model predictions for full GzLM. Red curves represent deviations from homogeneity.

To look more deeply into the spatial autocorrelation, I plotted the autocorrelation function (ACF) for the residuals of the full generalized linear model across all 4800 points (Figure 9). The ACF showed that there were high levels of similarity between points closer together, starting at an autocorrelation of 0.8 at lag 2 (Figure 9). I also looked at the ACF for each explanatory variable and along x and y coordinates. Several variables followed the same autocorrelation structure as the residuals: pcp0607, MaxT0506, and DistanceRo. The other variables either followed a similar pattern but with much lower similarity (DistanceRi and MinT0506) or had insignificant similarity across points (below the blue dotted line; sg\_gdd\_p1, X and Y; Appendix)

Chart

Description automatically generated

Figure 9. Autocorrelation plot for residuals of full GzLM.

### General additive model (GAM)

Because the generalized linear model had significant spatial autocorrelation I decided to fit a general additive model with a smoothing function across the latitude and longitude *(Eq. 2)*. The smooth terms were significant, mirroring the test for spatial autocorrelation (p < 2e-16). The full model included the same explanatory variables as the full GzLM *(Eq. 1).*

*(Eq. 2)*

### Check assumptions for GAM

I used DHARMa to check the residuals for the full GzLM, and found the residuals to be normal (Figure 10) and homogenous (Figure 11). One quantile for the test of homogeneity had significant deviation, however, this is a marked improvement to the GzLM, and I also assume the test is highly sensitive because of my large sample size. I, therefore, continued my analysis under the assumption of homogeneity.

Chart, line chart

Description automatically generated

Figure 10. QQ plot of DHARMa simulated residuals for full GAM.

Chart, line chart

Description automatically generated

Figure 11. DHARMa residuals against rank transformed model predictions for full GAM. Red curves represent deviations from homogeneity.

I also plotted the autocorrelation function (ACF) for the full GAM. The pattern in autocorrelation between points was the same as for the ACF of GzLM, but similarities started much lower, at 0.65 (Figure 9, Figure 12).

Chart, surface chart

Description automatically generated

Figure 12. Autocorrelation plot for the full GAM.

### Best GAM

I compared a set of a priori GAM *(Eq. 2-7;* Table 4*)* using Akaike’s Criterion (AIC). AIC calculates goodness of fit but penalizes complexity to minimize overfitting. I fit model three, including only tree species, because budworm are thought to preferentially defoliate balsam fir (). Models four and five, including maximum temperature in May and June and precipitation in June and July, were fit because these covariates have been shown to influence SBW populations (). Models four and five were fit with and without tree species as a covariate. Models six and seven are reduced models based on significant factors calculated from the full model, with and without the effect of tree species.

*(Eq. 3)*

*(Eq. 4)*

*(Eq. 5)*

*(Eq. 6)*

*(Eq. 7)*

Table 5. Degrees of freedom and AIC of each a priori GAM.

| Model | Equation | df | AIC |
| --- | --- | --- | --- |
| 2 (Full) | *Eq. 2* | 38.014 | 4843.057 |
| 3 | *Eq. 3* | 32.535 | 5166.595 |
| 4 | *Eq. 4* | 30.349 | 4936.426 |
| 5 | *Eq. 5* | 34.347 | 4925.907 |
| 6 | *Eq. 6* | 32.293 | 4843.238 |
| 7 | *Eq. 7* | 36.256 | 4839.866 |

The best fit model was a reduced model without tree species (model 6; Table 4). While model 6 did not have the lowest AIC, it was only ~ 4 higher than the lowest AIC of model 7, and model 6 was more parsimonious than model 7. Therefore, I will consider model 6 in my conclusions, which contained pcp0607, MaxT0506, DistanceRo, and DistanceRi as explanatory variables (Table 5).

Table 6. Summary of GAM *(Eq. 6)* Parameter coefficients and significance.

| Parameter | Estimate | Std. Error | z-value | p-value |
| --- | --- | --- | --- | --- |
| pcp0607 | 9.598e-02 | 1.116e-02 | 8.601 | < 2e-16 \*\*\* |
| MaxT0506 | -7.080e-01 | 6.796e-02 | -10.419 | < 2e-16 \*\*\* |
| DistanceRo | 6.665e-05 | 9.246e-06 | 7.208 | 5.67e-13 \*\*\* |
| DistanceRi | 3.918e-05 | 5.320e-06 | 7.365 | 1.77e-13 \*\*\* |

## Discussion:

Temp and precipitation as expected – line up with biological reasoning

Corridors interesting that important – why?

Tree species not – if analyzed on own only sp2 different

Must mean that variation bw species can be accounted for in environmental variable

## References:

**Bergeron, Y. and A. Leduc**. **1998**. Relationships between change in fire frequency and mortality due to spruce budworm outbreak in the southeastern Canadian boreal forest. *Journal of Vegetation Science* **9**: 492–500.

**Candau, J.-N. and R. Fleming**. **2005**. Landscape-scale spatial distribution of spruce budworm defoliation in relation to bioclimatic conditions. *Canadian Journal of Forest Research-revue Canadienne De Recherche Forestiere - CAN J FOREST RES* **35**: 2218–2232.

**Cooke, B. J. and J. Regniere**. **n.d.** Insect Defoliators as Periodic Disturbances in Northern Forest Ecosystems. 39.

**Department of Fisheries Forestry and Agriculture of Newfoundalnd and Labrador**. **2022**. Forest inventory program.

**Dymond, C. C., E. T. Neilson, G. Stinson, K. Porter, D. A. MacLean, D. R. Gray, M. Campagna and W. A. Kurz**. **2010**. Future spruce budworm outbreak may create a carbon source in eastern Canadian forests. *Ecosystems* **13**: 917–931.

**Fuentealba, A., A. Dupont, C. Hébert, R. Berthiaume, R. Quezada-García and É. Bauce**. **2019**. Comparing the efficacy of various aerial spraying scenarios using Bacillus thuringiensis to protect trees from spruce budworm defoliation. *Forest Ecology and Management* **432**: 1013–1021.

**Leroux, S. J., L. Charron, L. Hermanutz and J. Feltham**. **2021**. Cumulative effects of spruce budworm and moose herbivory on boreal forest ecosystems. *Functional Ecology* **35**: 1448–1459.

**Leroux, S. J., Y. F. Wiersma and E. Vander Wal**. **2020**. Herbivore impacts on carbon cycling in boreal forests. *Trends in Ecology & Evolution* **35**: 1001–1010.

**Li, M., D. A. MacLean, C. R. Hennigar and J. Ogilvie**. **2020**. Previous year outbreak conditions and spring climate predict spruce budworm population changes in the following year. *Forest Ecology and Management* **458**: 117737.

**Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, L. Gomis, *et al.*** **2021**. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

**McKenney, D. W., M. F. Hutchinson, P. Papadopol, K. Lawrence, J. Pedlar, K. Campbell, E. Milewska, R. F. Hopkinson, D. Price and T. Owen**. **2011**. Customized Spatial Climate Models for North America. *Bulletin of the American Meteorological Society* **92**: 1611–1622. American Meteorological Society.

**Morin, H., R. Gagnon, A. Lemay and L. Navarro**. **2021**. Chapter Thirteen - Revisiting the relationship between spruce budworm outbreaks and forest dynamics over the Holocene in Eastern North America based on novel proxies. In: *Plant Disturbance Ecology (Second Edition)* (E. A. Johnson and K. Miyanishi, eds), pp. 463–487. Academic Press, San Diego.

**Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, *et al.*** **2011**. A large and persistent carbon sink in the world’s forests. *Science. 333: 988-993.* **333**: 988–993.

**R Core Team**. **2020**. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

**Seddon, N., A. Chausson, P. Berry, C. A. J. Girardin, A. Smith and B. Turner**. **2020**. Understanding the value and limits of nature-based solutions to climate change and other global challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190120. Royal Society.

**Sterner, T. E. and A. G. Davidson**. **1980**. Forest insect and disease conditions in Canada. *Catal*. Fo21-1/1980E, Canadian Forestry Service.