SPATIAL PATTERNS AND FACTORS INFLUENCING SPRUCE BUDWORM INFESTATION IN EASTERN CANADA FORESTS

by

Mingke Li

Bachelor of Geographic Information Science, Nanjing Forestry University, 2016

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science in Forestry

in the Graduate Academic Unit of Forestry and Environmental Management

Supervisor:	David A. MacLean, PhD, FOREM
Examining Board:	Rob Johns, PhD, CFS Kara Costanza, PhD, FOREM

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

August 2019

ABSTRACT

A spruce budworm (*Choristoneura fumiferana* Clem.; SBW) outbreak in Québec spread southward into New Brunswick in 2014. This thesis used spatial analyses of 5 years of SBW population data in northern New Brunswick and tree defoliation data in Québec to examine spatial patterns and factors influencing SBW infestation. Local previous-year SBW population level, proximity to outbreak hot-spots, and April degree-days were important in predicting SBW population levels in New Brunswick, although relationships were inconsistent across years. Models incorporating spatial stuctures explained 68–79% of the annual variance, and performed better than non-spatial models. A combined-year model with $R^2 = 0.53$ consistently underestimated upcoming-year populations. Defoliation patterns quantified in 57 plots in Québec were clustered in 28-47% of cases, which had higher plot-level defoliation and higher deviations. Plot-level defoliation and basal area explained 80% of the variance in individual-tree-defoliation. The thesis contributed to efficient sampling allocation and insecticide treatment targeting infestation.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. David MacLean, for helping to formulate the overarching research goals, guiding the research process, and reviewing multiple drafts throughout the process of this thesis. I am grateful to other members of my advisory committee, Dr. Chris Hennigar and Mr. Jae Ogilvie, who provided helpful suggestions and feedback during the research. Thanks also to Dr. John Kershaw who offered his expertise on statistical modeling. I wish to thankfully acknowledge contributions from the Healthy Forest Partnership Spruce Budworm Early Intervention Strategy Research Group, New Brunswick Department of Energy and Resource Development, Gouvernement du Québec, Ministère des Forêts, de la Faune et des Parcs, La société de Protection des Forêts contre les Insectes et Maladies, and the Forest Watershed Research Centre at the University of New Brunswick for providing the insect population data, insecticide spray treatment data, and topographic data. I appreciate the excellent work of all involved in the defoliation data collection in the field plots used in Chapter 3: Shawn Donovan, Sean Lamb, Bo Zhang, Rebecca Landry, David Alton, Kerstin Trainor, and those who assisted with the field work before I joined the lab. Special thanks to all of those have supported, accompanied, and encouraged me during my study-abroad journey: my parents, my fellow students, Ken Yeamans, and Barbara Prince.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: GENERAL INTRODUCTION

1.1. Problem Statement

The spruce budworm (*Choristoneura fumiferana* Clem.; SBW) is one of the most destructive insects in North America's conifer forests. SBW outbreaks typically occur every 35 years, and are historically the most important driver of forest structure development in eastern Canadian forests (Baskerville 1975; Royama 1984). Effective management of SBW outbreaks should account for the dynamics of stands and insect populations, especially during an initial stages of an outbreak. Knowledge of factors influencing SBW epidemics, expanding tendency of SBW outbreaks, and prediction of high-level SBW populations is important to effectively manage SBW.

After SBW outbreaks are well established, Baskerville and MacLean (1979) observed that SBW-caused tree mortality in immature balsam fir (*Abies balsamea* L. Mill.) tended to have a strongly contagious spatial distribution. Given that tree mortality is strongly related to cumulative defoliation (Erdle and MacLean 1999; Chen et al. 2017), a clustered spatial distribution of dead trees may reflect higher defoliation in those trees than in surrounding trees. However, no previous studies have determined spatial patterns of SBW defoliation among trees. Although the spatial variability at multiple scales has been studied, shoot, branch, tree, plot, stand, and landscape included, the spatial inter-relationships of SBW defoliation levels of trees within a plot is not well understood (MacLean and Lidstone 1982; Hardy et al. 1983; Gray and MacKinnon 2006). Since relationships between tree mortality and defoliation are non-linear, effects of defoliation on stand growth reduction and mortality will be underestimated or overestimated if predictions are made by average defoliation per plot or stand, especially when some of

the trees suffer substantially higher or lower levels of defoliation. Such knowledge of spatial defoliation patterns contributes to projecting stand growth reduction or tree mortality in distance-dependent or tree-list growth models (e.g., Lamb et al. 2018).

1.2. Background

1.2.1. Spruce budworm overview

Eastern spruce budworm is native to eastern North America and has been found throughout the range of its primary host trees, balsam fir and spruce (*Picea* spp. A. Dietr.), for thousands of years (Balch 1958). After emerging from eggs in August, the first instar larvae immediately disperse within trees or stands by wind (Nealis 2016). At suitable sites, they spin silken shelters in twig and bark crevices, molt to the second instar larvae (L2), and overwinter in hibernacula until the following spring (Nealis 2016). Emerging from hibernacula, SBW larvae start feeding on new needles and buds (MacLean 1984). These L2 SBW larvae cause serious defoliation among stands and forests when they occur in high population levels (Blais 1985).

Defoliation resulting from SBW feeding causes two main consequences, tree growth loss and mortality (MacLean 1984). Defoliation and damage are more severe near the tops of host trees (Mott et al. 1957; Virgin et al. 2018). Defoliation after several consecutive years can lead to decreased radial growth (Alfaro et al. 1982), reduction or cessation of height growth (Baskerville and MacLean 1979), and reduced tree volume (Piene 1980). Stands of balsam fir start to die after 3 to 5 years of severe defoliation, while spruce stands can withstand 5 to 7 years of severe defoliation before mortality begins (Blais 1985; Ostaff and MacLean 1989). Defoliation level varies as a function of

the SBW population (Erdle and MacLean 1999; Chen et al. 2017), host species (Hennigar et al. 2008), tree age classes (MacLean 1980), and site conditions (MacLean and MacKinnon 1997), and this determines the amount of tree mortality.

Outbreaks of SBW recur periodically on 30-65 year return intervals (e.g., Royama 1984, Jardon et al. 2003, Boulanger and Arsenault 2004, Fraver et al. 2007, Boulanger et al. 2012). A total of 21 regional outbreaks were recorded in the 20th century in eastern Canada (Blais 1983) as parts of the three major 1910s, 1950s, and 1970s-1980s outbreaks, in comparison with nine regional outbreaks noted via dendrochronology in the same regions the 19th century. However, what initiates SBW outbreaks remain unclear, although exogenous climate forces, acting as the Moran effect, can act to synchronize oscillating populations over large areas (Myers 1998). Damage to forests caused by SBW and the resulting economic losses has been estimated. For example, from 1967 to 1981, 23 million ha of forest were defoliated by SBW in eastern North America, and the outbreak was still ongoing (Kettela 1983). From 1977 to 1981, 44 million m^3 per year of tree volume was lost in Canada due to SBW outbreaks (Sterner and Davidson 1982). In addition to these direct impacts, SBW outbreaks also influence wildlife and other values of forest landscapes.

1.2.2. Theories of spatial synchrony and outbreaks of forest insects

The spatial synchrony of forest insect populations refers to geographically disjunctive populations tending to be temporally correlated (Liebhold et al. 2004). Large-scale synchronous increase in insect population density during one year and persistence over subsequent years can lead to rising insect populations and an outbreak initiation

(Liebhold et al. 2012; Bouchard et al. 2018). However, reasons for spatial synchrony are complicated because different ecological processes can contribute to similar dynamic patterns, and multiple biotic or abiotic forces can act simultaneously to cause an insect outbreak (Nelson et al. 2004; Liebhold et al. 2012). Understanding the causal mechanisms behind large-scale synchrony of forest insect populations has important implications for forecasting of outbreak onset and how insect populations can best be managed. Three main causes of spatial synchrony have been proposed: regional stochasticity, dispersal, and trophic interactions (Bjørnstad et al. 1999; Liebhold et al. 2004).

First, Moran (1953) proposed that spatially correlated forces (e.g., climate perturbations), as density-independent factors, were able to turn local population fluctuations into synchrony (i.e., "Moran effect"). An investigation of gypsy moth (*Lymantria dispar* L.) outbreaks demonstrated that spatially correlated weather conditions can lead to synchronous local populations (Williams and Liebhold 1995). 140 outbreaks of 26 forest insect species from 1932 to 1992 were reviewed by Myers (1998), demonstrating the reliability of the theory of Moran effect. In the case of SBW, the Moran effect combined with a high rate of adult dispersal can lead to spatially synchronized SBW outbreaks (Williams and Liebhold 2000). Second, dispersal was proposed as a factor of synchronous dynamics across large regions by previous literature (Liebhold et al. 2012; Myers and Cory 2013). Density-independent exchanges of even a small proportion of spatially-disjunct populations can result in synchrony according to mathematical models (Barbour 1990; Bjørnstad and Bolker 2000; Liebhold et al. 2004). However, some studies argued that there were weak relationships between the dispersal

capability of an insect species and the distance over which insect populations would show synchrony (Peltonen et al. 2002). Third, trophic interaction, including bottom-up and top-down trophic factors, is another force acting on spatial synchrony of insect populations. Host tree species undoubtedly have an impact on synchronous dynamics of insects: positive relationships between host plants and insect population dynamics have been demonstrated for gypsy moth (Bjørnstad et al. 2010), forest tent caterpillar (*Malacosoma disstria,* Cooke et al. 2011), autumnal moth (*Epirrita autumnata,* Klemola et al. 2006), and SBW (Bouchard et al. 2018). Nomadic predators, parasitoids, and pathogens can potentially influence insect populations (Liebhold et al. 2012).

For SBW, how an outbreak starts has been debated by two concurrent hypotheses: the 'oscillatory hypothesis' and the 'double-equilibrium hypothesis' (Pureswaran et al. 2016). Under the 'oscillatory hypothesis', the basic population oscillation is mainly influenced by parasitoids and diseases, as well as predator-prey dynamics (Royama 1984). An outbreak starts when the prey density is at its lowest, following which a rapid growth of SBW populations would occur. In contrast, the 'double-equilibrium hypothesis' argued that SBW populations are maintained at a lower equilibrium density by heavy mortality until the mortality rates decline due to forest maturation, highly favorable climates, or moths' immigration, and then insect populations reach the upper equilibrium density (Morris 1963; Ludwig et al. 1978).

Throughout all kinds of hypotheses explaining mechanisms underlying forest insect population dynamics, climate was frequently determined to directly or indirectly influence insect population via food availability, vulnerability to disease, migration activities, and trophic relationships between host plants and natural enemies (Royama

1984; Régnière and Lysyk 1995; Myers and Cory 2013; Bouchard et al. 2018). At first, the Moran effect was believed to be a direct result of climate influence (Williams and Liebhold 1995). However, scientists noted that weather, as an exogenous effect, can indirectly synchronize insect populations by affecting the resource synchrony (Haynes et al. 2009). In fact, relationships between climate factors and insect outbreaks are undoubtedly complicated, and studies of such relationships are not able to achieve an agreement (Gray 2008). In the climate change context, relationships between climate and insect population synchrony are more complicated than what was proposed by Moran (1953), not only because of the spatial heterogeneities and nonlinearities, but also because of the trophic chains affected both directly and indirectly by spatially correlated weather conditions (Bouchard et al. 2018).

1.2.3. Spatial analyses in the ecological context

The developing applications of spatial analysis in the field of ecology in the last few decades derived from growing awareness of spatial structures in ecological thinking, constant revisiting of landscapes' spatial heterogeneity, and the availability of platforms to perform spatial analyses (Dale and Fortin 2014). As one of the core concepts in spatial analysis, spatial autocorrelation has attracted attention from ecologists. Spatial autocorrelation shows the correlation among variables across georeferenced space, defined as the observed value of the variable at one locality being dependent on the values at neighboring localities (Sokal and Oden 1978). Sources of spatial autocorrelation in ecological data can be causal interactions within the observed variable itself (i.e., inherent autocorrelation), or the observed variable is functionally dependent on an underlying

variable which is autocorrelated (i.e., induced autocorrelation; Dale and Fortin 2014). For instance, the spatial-autocorrelated population of one species can be caused by its mobility and dispersal, or caused by spatial-autocorrelated moisture or temperature which have impacts on the species population (Dale and Fortin 2014). Traditionally, spatial dispersion patterns in ecology were studied by the variance-mean-based relationship between samples which omitted the spatial location of the studied samples. However, spatial autocorrelation determines the degree of dependence of spatially-related data, offering a more efficient measure of spatial dependence than the traditional approaches (Papadopoulos et al. 2003).

Analyses of spatial ecological patterns include two families of methods: point pattern analysis and surface pattern analysis (Legendre 1993). Point pattern analysis aims to determine the distribution of individual objects through space, i.e., whether the location of studied data points is random, clustered, or dispersed (Ebdon 1985). It helps to infer what ecological process may have created the observed spatial structure. The widely used point pattern statistics include Clark and Evans (1954), Pielou (1959), and Ripley (1979). On the other hand, surface pattern analysis aims to determine the distribution of a spatially continuous phenomenon, i.e., whether the measured variable at one site is dependent on values at neighboring sites (Sokal and Oden 1978). Under the surface pattern analysis family, global spatial statistics estimate the intensity of spatial dependence for the entire study area and summarize it with a single value, while local spatial statistics allow for the decomposition of global or general statistics into the contribution by each individual observation. Commonly used global statistics include global Moran's *I* (1950), global Geary's *c* (1954), and Getis-Ord General *G* (1992) ; local statistics include local Moran's *I*, local Geary's *c*, and Getis-Ord *G** (Anselin 1995; Ord and Getis 1995).

Spatial analysis techniques have been applied efficiently to plant and insect related research, contributing to inferences of selection, migration, drift, and isolation by distance (Reed and Burkhart 1985; Czaplewski and Reich 1994; Papadopoulos et al. 2003; Ryan et al. 2004; Boiteau 2005). Knowledge of such spatial patterns can provide insights about the underlying ecological processes that generate them. Understanding the degree of spatial autocorrelation and spatial scale of ecological variables is a precondition to relate the ecological response to the covariates (Dale and Fortin 2014). Determining spatial patterns also has important application in current forest management. For example, strategies of number and location of insect sampling, as well as frequency of insect trapping, were traditionally based on logistics, but it will be more efficient and reasonable to base these on bionomics and expected spatial distribution of the target insects (Ryan et al. 2004).

1.2.4. Spatial regression with ecological data

Spatial autocorrelation is one of the natures of ecological data. When modeling ecological data, ignoring spatial autocorrelation may result in statistical issues, because the assumption of independence of the errors is often violated (Dormann et al. 2007). Such issues will not occur if the intrinsic causes of spatial autocorrelation in the response do not exist, and the extrinsic causes of spatial structures in the response can be fully explained by the spatial structures in the predictors (Beale et al. 2010). However, these conditions are rarely met concurrently in actual studies. Hence, residuals are expected to have spatial dependence if spatial autocorrelation is not dealt with in the modeling process, and consequently, Type I errors will increase and the estimation of regression parameters will be unpredictably biased (Dormann et al. 2007; Bini et al. 2009).

Spatial regressions are extensions of the traditional statistical technique with modifications required to take account of spatial dependence in the data (Dale and Fortin 2014). Additionally, more and more new spatial models and methodologies are being introduced to ecologists. Spatial models can be classified into three categories by the way spatial effects are accounted for in the models: 1) spatial effects are in an error term, e.g., Generalized Least Squares with a Spherical function for the error matrix, 2) spatial effects are incorporated as covariates, e.g., Spatial Filters and Generalized Additive Models, and 3) transformed response data into spatial independent values, e.g., Wavelet Revised Models (Dormann et al. 2007; Beale et al. 2010). Performance of these spatial models has been estimated and compared in several papers, which gives ecologists evidence of how to select from these models in different study situations (e.g., Dormann et al. 2007; Kissling and Carl 2008; Bini et al. 2009; Beale et al. 2010).

Many studies incorporating spatial autocorrelation have been carried out among different fields, including wildlife habitat, plant species distribution, gene frequency distribution, land use change over time, etc. (e.g., Augustin et al. 1996, 1998; Knapp et al. 2003; Syartinilia and Tsuyuki 2008). Results of these studies showed the stronger forecasting ability and better overall accuracy of spatial models compared to traditional non-spatial models. Spatial models are useful in exploring the possible relationships between spatially referenced responses and covariates, predicting response values at unsurveyed locations, and estimating global characteristics of the object distribution (Augustin et al. 1998).

1.3. Study Description

This project consists of two distinct subprojects, both involving spatial analyses of SBW dynamics at different scales. The first study area was in northern New Brunswick, amounting to 3,730,000 ha, covering about one-half of the province. The study was based on L2 per branch SBW population data from 2013 to 2018 jointly collected by New Brunswick Department of Energy and Resource Development (NBERD) and forest industry staff from Forest Protection Limited, J.D. Irving, Limited, Acadian Timber Corporation, and Fornebu Lumber Company Inc.. During this period, the study area transitioned from very low SBW to the initial stages of a SBW outbreak, with the first defoliation in NB forecast in 2015 by NBERD. The question addressed was how spatial patterns and factors influence the spread of SBW L2 samples over time.

The second study area was in the central Gaspé Peninsula region of Québec, where plots were measured annually by Dr. David MacLean's lab for individual tree current-year defoliation from 2014 to 2018. The spruce budworm outbreak in Québec was expanding continuously according to the annual report of SBW infested area (QMRNF 2018), began in the Gaspé region in 2012, and defoliation reached the study area in 2014 . A portion of the studied plots were protected by aerial spraying of *Bacillus thuringiensis* biological insecticide to control defoliation over the years. The question of interest here was determining within-plot spatial patterns of tree locations and defoliation.

Throughout the two subprojects, geostatistics and spatial analysis were conducted, including spatial pattern analysis, spatial autocorrelation analysis, spatial overlay analysis, and spatial autoregressive modeling. This thesis made an attempt to answer ecological questions by geospatial analysis methods.

1.4. Objectives

The overall objectives of this project were to 1) determine potential effects of forest composition, climatic, topographic, site quality, insecticide treatment, and previous-year outbreak conditions on SBW populations during the outbreak initiation in northern New Brunswick, and 2) evaluate spatial patterns of SBW defoliation of individual trees within plots in Gaspé, Québec.

The specific objectives of the first subproject (Chapter 2) were to:

1) determine factors that influenced SBW L2 populations from 2014 to 2018, the outbreak initiation phase in northern New Brunswick; and

2) test if the L2 population in the subsequent year could be predicted by L2 distribution in the preceding year and environmental factors.

The specific objectives of the second subproject (Chapter 3) were to:

1) determine the spatial distribution of tree stems within studied plots;

2) detect and quantify spatial aggregation of inter-tree and intra-plot defoliation; and

3) test whether the individual tree defoliation could be predicted by tree and plot characteristics.

1.5. Thesis Structure

This thesis is composed of four chapters. Following this introduction, Chapter 2, *Previous year population level, proximity to high population sites, and spring climate predict SBW population changes in the following year*, used spatial overlay and spatial modeling to reveal factors influencing SBW population and to predict subsequent-year budworm population. Chapter 3, *Spatial-temporal patterns of spruce budworm*

defoliation within plots in Québec, investigated the spatial-temporal patterns of SBW defoliation of trees within plots over 5 years during the current SBW outbreak in Québec. Chapter 4 summarizes the main results and conclusions of the project, and discusses the main applications and potential value of the project results in SBW control management. It also points out the possible limitations of the two subprojects, and suggests improvement and follow-up studies.

1.6. References

Alfaro, R.I., Van Sickle, G.A., Thomson, A.J., and Wegwitz, E. 1982. Tree mortality and radial growth losses caused by the weatern spruce budworm in a Douglas-fir stand in British Columbia. Can. J. For. Res. **12**(4): 780-787.

Anselin, L. 1995. Local indicators of spatial association-LISA. Geogr. Anal. **27**(2): 93-115.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. J. Appl. Ecol. **33**(2): 339-347.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. 1998. The role of simulation in modelling spatially correlated data. Environmetrics **9**: 175-196.

Balch, R.E. 1958. The spruce budworm in New Brunswick. Science Service, Forest Biology Division, Dept. of Agriculture, Ottawa, ON, Canada.

Barbour, D.A. 1990. Synchronous fluctuations in spatially separated populations of cyclic forest insects. *In* Population dynamics of forest insects. *Edited by* A.D. Watt, S.R. Leather, M.D. Hunter, and N.A.C. Kidd. Intercept, Andover, Hampshire, UK. pp. 339-346.

Baskerville, G.L. 1975. Spruce budworm: super silviculturist. For. Chron. **51**: 138-140.

Baskerville, G.L. and MacLean, D.A. 1979. Budworm-caused mortality and 20-year recovery in immature balsam fir stands. Can. For. Serv., Fredericton, NB, Canada. Inf. Rep. M-X-102.

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J., and Elston, D.A. 2010. Regression analysis of spatial data. Ecol. Lett. **13**: 246-264.

Bini, L.M., Diniz-filho, J.A.F., Rangel, T.F.L.V.B., Akre, S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araújo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Böhning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Jr, P.D.M., David, S., Ferrer-castán, D., Field, R., Filloy, J., Fleishman, E., Jose, F., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araijo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Marco, P.D., Dobkin, D.S., Ferrer-castain, D., Field, R., Filloy, J., Fleishman, E., Gmez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, J., Len-cortes, J.L., Lobo, J.M., Montoya, D., Morales-castilla, I., Moreno, J.C., Oberdorff, T., Olalla-t, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R., and Hawkins, B.A. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography **32**(2): 193-204.

Bjørnstad, O.N. and Bolker, B. 2000. Canonical functions for dispersal-induced synchrony. Proc. R. Soc. Lond. B. **267**: 1787-1794.

Bjørnstad, O.N., Ims, R.A., and Lambin, X. 1999. Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends Ecol. Evol. **14**(11): 427-432.

Bjørnstad, O.N., Robinet, C., and Liebhold, A.M. 2010. Geographic variation in North American gypsy moth cycles: subharmonics, generalist predators, and spatial coupling. Ecology **91**(1): 106-118.

Blais, J.R. 1983. Trends in the frequency, extent, and severity of spruce budworm outbreaks in eastern Canada. Can. J. For. Res. **13**(4): 539–547.

Blais, J.R. 1985. Impact of the spruce budworm on balsam fir and white spruce in the Laurentian reserve, Québec: an interim report. Canadian Forestry Service, Sainte-Foy, Québec. Inf. Rep. LAU-X-68.

Boiteau, G. 2005. Within-field spatial structure of colorado potato beetle (*Coleoptera: Chrysomelidae*) populations in New Brunswick. Environ. Entomol. **34**(2): 446-456.

Bouchard, M., Régnière, J., and Therrien, P. 2018. Bottom-up factors contribute to large-scale synchrony in spruce budworm populations. Can. J. For. Res. **48**(3): 277-284.

Boulanger, Y. and Arseneault, D. 2004. Spruce budworm outbreaks in eastern Québec over the last 450 years. Can. J. For. Res. **34**(5): 1035-1043.

Boulanger, Y., Arseneault, D., Morin, H., Jardon, Y., Bertrand, P., and Dagneau, C. 2012. Dendrochronological reconstruction of spruce budworm (Choristoneura fumiferana) outbreaks in southern Québec for the last 400 years. Can. J. For. Res. **42**(7): 1264-1276.

Chen, C., Weiskittel, A., Bataineh, M., and MacLean, D.A. 2017. Even low levels of spruce budworm defoliation affect mortality and ingrowth but net growth is more driven by competition. Can. J. For. Res. **47**(11): 1546-1556.

Clark, P.J. and Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology **35**(4): 445-453.

Cooke, B.J., MacQuarrie, C.J.K., and Lorenzetti, F. 2011. The dynamics of forest tent caterpillar outbreaks across east-central Canada. Ecography **34**: 001-014.

Czaplewski, R.L. and Reich, R.M. 1994. Spatial autocorrelation in growth of undisturbed natural pine stands across Georgia. For. Sci. **40**(2): 314-328.

Dale, M.R. and Fortin, M.J. 2014. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Dormann, C.F., Mcpherson, J.M., Arau, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Ohlemu, R., Peres-neto, P.R., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**: 609-628.

Ebdon, D. 1985. Statistics in geography: a practical approach. Basil Blackwell Ltd., New York, NY, US.

Erdle, T.A. and MacLean, D.A. 1999. Stand growth model calibration for use in forest pest impact assessment. For. Chron. **75**: 141-152.

Fraver, S., Seymour, R.S., Speer, J.H., and White, A.S. 2007. Dendrochronological reconstruction of spruce budworm outbreaks in northern Maine, USA. Can. J. For. Res. **37**(3): 523-529.

Geary, R.C. 1954. The contiguity ratio and statistical mapping. The Incorporated Statistician **5**(3): 115-146.

Getis, A. and Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. **24**(3): 189-206.

Gray, D.R. 2008. The relationship between climate and outbreak characteristics of the spruce budworm in eastern Canada. Clim. Change **87**: 361-383.

Gray, D.R. and MacKinnon, W.E. 2006. Outbreak patterns of the spruce budworm and their impacts in Canada. For. Chron. **82**(4): 550-561.

Hardy, Y.J., Lafond, A., and Hamel, L. 1983. The epidemiology of the current spruce budworm outbreak in Québec. For. Sci. **29**(4): 715-725.

Haynes, K.J., Liebhold, A.M., Fearer, T.M., Wang, G., Norman, G.W., and Johnson, D.M. 2009. Spatial synchrony propagates through a forest food web via consumer-resource interactions. Ecology **90**(11): 2974-2983.

Hennigar, C.R., MacLean, D.A., Quiring, D.T., and Kershaw, J.A. 2008. Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. For. Sci. **54**(2): 158-166.

Jardon, Y., Morin, H., and Dutilleul. P. 2003. Periodicity and synchronism of outbreaks of spruce budworm in Québec. Can. J. For. Res. **33**(10): 1947-1961.

Kettela, E.G. 1983. A cartographic history of spruce budworm defoliation 1967 to 1981 in eastern North America. Canadian Forestry Service, Environment Canada., Ottawa, ON, Canada. Inf. Rep. DPC-X-14.

Kissling, W.D. and Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecol. Biogeogr. **17**: 59-71.

Klemola, T., Huitu, O., and Ruohomaki, K. 2006. Geographically partitioned spatial synchrony among cyclic moth populations. Oikos **114**(2): 349-359.

Knapp, R.A., Matthews, K.R., Preisler, H.K., and Jellison, R. 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. Ecol. Appl. **13**: 1069-1082.

Lamb, S.M., MacLean, D.A., Hennigar, C.R., and Pitt, D.G. 2018. Forecasting forest inventory using imputed tree lists for LiDAR grid cells and a tree-list growth model. Forests 9: 167.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology **74**(6): 1659-1673.

Liebhold, A., Koenig, W.D., and Bjørnstad, O.N. 2004. Spatial synchrony in population dynamics. Annu. Rev. Ecol. Evol. Syst. **35**: 467-490.

Liebhold, A.M., Haynes, K.J., and Bjørnstad, O.N. 2012. Spatial synchrony of insect outbreaks. *In* Insect outbreaks revisited. *Edited by* P. Barbosa, D.K. Letourneau, and A.A. Agrawal. Blackwell Publishing Ltd., Chichester, UK. pp. 113-125.

Ludwig, D., Jones, D.D., and Holling, C.S. 1978. Qualitative analysis of insect outbreak systems: the spruce budworm and the forest. J. Anim. Ecol. **47**(1): 315-332.

MacLean, D.A. 1980. Vulnerability of fir-spruce stands during uncontrolled spruce budworm outbreaks: a review and discussion. For. Chron. **56**: 213-221.

MacLean, D.A. 1984. Effects of spruce budworm outbreaks on the productivity and stability of balsam fir forests. For. Chron. **60**: 273-279.

MacLean, D.A. and Lidstone, R.G. 1982. Defoliation by spruce budworm: estimation by ocular and shoot-count methods and variability among branches, trees, and stands. Can. J. For. Res. **12**(3): 582-594.

MacLean, D.A. and MacKinnon, W.E. 1997. Effects of stand and site characteristics on susceptibility and vulnerability of balsam fir and spruce to spruce budworm in New Brunswick. Can. J. For. Res. **27**(11): 1859-1871.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika **37**: 17-23.

Moran, P.A.P. 1953. The statistical analysis of the Canadian Lynx cycle. Aust. J. Zool. **1**(3): 291-298.

Morris, R.F. 1963. The dynamics of epidemic spruce budworm populations. Mem. Entomol. Soc. Can. **95**(S31): 7-12.

Mott, D.G., Nairn, L.D., and Cook, J.A. 1957. Radial growth in forest trees and effects of insect defoliation. For. Sci. **3**: 286-304.

Myers, J.H. 1998. Synchrony in outbreaks of forest Lepidoptera: a possible example of the Moran effect. Ecology **79**(3): 1111-1117.

Myers, J.H. and Cory, J.S. 2013. Population cycles in forest Lepidoptera revisited. Annu. Rev. Ecol. Evol. Syst. **44**: 565-592.

Nealis, V.G. 2016. Comparative ecology of conifer-feeding spruce budworms (Lepidoptera: Tortricidae). Can. Entomol. **148**: S33-S57.

Nelson, W.A., McCauley, E., and Wimbert, J. 2004. Capturing dynamics with the correct rates: inverse problems using semiparametric approaches. Ecology **85**(4): 889-903.

Ord, J.K. and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. **27**(4): 286-305.

Ostaff, D.P. and MacLean, D.A. 1989. Spruce budworm populations, defoliation, and changes in stand condition during an uncontrolled spruce budworm outbreak on Cape Breton Island, Nova Scotia. Can. J. For. Res. **19**(9): 1077-1086.

Papadopoulos, N.T., Katsoyannos, B.I., and Nestel, D. 2003. Spatial autocorrelation analysis of a *Ceratitis capitata* (Diptera: Tephritidae) adult population in a mixed deciduous fruit orchard in northern Greece. Environ. Entomol. **32**(2): 319-326.

Peltonen, M., Liebhold, A.M., Bjørnstad, O.N., and Williams, D.W. 2002. Spatial synchrony in forest insect outbreaks: roles of regional stochasticity and dispersal. Ecology **83**(11): 3120-3129.

Pielou, E.C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. J. Ecol. **47**(3): 607-613.

Piene, H. 1980. Effects of insect defoliation on growth and foliar nutrients of yong balsam fir. For. Sci. **26**(4): 665-673.

Pureswaran, D., Johns, R.C., Heard, S.B., and Quiring, D. 2016. Paradigms in eastern spruce budworm (Lepidoptera: Tortricidae) population ecology: a century of debate. Env. Entomol. **45**(6): 1333-1342.

QMRNF. 2018. Aires Infestées par la Tordeuse des Bourgeons de l'épinette au Québec en 2018. Québec Ministère des Ressources Naturelles et de la Faune. Available from http://www.mffp.gouv.qc.ca/publications/forets/fimaq/insectes/tordeuse/TBE_2018_P.pdf [accessed 2 January 2019].

Reed, D.D. and Burkhart, H.E. 1985. Spatial autocorrelation of individual tree characteristics in loblolly pine stands. For. Sci. **31**(3): 575-587.

Régnière, J. and Lysyk, T. 1995. Population dynamics of the spruce budworm, *Choristoneura fumiferana*. *In* Forest Insect Pests in Canada. *Edited by* J.A. Armstrong and W.G.H. Ives. Canadian Forest Service, Science and Sustainable Development Directorate, Ottawa, ON, Canada. pp. 95-105.

Ripley, B.D. 1979. Tests of "randomness" for spatial point patterns. J. Royal Stat. Soc. **41**(3): 368-374.

Royama, T. 1984. Population dynamics of the spruce budworm *Choristoneura fumiferana*. Ecol. Monogr. **54**(4): 429-462.

Ryan, P.A., Lyons, S.A., Alsemgeest, D., Thomas, P., and Kay, B.H. 2004. Spatial statistical analysis of adult mosquito (Diptera: Culicidae) counts: an example using light trap data, in Redland Shire, southeastern Queensland, Australia. J. Med. Entomol. **41**(6): 1143-1156.

Sokal, R.R. and Oden, N.L. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10**: 199-228.

Sterner, T.E. and Davidson, A.G. 1982. Forest insect and disease conditions in Canada 1981. Environment Canada, Canadian Forestry Service, Headquarters, Ottawa, ON, Canada.

Syartinilia and Tsuyuki, S. 2008. GIS-based modeling of Javan Hawk-Eagle distribution using logistic and autologistic regression models. Biol. Conserv. **141**: 756-769.

Virgin, G.V.J., MacLean, D.A., and Kershaw, J.A. 2018. Topkill and stem defects initiated during an uncontrolled spruce budworm outbreak on Cape Breton Island, Nova Scotia. Forestry **91**: 63-72.

Williams, D.W. and Liebhold, A.M. 1995. Influence of weather on the synchrony of gypsy moth (Lepidoptera: Lymantriidae) outbreaks in New England. Environ. Entomol. **24**(5): 987-995.

Williams, D.W. and Liebhold, A.M. 2000. Spatial synchrony of spruce budworm outbreaks in eastern North America. Ecology **81**(10): 2753-2766.

CHAPTER 2: PREVIOUS YEAR POPULATION LEVEL, PROXIMITY TO HIGH POPULATION SITES, AND SPRING CLIMATE PREDICT SPRUCE BUDWORM POPULATION CHANGES IN THE FOLLOWING YEAR

Planned for submission to Forest Ecology and Management

Authors: Mingke Li, David MacLean, Chris Hennigar, and Jae Ogilvie

2.1. Abstract

We determined the effects of local spruce budworm (*Choristoneura fumiferana* Clem.; SBW) population level, proximity to sites with high SBW populations, insecticide spray, and environmental variables on SBW populations from 2014 to 2018, the outbreak initiation period in northern New Brunswick, Canada. SBW second instar larvae (L2) per branch population data collected at 1100-2000 sample points per year were used to create annual interpolated population rasters. Fishnet sample points extracted from these rasters were overlaid with georeferenced layers of 46 possible predictor variables including forest composition, climate, topography, site quality, and insecticide treatment. Results showed that local SBW population in the previous year, proximity to sites with high SBW populations, and early spring climate were consistently the most important predictors over the 5 study years. Simultaneous autoregressive models were used to address spatial autocorrelation when forecasting the SBW L2 population, and a linear mixed effects model was fit to aggregate data for 2015-2018 combined. The models reduced spatial dependence in the residuals, and explained 68–79% of the annual variance in L2 population levels and 53% of the variance over all 4 years combined. Sensitivity analysis showed that locations with 5-10 more SBW L2 per branch than observed values, or that were 20-40 km closer to high population sites in the previous year could have as many as 24 more L2 in the current year. Cumulative degree days in April helped to estimate the upper and lower bounds of the population. The expansion and retraction of the current SBW outbreak initiation were mathematically described in this study based on the insect population data. Understanding the variables that influence SBW outbreak initiation and how variables impact insect populations can help design

small area target-specific insecticide spray applications and focus subsequent-year sampling design on predicted outbreak hot spots and areas detected in defoliation surveys.

Keywords: *Choristoneura fumiferana*, population prediction, outbreak initiation, simultaneous autoregression, spatial autocorrelation

2.2. Introduction

Synchrony in insect population fluctuations often occurs at the landscape scale. Such large-scale synchronous increases in insect population density during one year and persistence over subsequent years can lead to rising insect populations and outbreak initiation (Royama et al. 2005; Liebhold et al. 2012; Bouchard et al. 2018). Although causes of spatial synchrony of insect outbreaks and local population fluctuations remain unclear, several underlying mechanisms have been proposed, including regional stochasticity, dispersal, and trophic interactions (Régnière and Lysyk 1995; Myers 1998; Williams and Liebhold 2000; Liebhold et al. 2012). Efforts have been made to understand how seasonal drought accelerates outbreak initiation of insect pests (e.g., spruce beetle (*Dendroctonus rufipennis*); Hart et al. 2017)), how stands with more host species abundance attract insect pests' initial attack (e.g., mountain pine beetle (*Dendroctonus ponderosae* Hopkins); Klutsch et al. 2009), and how parasitism produces regional insect pest population patterns during the outbreak onset phase (e.g., western tussock moth (*Orgyia vetusta*); Maron et al. 2001). From a forest resource management viewpoint, the spread of a forest insect outbreak can be controlled or slowed down by altering forest structures if forest characteristics have main effects on the outbreak onset and development (Robert et al. 2012). Alternative strategies will be required if influence from biotic factors are absent (e.g., Bouchard and Auger 2014).

Major spruce budworm (*Choristoneura fumiferana* Clem.; SBW) outbreaks in balsam fir (*Abies balsamea* L. Mill.) and spruce (*Picea* spp. A. Dietr.) forests are the dominant natural disturbance in eastern North America. For example, outbreaks in the 1910s, 1940s, and 1970s damaged 10, 25, and 55 million ha, respectively, across eastern Canada

(Blais 1983). These outbreaks followed about a 30-65 year period, depending upon the region (e.g., Royama 1984; Gray and MacKinnon 2007), and SBW populations in northern New Brunswick, Canada typically rose shortly after the outbreak spread through Québec, Canada from the west (Kettela 1983; Hardy et al. 1986). Another SBW outbreak began in about 2006 in northern Québec, reaching 8.2 million ha of defoliation by 2018 (QMRNF 2018). SBW population increases spread southward into the Gaspé-Bas St. Laurent region just to the north of New Brunswick in 2011, and reached New Brunswick in 2014. In this paper, we analyze the spatial patterns of SBW population increases in New Brunswick from 2013 to 2018, and determine the possible influence of forest composition, climate, topography, and site quality on SBW population increases, assessed using overwintering second instar larvae (L2) sampling on approximately 4,500-6,000 branches per year.

The above four categories of variables were selected based on demonstrated effects on SBW outbreak dynamics: forest composition, climate, topography, and site quality. First, high SBW population levels tended to occur when eggs were deposited in dense and mature softwood stands rather than young and open mixed stands (Greenbank 1957). Higher abundance of balsam fir corresponded to more severe SBW defoliation (Bouchard and Auger 2014), while higher abundance of hardwoods (broadleaved trees) resulted in less severe defoliation (Su et al. 1996; Zhang et al. 2018), reduced tree volume loss (Needham et al. 1999), and decreased radial growth reduction (Campbell et al. 2008). Second, dry and warm summers benefit larval survival, and potentially lead to higher larval populations during the flight season (Greenbank 1957; Royama 1984; Régnière and Nealis 2007). High frequencies of defoliation were associated with dry Junes and

cool springs (Candau and Fleming 2005). Spring and summer degree days showed a strong influence on outbreak duration, severity, and spatial variability as well; locations with warmer spring temperatures and higher cumulative degree-days generally experienced shorter outbreak duration and lower severity (Gray 2008, 2013). Third, proximity to previously defoliated areas had a positive influence, while elevation had a negative influence, on SBW defoliation during the onset of the latest SBW outbreak in Québec (Bouchard and Auger 2014). Fourth, stands on moist/rich sites had 19% higher defoliation than stands on wet/poor sites (MacKinnon and MacLean 2003). Compared to mesic (averaged 45% mortality) and sub-hydric fir stands (27%), xeric (85%) and hydric fir stands (75%) were more vulnerable to tree mortality caused by SBW outbreaks (Dupont et al. 1991). Research on a SBW outbreak in northern British Columbia, western Canada during 1985-1996 noted that increasing stand volume, higher current needle biomass, and proximity to the nearest river or nearest defoliation led to increased likelihood of the onset of outbreaks (Magnussen et al. 2004).

An experiment to suppress rising SBW populations before major defoliation occurs, termed an ‘early intervention strategy’ (EIS), has been underway in New Brunswick, Canada from 2014 to 2018 (MacLean et al. 2019). This is the first attempt of area-wide management (all areas within the jurisdiction of the province of New Brunswick) of an endemic forest insect population. The EIS approach includes intensive monitoring of overwintering SBW to detect ‘hot spots’ of low but rising populations and targeted insecticide treatment to prevent spread. Following 5 years of over 420,000 ha of EIS treatments of low but increasing SBW populations, SBW L2 levels across northern New Brunswick were considerably lower than populations in adjacent Québec (MacLean et al.

2019). SBW populations in blocks treated with *Bacillus thuringiensis* or tebufenozide insecticide were consistently reduced and generally did not require treatment in the subsequent year. SBW populations observed in a given year are the basis for insecticide treatment strategies in the following year (MacLean et al. 2019). Since SBW population increased up to 2017, areas requiring treatment increased up to 2018. However, SBW populations showed over 90% reductions in the autumn 2018 samples that indicated 2019 population levels, resulting in a decline of area that needed treatment in 2019 to only 10,000 ha (Healthy Forest Partnership 2019). Following the initial 5 years of tests, EIS appears to be effective in reducing the SBW outbreak (MacLean et al. 2019). Forecasting regions with high-level SBW populations in the following year is a critical basis for monitoring sampling, insecticide treatment, or other pest management tactics, and understanding the relationship of stand, site, and climate factors to population increases is an important component.

When performing statistical tests on ecological data, such as our SBW L2 population data in northern New Brunswick from 2014-2018, the assumption of independence of residuals is often violated, increasing Type I errors and biasing the estimation of regression parameters (Dormann et al. 2007; Beale et al. 2010). Lack of independence in the residuals can arise because either the response or the predictor variables are spatially autocorrelated, i.e., objects that are closer to each other have a tendency to be more similar than those that are further apart (Sokal and Oden 1978; Dale and Fortin 2014). Spatially-structured residuals may be due to: 1) omission of important predictors in the model, 2) inappropriate model specification, or 3) a mismatch between the spatial patterns of the response and predictors (Dale and Fortin 2014). Accordingly, solutions to

avoid statistical issues resulting from spatial autocorrelation include: 1) incorporating all necessary ecological predictors to fit the model, 2) adopting generalized linear mixed models where a random effect is used to account for the effects of location, and 3) spatial regression that includes 'space' explicitly as an additional variable (Beale et al. 2010; Dale and Fortin 2014). Spatial regression takes into account spatial dependence in the data by adding a lagged response variable (autoregression; Anselin 1988) or lagged covariates (i.e. autocovariate regression; Augustin et al. 1996).

The objectives of this study were to: 1) determine variables that influenced SBW L2 populations from 2014 to 2018, the outbreak initiation phase in northern New Brunswick, and 2) test if the L2 population in the subsequent year could be predicted by L2 distribution in the preceding year and environmental variables. Based on previous research, we included four categories of environmental/site variables that might influence SBW outbreak initiation: 1) forest composition, including percentage of balsam fir and spruce host species and hardwoods; 2) spring and summer climatic conditions (monthly average temperature, monthly cumulative degree-days, and monthly total precipitation); 3) topographic characteristics (elevation, slope, and aspect), and 4) site quality (depth to water (Murphy et al. 2007) and biomass growth index (Hennigar et al. 2017)). We also included local SBW L2 population levels from the previous year, proximity to sites with high SBW populations in the previous year, and insecticide spray treatments in the previous two years as influencing variables. Determining variables that influence SBW outbreak initiation can help to better understand how outbreaks start and facilitate insect management, insecticide treatments, and field sampling design.

2.3. Methods

2.3.1. Study area

The study area was in northern New Brunswick, Canada, covering approximately 64° 30′ to 69° 0′ W and 46° 30′ to 48° 0′ N, and amounting to a 3,730,000 ha area (Figure 2.1a). Species composition of New Brunswick forests is approximately 68% softwood and 32% hardwood (Erdle 2008). SBW host species spruce and balsam fir together comprise more than half of the province-wide forest (55%), followed by red maple (*Acer rubrum* L.) and sugar maple (*Acer saccharum* Marsh.) making up about 15% of the forest. Roughly 20% of the forest is younger than 20 years old, resulting from harvesting; 45% of the forest is older than 60 years old (Erdle 2008). In northern New Brunswick, balsam fir and spruce are the largest species groups (Erdle 2008).

2.3.2. Data collection and preparation

SBW L2 data in New Brunswick from 2013 to 2018 were provided by New Brunswick Department of Energy and Resource Development (NBERD) and the Healthy Forest Partnership Early Intervention Strategy (EIS) Research group. NBERD and forest industry staff from Forest Protection Limited, J.D. Irving, Limited, Acadian Timber Corporation, and Fornebu Lumber Company Inc. jointly collected data from a large number of sample points around New Brunswick. Each L2 point included a sample of one mid-crown branch from each of three balsam fir or three spruce trees (2013-2015), or a mix of balsam fir and spruce representing the stand condition (2016-2018). From 2013 to 2018, 1136, 1503, 1561, 1649, 1964, and 1851 L2 points were sampled per year, respectively. NBERD processed the branch samples by washing them with a sodium

hydroxide solution, and filtering and counting the amount of SBW L2 whose hibernacula were destroyed by the washing solution (Miller et al. 1971; MacLean et al. 2019). The average number of L2 per branch was classified into six classes: nil, trace, low, moderate, high, and extreme (0, 0.1-3.5, 3.6-6.5, 6.6-20.5, 20.6-40.5, and > 40.5 L2 per branch). Moderate or higher classes (> 6.5 L2 per branch) are particularly of concern, since this is the threshold used for insecticide intervention under the EIS approach (MacLean et al. 2019). Across northern New Brunswick, L2 populations increased consistently up to 2017, but declined dramatically in 2018 (Figure 2.1b). SBW L2 data in the adjacent Gaspé-Bas St. Laurent region of Québec from 2013 to 2018 were included in the analyses as well, helping to determine the effect of proximity to moderate or higher levels of L2 population in the previous year, especially for the sample points near the boundary of the two provinces. SBW L2 data for Québec were provided by Gouvernement du Québec, Ministère des Forêts, de la Faune et des Parcs and La Société de Protection des Forêts contre les Insectes et Maladies. L2 abundance was counted on one < 75 cm mid-crown branch sampled from each of three balsam fir or spruce trees. L2 populations in the Gaspé-Bas St. Laurent region continuously increased from 2013 to 2018 (Figure 2.1b).

Point layers of the raw L2 sample data in both New Brunswick and Québec were interpolated into raster layers at 20 m resolution using Inverse Distance Weighted (IDW) methods (Watson and Philip 1985). We tested the layer using data from the nearest 6, 12, and 24 points in calculating the value of the interpolated cell, and found that resulting datasets gave similar results in Gradient Boosting Machine (GBM) tests and statistical models, with a tendency that models with fewer interpolated points performed slightly less well in terms of goodness-of-fit indicators. Therefore, we used results based on 12

points as inputs for the IDW interpolation process. Fishnet L2 points, which are the points at the crosses of a systematic lattice, for each year from 2013-2018 were extracted from the interpolated raster layers at a 2 km interval. A total of 9183 fishnet points were extracted within the study area in each year. Both raw L2 point layers and fishnet L2 layer data were analyzed in this study.

Datasets for all the other influencing variables were available through various sources, and necessary preprocessing was done before analyses, as described in Table 2.1. These included forest composition, climate, proximity to high L2 population sites in the previous year, topography and site quality variables, the previous year SBW L2 population, and insecticide spray treatments in the previous two years (Table 2.1). All raster data used in the research were at 20 m resolution.

2.3.3. Spatial overlay analysis

Annual SBW L2 point layers from 2014 to 2018, including both the raw sample point layers and fishnet point layers, were overlaid with geo-referenced layers of possible influencing variables (Table 2.1), including monthly average temperature from April to September, monthly cumulative degree days (> 5 ℃) from April to September, monthly total precipitation from April to September, elevation, slope, aspect, depth to water (Murphy et al. 2007), and biomass growth index (Hennigar et al. 2017). Previous local SBW population for Year N was extracted from the interpolated raster in Year N-1. Proximity to moderate or higher levels of SBW L2 populations in Year N-1 was computed in three ways: distance to the nearest raw sample point with L2 > 6.5 per branch (ProxRaw), distance to the nearest fishnet point with L2 > 6.5 per branch

(ProxGrid), and distance to the nearest cell with L2 > 6.5 per branch (ProxCell; Table 2.1). SBW L2 data from the Gaspé-Bas St. Laurent, Québec region were incorporated in the proximity calculation, although only those L2 points within the northern New Brunswick study area were used in the statistical analyses. Insecticide spray history was included as a dummy variable, with four numbers representing whether a sample point fell within the spray blocks in the previous two years, in one of the two preceding years, or none of two years. Tree species proportions of balsam fir, spruce, and hardwood were calculated within 50 m circular buffer areas for each L2 point, using the dominant forest layer and an area-based weighted average of species proportion of polygons generated from the LandBase shapefile intersected with each circular buffer area (details of variables and data sources are given in Table 2.1). All spatial analyses above were done using ArcMap 10.4 (ESRI, Redlands, CA, USA).

2.3.4. Variable importance analyses

With all 46 predictor variables included (Table 2.1), GBM analysis (Ridgeway 2007) was used to determine the most influential predictors in forecasting the SBW L2 population, based on both raw sample points and fishnet points. As a machine learning technique, gradient boosting determines the performance of decision trees using gradients in the loss function in a sequential fashion (Friedman 2001). GBM analysis reports the relative influence of analyzed variables according to frequencies that a variable was selected over all splits, weighted by the squared improvement to the model in each split (Elith et al. 2008). The “caret” package (Kuhn 2008) within R version 3.4 (R Development Core Team 2018) was used for the GBM analysis. Correlation tests were

also done for all predictor variables, and highly correlated predictor variables ($r \geq 0.7$) were avoided in the models. The generally most important uncorrelated variables ($r < 0.7$) across all years selected from the GBM-resulted variable ranks and their interactions were used in fitting models.

2.3.5. Spatial regression models

Simultaneous autoregressive (SAR) models (Wall 2004) were implemented to predict SBW L2 population levels in each year from 2014 to 2018. SAR models assume that the response variable depends on not only the predictor variables, but also the spatial neighborhood relationships among all samples in the study area, which were implemented as a $n \times n$ weights matrix in the models (Haining 2003). The specification of the spatial weights usually starts from a binary neighbors list in which objects are either listed as neighbors or are absent, then is further weighted to give less-distant neighbors more weight (Hoef et al. 2018). The weights matrix consists of zeros on the diagonal and weights for the neighbors in the off-diagonal positions (Hoef et al. 2018). Based on results of Kissling and Carl (2008), who compared SAR model performance with 3240 different combinations of model settings, row-standardization (i.e., "W" coding style) was used to code the spatial weights matrix in this study.

We fitted SAR models using only the L2 fishnet points because, compared to the irregular distribution of L2 raw sample points, fishnet sample points have a regular lattice of point sites, which guarantees that all orders of neighborhood sizes remain the same for all samples. This benefits the comparison between models across different years, since in this case, a common neighborhood size can be used in different-year SAR models. In

addition, predicting regions with high L2 populations, instead of points with high L2 populations, has more guiding value in insect sampling and management. Given that our fishnet points were at 2 km intervals, we defined the upper neighbor boundary as 5 km to include eight adjacent neighbors, i.e., first order neighbors, for each sample point, as advised by Kissling and Carl (2008). SAR models take three forms: 1) lagged response model which assumes that the autoregressive processes merely occurs in the response variable, 2) lagged mixed model which assumes such processes occur in both response and predictor variables, and 3) spatial error model which assumes that the process only occurs in the error terms (Anselin 1988). We applied the second form, SAR_{mix} model, because the response variable SBW L2 was likely to have both inherent spatial autocorrelation (i.e., derived from interactions within the observed variable itself) and induced spatial autocorrelation (i.e., the observed variable is functionally dependent on an underlying variable which is autocorrelated). SAR_{mix} models take the form (Anselin 1988):

$$Y = \rho WY + WX\gamma + X\beta + \varepsilon$$

where $X\beta + \varepsilon$ are the standard terms in ordinary least squares (OLS) regression: β is a vector representing the slopes associated with the predictor-matrix X, and ε is a vector of the error terms. ρW is a term describing spatial autocorrelation in the response variable vector Y, where ρ is the autoregression parameter, and W is the spatial weights matrix. $WX\gamma$ describes the regression coefficients γ of the spatially lagged predictors WX. SAR models were implemented using the "spdep" package (Bivand 2006), and the R script used to fit SAR models was derived from the appendix of Dormann et al. (2007). The top predictor variables determined by GBM analysis and their interactions were

included in the full models, and then any non-significant variables or interactions, assessed by *t* tests ($\alpha = 0.05$), were dropped in the reduced models.

Spatial dependence of SAR model residuals was investigated using correlograms of Moran's *I* and maps of residuals. Since parameters of SAR models are estimated by the maximum likelihood method (Wall 2004), the likelihood-based measure of goodness-of-fit, log likelihood, was appropriate for SAR models (Lichstein et al. 2002; Tognelli and Kelt 2004). For the same reason, we used the Nagelkerke pseudo R^2 (Nagelkerke 1991), which is based on likelihood, instead of the traditional R^2, to compare model performance. Nagelkerke pseudo R^2 was computed according to the formula (Nagelkerke 1991):

$$R^2 = 1 - \exp[-2/n(l_A - l_0)]$$

where l_A is the log likelihood of the model to be tested, l_0 is the log likelihood of the null model that merely contains the intercept, and n is the sample size.

In each year from 2014 to 2018, OLS models were fitted with the same predictor variables and interactions to compare with the SAR_{mix} models in terms of spatial dependence as well as overall model performance. The Nagelkerke pseudo R^2 formula above yields the identical value as the traditional R^2 for OLS models (Lichstein et al. 2002). All correlograms in this study were created using functions in the "ncf" package (Bjørnstad et al. 1999) with 500 permutations for each test. Then the significance levels of the coefficient at each lag distance were adjusted by progressive Bonferroni correction, where the Bonferroni-corrected significance level was computed for each distance class separately (Legendre and Legendre 2012):

$$a'(d) = a/d'$$

where a is the commonly used significance level, i.e., $a = 0.05$, d' is the number of tests actually performed up to the specific distance class, a' is the adjusted significance level, and d is the distance class of interest.

Finally, we applied models forecasting the L2 population in Year N to the datasets in the subsequent year using the "predict.sarlm" function (Goulard et al. 2017) in the package "spdep" (Bivand 2006). We also carried out sensitivity analysis on each year's model, by comparing the predicted L2 population under different scenarios systematically varying values of each of three predictor variables while controlling the other two predictors unchanged, to test the degree that each predictor variable influenced the response variable. The predicted fishnet point values were then interpolated as raster layers using the IDW method, and compared with the observed L2 population distributions.

2.3.6. Combined year model

To predict L2 population levels in the coming year using a generalized model formula, linear mixed effects (LME) models were fitted, combining all years' data with year as a random effect and maximum likelihood as the parameter estimation using the "lme" function in the package "nlme" (Lindstrom and Bates 1988). The effects of spatial autocorrelation were addressed using the "corr" argument in the model settings. The L2 population data was log-transformed to stabilize the variance and to improve the normality of the residuals. The full model was fitted using the same predictors and their interactions used in the SAR_{mix} models, with random intercepts and random slopes on all three predictors. Then a reduced model was generated by dropping non-significant

variables and interactions, assessed by *t* tests ($\alpha = 0.05$), with fewer random components in slopes by comparing the full model and the reduced models using likelihood ratio tests. Assumptions of residual normality and homoscedasticity were evaluated using residual plots. To be comparable to each year's SAR_{mix} model, the goodness-of-fit of the combined-year (2015-2018) model was also evaluated by log likelihood and Nagelkerke pseudo R^2. The reduced model was used to predict L2 population from 2015 to 2018, with only the fixed effects included, and compared with the observed L2 population levels L2 populations in 2019 were also predicted using the reduced combined-year model, once climate data needed in the model were available.

2.4. Results

2.4.1. Relative influence of predictor variables

Results of GBM tests showed that local previous-year SBW L2 population levels, proximity to high population sites in the previous year, and climate conditions in spring (April or May) were the most important variables to predict SBW L2 population levels from 2015 to 2018 (Figure 2.2). Specifically, distance to the nearest fishnet point with L2 > 6.5 per branch in the previous year was the most important predictor in 2015 (relative influence = 41%), 2017 (30%), and 2018 tests (16%; Figure 2.2b,d,e). Distance to the nearest raw sample point with L2 > 6.5 per branch in the previous year, which was highly correlated with the distance to the nearest fishnet point (r = 0.9), was the most important predictor in 2016 (relative influence = 24%; Figure 2.2c). Local previous-year L2 population levels was the second most important variable in 2015 (relative influence = 8%) and 2016 (20%), and the third in 2017 (14%) and 2018 (10%; Figure 2.2c-e).

Cumulative degree days in April ranked in the top five variables in all years from 2014 to 2018 (Figure 2.2). In 2014, results showed two other top predictor variables: monthly mean temperature in May (11%) and biomass growth index (10%; Figure 2.2a). Because the L2 population was < 6.5 per branch for all sample points in New Brunswick in 2013, proximity to high SBW populations were all calculated based on distances to the adjacent Gaspé-Bas St. Laurent region. In addition, there were no insecticide spray treatments carried out in 2012 or 2013, so this variable was not included in the 2014 analysis. However, insecticide spray treatment was not included among the top five influencing variables in any of the five years.

To generalize the model form over the years into a form that could be used for multi-year prediction, we used the variables local L2 population in the previous year (L2_PreYear), distance to the nearest fishnet point with L2 > 6.5 per branch in the previous year (ProxGrid), and cumulative degree days in April (DD_04) and their interactions to fit the 2015-2018 models, predicting L2 population over the 4 years. Figure 2.3 shows the trends of these three predictor variables from 2015 to 2019. Cumulative degree days in April consistently increased from 2015 to 2018, but decreased in 2019 (Figure 2.3a); 2015 and 2019 had noticeably cold springs. Previous-year L2 population levels kept increasing from 2015 to 2018 (Figure 2.3b). With the L2 population expansion (Figure 2.3b), there was increasing area with low ProxGrid (0-20 km to the nearest fishnet point with L2 > 6.5 per branch) from 2015 to 2018 (Figure 2.3c). For the 2014 model, we used cumulative degree days in April (DD_04), monthly mean temperature in May (Temp_05), biomass growth index (BGI), and their interactions to predict L2 population levels.

2.4.2. Performance of spatial regression models and combined-year (2015-2018) LME models compared to OLS models

Over the 5 study years, SAR_{mix} models explained 68% to 79% of the variance in the SBW L2 populations, whereas OLS models only explained 6% to 35% of the variance (Table 2.2). SAR_{mix} models had higher log likelihood ratios than OLS models in each year, representing a better fit of the data by SAR_{mix} models (Table 2.2). The predictor DD_04 was non-significant in the 2015-2017 models ($\alpha = 0.05$), while the interaction term of DD_04 and L2_PreYear was significant in years 2015, 2016, and 2018 ($\alpha = 0.05$). Moreover, OLS residuals showed positive spatial autocorrelation up to a distance of 100 km, with the Moran's *I* ranging from 0 to about 0.6 over the 5 years (Figure 2.4a), providing evidence that the assumption of independently distributed residuals was violated. OLS residuals were spatially clustered across all years, with clumps of positive residuals and negative residuals (Figure 2.4b). L2 population level, the response variable, was also spatially autocorrelated up to a distance of 100 km across the years, suggesting that the spatially-correlated L2 population was one of the possible causes of the spatial-correlated residuals at roughly the same spatial scale (Figure 2.4a). However, in contrast to OLS models, correlograms and maps of SAR_{mix} model residuals showed very low spatial autocorrelation: Moran's *I* was less than 0.05 for all tested lag-distances in all study years (Figure 2.4a), and much less spatial heterogeneity was exhibited (Figure 2.4c). Being able to address the spatial autocorrelation in model residuals indicated the suitability of SAR_{mix} models for the ecological data analyzed in this study.

To aggregate multiple years' data and generate a generalized model form, the LME model was fit combining 2015-2018 data. Data in 2014 were excluded when fitting the

2015-2018 combined-year LME model because the top predictors for 2014 data differed from those for 2015-2018 data according to GBM tests (Figure 2.2). The predictor variable L2_PreYear, the interaction term DD_04 * L2_PreYear, and the interaction term L2_PreYear * ProxGrid were kept in the reduced LME model. The reduced LME model with year as a random effect had random intercepts and slopes for the predictor L2_PreYear, because the random slopes on the other two variables DD_04 and ProxGrid did not show improvement of model performance based on likelihood ratio tests. The 2015-2018 combined-year LME model explained 53% of the variance in SBW L2 population levels, versus only 22% of the variance explained by the OLS model (Table 2.2). The log likelihood ratio of the LME model was also higher than the OLS model, suggesting a better goodness-of-fit (Table 2.2).

2.4.3. Relationships of SBW L2 population levels to top predictor variables

We used sensitivity analyses to systematically vary values of the three predictor variables and determine the influence on predicted SBW L2 populations. Results for the variable DD_04 showed that higher L2 population forecasts were associated with lower degree days in 2016 and 2017 (Figure 2.5a), during which years the observed values of DD_04 averaged 21.2 °C·d (σ = 5.7 °C·d) and 23.3 °C·d (σ = 9.4 °C·d), respectively. In 2016, the observed DD_04 -20 °C·d and -10 °C·d scenarios averaged 5.2 and 2.7 L2 per branch increases in the L2 population predictions. In the same year, the observed DD_04 +10 °C·d and +20 °C·d scenarios averaged 1.1 and 1.3 L2 per branch decreases in the population predictions (Figure 2.5a). In 2017, scenarios of the observed DD_04 -20 °C·d and -10 °C·d, and +10 and +20 °C·d averaged 1.8 and 1.0 L2 per branch increases, and

0.6 and 1.1 L2 per branch decreases, respectively (Figure 2.5a). In 2015, a year with a much cooler spring ($\bar{x}$ = 9.1 °C·d, σ = 4.6 °C·d), predicted L2 populations were more sensitive to increases in L2_PreYear than to DD_04 (Figure 2.5a,b).

Sensitivity analyses of L2_PreYear indicated that locations with higher L2 in Year N-1 had more L2 in Year N, although the magnitude of the sensitivity was reduced in 2017 and even lower in 2018 (Figure 2.5b). Because the L2_PreYear used in the sensitivity analyses was set to zero when varying the observed L2 by -5 or -10 produced a negative value, many values of the predictor L2_PreYear were zeros in the actual sensitivity tests. Hence, sensitivity to the predictor L2_PreYear was mostly reflected in the original L2_PreYear +5 or +10 cases, especially in 2015 and 2016 (Figure 2.5b). Specifically, systematically increased L2_PreYear of five larvae per branch averaged 9.8 L2 per branch increases in 2015 L2 forecasts, 12.1 in 2016, 2.1 in 2017, and 0.1 in 2018 (Figure 2.5b). Systematically increased L2_PreYear of 10 larvae per branch averaged 19.5, 24.1, 4.1, and 0.2 per branch increases in L2 forecasts in 2015, 2016, 2017, and 2018, respectively (Figure 2.5b). Moreover, in 2015, 2017, and 2018, locations closer to high L2 sites in Year N-1, i.e. lower values of ProxGrid, were generally associated with higher predicted L2 population in Year N (Figure 2.5c). Compared to the other years, forecasts of 2017 L2 populations were more sensitive to the predictor ProxGrid (Figure 2.5c). Systematically increased ProxGrid of 40 km averaged 0.05, 1.02, and 0.16 L2 per branch decrease in 2015, 2017, and 2018, respectively. In 2016, systematically increased ProxGrid led to slightly higher L2 population predictions: average 0.08 and 0.13 L2 per branch increases in ProxGrid +20 and +40 km cases. A 40 km decrease in ProxGrid contributed to at most an average of 1.2 L2 per branch increase (2017) among the years

2015-2018. Across all tested years, predicted 2018 L2 populations showed less sensitivity to all three predictors, probably because of the relatively low and evenly-distributed L2 observation values in this year ($\bar{x} = 0.3$, $\sigma = 0.7$; Figure 2.1b).

2.4.4. Comparison of observed and predicted SBW L2 populations

Generally, the Year N SAR_{mix} models and combined-year LME model underestimated SBW L2 population levels in Year N, especially at high population levels, from 2015 to 2018. The Year N-1 SAR_{mix} models were more accurate in predicting Year N L2 populations compared to the combined-year (2015-2018) LME model which had a greater bias in L2 population forecasts. In comparison with the observed SBW L2 population distribution (Figure 2.6a), areas with moderate or higher levels of L2 population (L2 > 6.5 per branch) were underestimated by Year N SAR_{mix} models by 14620 (72%), 8140 (7%), 15470 (11%), and 2750 ha (100%) less than the observed area from 2015 to 2018, respectively (Figure 2.6b). In 2015 and 2016, areas with moderate or higher levels of L2 population were underestimated by Year N-1 SAR_{mix} models by 18,856 ha in 2016, and no area with L2 > 6.5 per branch was predicted in 2015 (the SAR_{mix} model in 2014 used different predictor variables from the other three years; Table 2.2; Figure 2.6a,c). However, area with moderate or higher levels of L2 populations were substantially overestimated, by 510,382 and 29,278 ha, by Year N-1 SAR_{mix} models in 2017 and 2018 (Figure 2.6a,c).

The combined-year (2015-2018) LME model did a poor job of predicting moderate or higher L2 populations in any of the sample years (Figure 2.6d). All areas (3,730,000 ha) were predicted to have SBW L2 populations less than 6.5 per branch in 2015 and 2016 by

the combined-year (2015-2018) LME model, whereas actual observed areas were 20,191 ha and 122,738 ha, respectively. The area with moderate or higher L2 population was predicted to be 138,560 ha less than observed in 2017 and 568 ha more than the observed area in 2018 (Figure 2.6a,d), and in 2018 the predicted location differed from the actual location. We used both the 2015-2018 combined-year LME model and the annual 2015-2018 SAR_{mix} models to predict moderate or higher SBW L2 populations in 2019 (Figure 2.6d,e). Except for the 2016 SAR_{mix} model (which hugely overestimated area with L2 > 6.5 per branch at 450,000 ha), the area of moderate or higher L2 populations predicted for 2019 was low, 0-256 ha, because input values of the predictor variable L2_PreYear, i.e., L2 populations in 2018, were almost all at the nil or trace level.

2.5. Discussion

2.5.1. Forecasts of SBW population levels in the coming year

For predicting SBW L2 population levels in the following year, the local previous-year insect population variable had the greatest importance in this study. Locations with higher SBW L2 populations, located closer to high population sites, had higher predicted L2 levels in the subsequent year. Sensitivity analyses showed that increasing the L2 population by five to ten per branch at a site in a given year would result in the L2 population at that site in the subsequent year increasing by 12 to 24 per branch. Being 40 km closer to or further from locations with > 6.5 L2 per branch resulted in only having 1 more or 1 less L2 per branch in the next year. Rising populations due to large-scale pulses of resource or important variables in a given year would potentially remain at the same level over the following years, which can contribute to an outbreak initiation

(Bouchard et al. 2018). Although SBW populations experienced over 90% reductions unexpectedly in 2018, the reason for the decline is currently unknown, although study is continuing, and it may well represent a temporary annual reduction in an increasing population trend (MacLean et al. 2019). The quantified relationships found, however, were not consistent across all years. Thus, a model fit in Year N-2 could not be used to reliably predict L2 populations in Year N. Moreover, populations in Year N predicted by the Year N-1 SAR_{mix} models were generally more accurate than the population predicted by the combined-year (2015-2018) LME model, which consistently underestimated L2 populations. The effect of cumulative degree days in April in the predicted year models may help to characterize the upper and lower population bounds when forecasting the SBW population.

2.5.2. Effects of environmental variables on outbreak initiation

Early spring climate conditions were generally the most important influencing variable among all the environmental variables analyzed in this study. Specifically, higher L2 populations were associated with lower early spring degree days in 2016 and 2017. Nonetheless, effects of spring degree days on L2 population differed for different previous-year outbreak conditions, including local population and proximity to high population locations, as suggested by the statistically significant interaction terms. Previous SBW outbreaks in New Brunswick beginning in 1912 and 1949 tended to occur several years after outbreaks occurred in adjacent Québec, and the role of moth dispersal was believed to be important (Greenbank 1957). In our case, the possible effect of moth dispersal from the current SBW epidemic in Québec was incorporated by including

proximity to high L2 population sites in Québec for sample points in the northern New Brunswick study area. The current SBW outbreak onset in New Brunswick expanded from the north, where the climate is relatively colder.

In previous studies, locations that experienced cooler springs were found to have higher defoliation frequencies from 1967 to 1998 in Ontario, and longer outbreak duration and higher severity from 1961 to 1990 in Québec (Candau and Fleming 2005; Gray 2013). It was indicated that a cooler April, typically colder than about 15 °C, can potentially lead to late L2 emergence from hibernacula with longer duration remaining as L2 before reaching the third instar (Rose and Blais 1954). The lower frequencies of defoliation with warm spring temperatures (Candau and Fleming 2005) were suggested to be caused by early emerged larvae suffering from heavy mortality later due to the exposure to late frosts and a frozen-ground-caused phenological mismatch. However, from a macroscopic scale, northern New Brunswick experienced increasingly warm springs from 2015 to 2018, when observed L2 populations experienced increases until 2017 and an unexpected reduction in 2018. It has been reported that warmer climate conditions lead to a better phenological match between the larvae emerging from overwintering and the young developing foliage (Bouchard et al. 2018), and this better match results in lower larval starvation and less need to disperse, which results in lower larval mortality (Miller 1958). Extreme cold winter temperature, typically below -40 °C, can cause mortality of overwintering L2 (Blais 1958). We evaluated this possible effect, but meteorological data showed that very few of the sample points in our study area experienced this low threshold temperature in winter during the study years.

Although forest composition has also been viewed as an important variable influencing SBW survival by affecting the diversity and populations of SBW parasitoids (Zhang et al. 2018), or the phenological matches between host trees and the insects (Nealis and Régnière 2004), species composition did not appear as an important influencing variable on SBW population in this study. This may have been because the L2 data used in the research were collected from branches from each of three balsam fir or spruce trees, and the fishnet L2 data for modeling was interpolated from the L2 points, during which process the resultant values highly depended on the distances to the surrounding raw L2 points. Nonetheless, dense and mature softwood forest, favorable for adult SBW dispersal, is associated with SBW outbreaks (Greenbank 1957).

Topography and site quality did not show clear influences on SBW population during the analysis. Despite the fact that the insect populations > 6.5 L2 per branch were nearly all within insecticide spray blocks in the EIS project, and SBW populations within treated blocks were consistently reduced (MacLean et al. 2019), effects of insecticide spray treatments were not important in our analysis.

2.5.3. Regression analysis of spatial population data

Better performance and less spatial autocorrelation in residuals demonstrated the suitability of SAR models compared to classic OLS models for analyzing these ecological data. These results suggest that modeling forest insect population dynamics should account for spatial structures in both target-insect distribution and underlying environmental factors. Given the spatial dynamics and periodicity of defoliator systems, it is essential to carry out outbreak phase-dependent analyses and be careful about the

study scale selection. There are many different statistical approaches to studies of spatial ecological data, depending on study objectives, data characteristics, sample sizes, model types, sources of spatial autocorrelation, etc. (Dormann et al. 2007; Beale et al. 2010). Performance of different spatial models has been estimated and compared in several studies, which provide evidence of model application in different conditions (e.g., Dormann et al. 2007; Kissling and Carl 2008; Bini et al. 2009; Beale et al. 2010).

2.5.4. Applications of the study results

The onset of the current SBW outbreak from 2015 (first forecast of defoliation by NBERD) to 2018 in New Brunswick gave us a good opportunity to explore possible factors influencing the initiation of SBW outbreaks, and what factors contribute to the development of population increases. Understanding the relationships among variables, and possible mechanisms, can help forest and pest managers to forecast insect outbreak situations in the future and make better management decisions. In the SBW case, identifying subsequent-year high SBW population areas benefits better sampling design, for example, to reduce the number of sampling sites while still obtaining reasonable L2 density estimates for a given area. In theory, rather than having a consistent sample of L2 points across northern New Brunswick, it would be more efficient to stratify sampling intensity by likelihood of upcoming year L2 populations predicted to exceed the EIS treatment threshold of 6.5 L2 per branch. Additionally, our results suggested that previous-year outbreak conditions were important, along with other factors, in predicting current-year SBW population levels. Therefore, applying insecticides on those outbreak "hot spots" in the preceding year to prevent further growth of local SBW populations is

advisable, which corresponds to the current strategy used by EIS project (MacLean et al. 2019). Attention to such temporal correlation of populations should be given when defining management strategies for other insect pests with similar population dynamics or mobility as SBW. It should be noted that, however, that factors influencing insect outbreaks are phase-dependent (e.g., Bouchard and Auger 2014), and the relationships between factors determined during an outbreak initiation can be different during the development or collapse phases of the outbreak.

2.6. Conclusions

Using spatial regression models, we determined the variables that influenced insect populations during the increases the SBW populations occurring in northern New Brunswick from 2014 to 2018. Variables important in determining upcoming year SBW L2 populations included previous year local L2 populations, proximity to high L2 population locations, and early spring degree days. Nonetheless, since the relationships found and quantified in this study were inconsistent across years, the model fit in a given year may not be applicable to predict L2 populations two or more years afterwards. Influencing factors determined during the outbreak initiation in this study may weaken or vary in spread and collapse phases of the outbreak. The combined-year (2015-2018) LME model performed the worst and underestimated L2 populations in the forecasts. Other variables including forest species composition, topography, site quality, and insecticide spray history were not important. High SBW L2 populations tended to be at locations closer to or at the sites where the previous year local SBW population was high. Due to the fact that the current SBW outbreak in New Brunswick is expanding from the

north, adjacent Gaspé-Bas St. Laurent region of Québec, the most northern region in New Brunswick suffered from the epidemic earlier than southern regions of the province. This study mathematically describes the expansion and retraction of the current SBW outbreak, based on the insect population data, in northern New Brunswick over the initial 5 years. Our results help determine which variables influenced SBW outbreak initiations and relationships between the SBW populations and the influencing variables, and also demonstrate the suitability of spatial regression models for ecological data.

2.7. Acknowledgments

This research was funded by the Spruce Budworm Early Intervention Strategy project supported by the Healthy Forest Partnership, Atlantic Innovation Fund, and by Natural Resources Canada. We thankfully acknowledge contributions in providing the research data from the New Brunswick Department of Energy and Resource Development, Spruce Budworm Early Intervention Strategy Research group, Gouvernement du Québec, Ministère des Forêts, de la Faune et des Parcs, La Société de Protection des Forêts contre les Insectes et Maladies, and the Forest Watershed Research Centre within the Faculty of Forestry and Environmental Management at the University of New Brunswick. We also thank Luke Amos-Binks and Jeremy Gullison of NBERD who provided access to the SBW L2 data and the insecticide spray treatment data.

2.8. References

Anselin, L. 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht, Netherlands.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. J. Appl. Ecol. **33**(2): 339-347.

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J., and Elston, D.A. 2010. Regression analysis of spatial data. Ecol. Lett. **13**: 246-264.

Bini, L.M., Diniz-filho, J.A.F., Rangel, T.F.L.V.B., Akre, S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araújo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Böhning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Jr, P.D.M., David, S., Ferrer-castán, D., Field, R., Filloy, J., Fleishman, E., Jose, F., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araijo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Marco, P.D., Dobkin, D.S., Ferrer-castain, D., Field, R., Filloy, J., Fleishman, E., Gmez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, J., Len-cortes, J.L., Lobo, J.M., Montoya, D., Morales-castilla, I., Moreno, J.C., Oberdorff, T., Olalla-t, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R., and Hawkins, B.A. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography **32**(2): 193-204.

Bivand, R. 2006. Implementing spatial data analysis software tools in R. Geogr. Anal. **38**: 23-40.

Bjørnstad, O.N., Ims, R.A., and Lambin, X. 1999. Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends Ecol. Evol. **14**(11): 427-432.

Blais, J.R. 1958. Effects of 1956 spring and summer temperatures on spruce budworm populations (*Choristoneura fumiferana* Clem.) in the Gaspé peninsula. Can. Entomol. **90**: 354-361.

Blais, J.R. 1983. Trends in the frequency, extent, and severity of spruce budworm outbreaks in eastern Canada. Can. J. For. Res. **13**(4): 539–547.

Bouchard, M. and Auger, I. 2014. Influence of environmental factors and spatio-temporal covariates during the initial development of a spruce budworm outbreak. Landsc. Ecol. **29**: 111-126.

Bouchard, M., Régnière, J., and Therrien, P. 2018. Bottom-up factors contribute to large-scale synchrony in spruce budworm populations. Can. J. For. Res. **48**(3): 277-284.

Campbell, E.M., MacLean, D.A., and Bergeron, Y. 2008. The severity of budworm-caused growth reductions in balsam fir/spruce stands varies with the hardwood content of surrounding forest landscapes. For. Sci. **54**(2): 195-205.

Candau, J.-N. and Fleming, R.A. 2005. Landscape-scale spatial distribution of spruce budworm defoliation in relation to bioclimatic conditions. Can. J. For. Res. **35**(9): 2218-2232.

Dale, M.R. and Fortin, M.J. 2014. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Dormann, C.F., Mcpherson, J.M., Arau, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Ohlemu, R., Peres-neto, P.R., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**: 609-628.

Dupont, A., Bélanger, L., and Bousquet, J. 1991. Relationship between balsam fir vulnerability to spruce budworm and ecological site conditions of fir stands in central Québec. Can. J. For. Res. **21**(12): 1752-1759.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**: 802-813.

Erdle, T.A. 2008. Management alternatives for New Brunswick' s public forest: report of the New Brunswick task force on forest diversity and wood supply. New Brunswick Department of Natural Resources, Fredericton, NB, Canada. pp. 108.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5): 1189-1232.

Furze, S., Ogilvie, J., and Arp, P.A. 2017. Fusing digital elevation models to improve hydrological interpretations. J. Geogr. Inf. Syst. **9**: 558-575.

Goulard, M., Laurent, T., and Thomas-Agnan, C. 2017. About predictions in spatial autoregressive models: optimal and almost optimal strategies. Spatial Econ. Anal. **12**(2-3): 304-325.

Gray, D.R. 2008. The relationship between climate and outbreak characteristics of the spruce budworm in eastern Canada. Clim. Change **87**: 361-383.

Gray, D.R. 2013. The influence of forest composition and climate on outbreak characteristics of the spruce budworm in eastern Canada. Can. J. For. Res. **43**(12): 1181-1195.

Gray, D.R. and MacKinnon, W.E. 2007. Historical spruce budworm defoliation records adjusted for insecticide protection in New Brunswick, 1965–1992. J. Acad. Entomol. Soc. **3**: 1-6.

Greenbank, D.O. 1957. The role of climate and dispersal in the initation of outbreaks of the spruce budworm in New Brunswick II. The role of dispersal. Can. J. Zool. **35**: 385-403.

Haining, R. 2003. Spatial data analysis: theory and practice. Cambridge University Press, Cambridge, UK.

Hardy, Y., Mainville, M., and Schmitt, D.M. 1986. Spruce budworms handbook: an atlas of spruce budworm defoliation in eastern North America, 1938-1980. U.S. Department of Agriculture, Forest Service, Cooperative State Research Service, Washington, DC, USA.

Hart, S.J., Veblen, T.T., Schneider, D., and Molotch, N.P. 2017. Summer and winter drought drive the initiation and spread of spruce beetle outbreak. Ecology **98**(10): 2698-2707.

Healthy Forest Partnership. 2019. Research Area Map. Available from http://www.healthyforestpartnership.ca/en/research/what-where-and-when [accessed 3 May 2019].

Hennigar, C., Weiskittel, A., Allen, H.L., and MacLean, D.A. 2017. Development and evaluation of a biomass increment-based index for site productivity. Can. J. For. Res. **47**(3): 400-410.

Hoef, J.M.V., Hanks, E.M., and Hooten, M.B. 2018. On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spat. Stat. **25**: 68-85.

Kettela, E.G. 1983. A cartographic history of spruce budworm defoliation 1967 to 1981 in eastern North America. Canadian Forestry Service, Environment Canada., Ottawa, ON, Canada. Inf. Rep. DPC-X-14.

Kissling, W.D. and Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecol. Biogeogr. **17**: 59-71.

Klutsch, J.G., Negron, J.F., Costello, S.L., Rhoades, C.C., West, D.R., Popp, J., and Caissie, R. 2009. Stand characteristics and downed woody debris accumulations associated with a mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak in Colorado. For. Ecol. Manage. **258**(5): 641-649.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**(5): 1-26.

Legendre, P. and Legendre, L. 2012. Numerical ecology. Elsevier, Amsterdam, Netherlands.

Lichstein, J.W., Simons, T.R., Shriner, S.A., and Franzreb, K.E. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecol. Monogr. **72**: 445-463.

Liebhold, A.M., Haynes, K.J., and Bjørnstad, O.N. 2012. Spatial synchrony of insect outbreaks. *In* Insect outbreaks revisited. *Edited by* P. Barbosa, D.K. Letourneau, and A.A. Agrawal. Blackwell Publishing Ltd., Chichester, UK. pp. 113-125.

Lindstrom, M.J. and Bates, D.M. 1988. Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J. Am. Stat. Assoc. **83**(404): 1014-1022.

MacKinnon, W.E. and MacLean, D.A. 2003. The influence of forest and stand conditions on spruce budworm defoliation in New Brunswick, Canada. For. Sci. **49**(5): 657-667.

MacLean, D.A., Amirault, P., Amos-Binks, L., Carleton, D., Hennigar, C., Johns, R., and Régnière, J. 2019. Positive results of an early intervention strategy to suppress a spruce budworm outbreak after five years of trials. Forests **10**(5): 448.

Magnussen, S., Boudewyn, P., and Alfaro, R. 2004. Spatial prediction of the onset of spruce budworm defoliation. For. Chron. **80**(4): 485-494.

Maron, J. L., Harrison, S., and Greaves, M. 2001. Origin of an insect outbreak: escape in space or time from natural enemies? Oecologia **126**(4): 595-602.

Miller, C.A. 1958. The measurement of spruce budworm populations and mortality during the first and second larval instars. Can. J. Zool. **36**: 409-422.

Miller, C.A., Kettela, E.G., and McDougall, G.A. 1971. A sampling technique for overwintering spruce budworm and its applicability to population surveys. Canadian Forestry Service, Fredericton, NB, Canada. Inf. Rep. M-X-25.

Murphy, P.N., Ogilvie, J., Connor, K., and Arp, P.A. 2007. Mapping wetlands: a comparison of two different approaches for New Brunswick, Canada. Wetlands **27**: 846-854.

Myers, J.H. 1998. Synchrony in outbreaks of forest Lepidoptera: a possible example of the Moran effect. Ecology **79**(3): 1111-1117.

Nagelkerke, N.J.D. 1991. A note on a general definition of the coefficient of determination. Biometrika **78**(3): 691-692.

Nealis, V.G. and Régnière, J. 2004. Insect–host relationships influencing disturbance by the spruce budworm in a boreal mixedwood forest. Can. J. For. Res. **34**(9): 1870-1882.

Needham, T., Kershaw, J.A., MacLean, D.A., and Su, Q. 1999. Effects of mixed stand management to reduce impacts of spruce budworm defoliation on balsam fir stand-level growth and yield. North. J. Appl. For. **16**(1): 19-24.

QMRNF. 2018. Aires Infestées par la Tordeuse des Bourgeons de l'épinette au Québec en 2018. Québec Ministère des Ressources Naturelles et de la Faune. Available from http://www.mffp.gouv.qc.ca/publications/forets/fimaq/insectes/tordeuse/TBE_2018_P.pdf [accessed 2 January 2019].

R Development Core Team. 2018. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Régnière, J. and Lysyk, T. 1995. Population dynamics of the spruce budworm, *Choristoneura fumiferana*. *In* Forest Insect Pests in Canada. *Edited by* J.A. Armstrong and W.G.H. Ives. Canadian Forest Service, Science and Sustainable Development Directorate, Ottawa, ON, Canada. pp. 95-105.

Régnière, J. and Nealis, V.G. 2007. Ecological mechanisms of population change during outbreaks of the spruce budworm. Ecol. Entomol. **32**: 461-477.

Régnière, J., Saint-Amant, R., and Béchard, A. 2014. BioSIM 10 user's manual. Quebec,Canada, pp. 1-65.

Ridgeway, G. 2007. Generalized boosted models: a guide to the GBM package. R Package Vignette.

Robert, L.E., Kneeshaw, D., and Sturtevant, B.R. 2012. Effects of forest management legacies on spruce budworm (*Choristoneura fumiferana*) outbreaks. Can. J. For. Res. **42**(3): 463-475.

Rose, A.H. and Blais, J.R. 1954. A relation between April and May temperatures and spruce budworm larval emergence. Can. Entomol. **86**: 174-177.

Royama, T. 1984. Population dynamics of the spruce budworm *Choristoneura fumiferana*. Ecol. Monogr. **54**(4): 429-462.

Royama, T., MacKinnon, W.E., Kettela, E.G., Carter, N.E., and Hartling, L.K. 2005. Analysis of spruce budworm outbreak cycles in New Brunswick, Canada, since 1952. Ecology **86**(5): 1212-1224.

Sokal, R.R. and Oden, N.L. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10**: 199-228.

Su, Q., MacLean, D.A., and Needham, T.D. 1996. The influence of hardwood content on balsam fir defoliation by spruce budworm. Can. J. For. Res. **26**(9): 1620-1628.

Tognelli, M.F. and Kelt, D.A. 2004. Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. Ecography **27**: 427-436.

Wall, M.M. 2004. A close look at the spatial structure implied by the CAR and SAR models. J. Stat. Plan. Inference. **121**: 311-324.

Watson, D.F. and Philip, G.M. 1985. A refinement of inverse distance weighted interpolation. Geoprocessing **2**: 315-327.

Williams, D.W. and Liebhold, A.M. 2000. Spatial synchrony of spruce budworm outbreaks in eastern North America. Ecology **81**(10): 2753-2766.

Zhang, B., MacLean, D.A., Johns, R.C., and Eveleigh, E.S. 2018. Effects of hardwood content on balsam fir defoliation during the building phase of a spruce budworm outbreak. Forests **9**: 530.

Table 2.1: Predictor variables (abbreviations, units, and descriptions) included in Gradient Boosting Machine analysis to determine their relative importance in predicting L2 population in Year N.

Influencing Factor	Variable Abbreviation	Unit	Variable Description
Forest composition[1]	BF%	%	Weighted average of balsam fir proportion within 50 m buffer zone
	SP%	%	Weighted average of spruce proportion within 50 m buffer zone
	HW%	%	Weighted average of hardwood proportion within 50 m buffer zone
Climate conditions[2]	Temp_04 to _09	°C	Monthly temperature (Apr. to Sept.)
	Temp_av45	°C	Average temperature in Apr. and May
	Temp_av67	°C	Average temperature in Jun. and Jul.
	Temp_av89	°C	Average temperature in Aug. and Sept.
	Temp_av456	°C	Average temperature in Apr. to Jun.
	Temp_av789	°C	Average temperature in Jul. to Sept.
	DD_04 to _09	°C·d	Cumulative degree days (>5°C) (Apr. to Sept.)
	DD_45	°C·d	Cumulative degree days (>5°C) in Apr. and May
	DD_67	°C·d	Cumulative degree days (>5°C) in Jun. and Jul.
	DD_89	°C·d	Cumulative degree days (>5°C) in Aug. and Sept.
	DD_456	°C·d	Cumulative degree days (>5°C) in Apr. to Jun.
	DD_789	°C·d	Cumulative degree days (>5°C) in Jul. to Sept.
	Prcp_04 to _09	mm	Monthly total precipitation (Apr. to Sept.)
	Prcp_45	mm	Total precipitation in Apr. and May
	Prcp_67	mm	Total precipitation in Jun. and Jul.
	Prcp_89	mm	Total precipitation in Aug. and Sept.
	Prcp_456	mm	Total precipitation in Apr. to Jun.
	Prcp_789	mm	Total precipitation in Jul. to Sept.
Proximity to high L2 population sites in the previous year	ProxRaw	km	Distance to the nearest raw sample point with L2 > 6.5 per branch in Year N-1
	ProxGrid	km	Distance to the nearest fishnet point with L2 > 6.5 per branch in Year N-1
	ProxCell	km	Distance to the nearest raster cell with L2 > 6.5 per branch in Year N-1
Topography[3]	DEM	m	Elevation
	Slope	°	Slope values
	Aspect	°	Aspect values
Site quality	DTW[3]	m	Depth to water
	BGI[4]	kg $ha^{-1}yr^{-1}$	Biomass growth index
Previous local SBW population[5]	L2_PreYear	per branch	L2 population at certain sample sites in the preceding year
Spray treatment history[6]	TreatCode		If the point falls in the spray treatment blocks 0: not treated either in Year N-1 or Year N-2 1: treated in Year N-1, not treated in Year N-2 2: treated in Year N-2, not treated in Year N-1 3: treated both in Year N-1 and Year N-2

[1]Forest data was obtained from New Brunswick Energy and Resource Development as a LandBase polygon layer. The shapefile was intersected with 50 m circular buffer zones around each L2 fishnet point. Species composition was calculated as the weighted average of species proportion based on the area of intercepted polygons.

[2]Climate data was generated by BioSIM11 (Régnière et al. 2014) which used four climate databases: Canadian Climate Normals database; New Brunswick Fire Weather; New Brunswick Agriculture Weather;

and Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques (MDDELCC). Processes under BioSIM11 included enquiring daily weather data for New Brunswick, transforming daily data into monthly data, and generating climate data for sample points by interpolation.

[3]Elevation and depth to water data were provided by Forest Watershed Research Centre at the University of New Brunswick (Murphy et al. 2007; Furze et al. 2017), and were resampled into 20 m resolution for this study. Slope and aspect raster layers were generated from the elevation raster, and smoothed by focal statistics with 3x3 cells as neighborhood.

[4]Biomass growth index was created by FORUS Research (Hennigar et al. 2017).

[5]Previous local SBW population was extracted from the Year N-1 interpolated L2 raster.

[6]Spray treatment shapefiles in the previous two years were provided as polygon layers by Early Intervention Strategy Research group. Whether a sample point was previously treated or not was identified by intersecting the point with the treatment blocks in each year.

Table 2.2: Nagelkerke pseudo R^2 and log likelihood of L2 population predictions by ordinary least squares (OLS) and simultaneous autoregressive mixed (SAR_{mix}) models over the years from 2014 to 2018, and by the combined linear mixed effects (LME) model from 2015 to 2018. Predictor variable abbreviations are described in Table 2.1.

Year	**Regression Type**	**Predictors[1]**	**Nagelkerke pseudo R^2 [2]**	**Log Likelihood**
2014	SAR_{mix}	DD_04+Temp_05+BGI	0.75	-1342
	OLS		0.06	-5676
2015	SAR_{mix}	L2_PreYear+ProxGrid+ DD_04*	0.72	-8648
	OLS	L2_PreYear	0.29	-12965
2016	SAR_{mix}	L2_PreYear+ProxGrid+ DD_04*	0.79	-13528
	OLS	L2_PreYear+ L2_PreYear* ProxGrid	0.35	-18802
2017	SAR_{mix}	L2_PreYear+ProxGrid+ DD_04* ProxGrid	0.77	-15367
	OLS		0.35	-20063
2018	SAR_{mix}	DD_04+L2_PreYear+ DD_04* L2_PreYear+	0.68	-3916
	OLS	DD_04* ProxGrid+L2_PreYear* ProxGrid	0.14	-8437
2015-2018	LME	L2_PreYear+ DD_04* L2_PreYear+	0.53	-13858
	OLS	L2_PreYear* ProxGrid	0.22	-17833

[1]Only significant variables and interactions were kept in the reduced models ($p < 0.05$).
[2]For OLS models, Nagelkerke pseudo R^2 yields the identical value as the traditional R^2, i.e., the proportion of the variance for a dependent variable explained by independent variables.

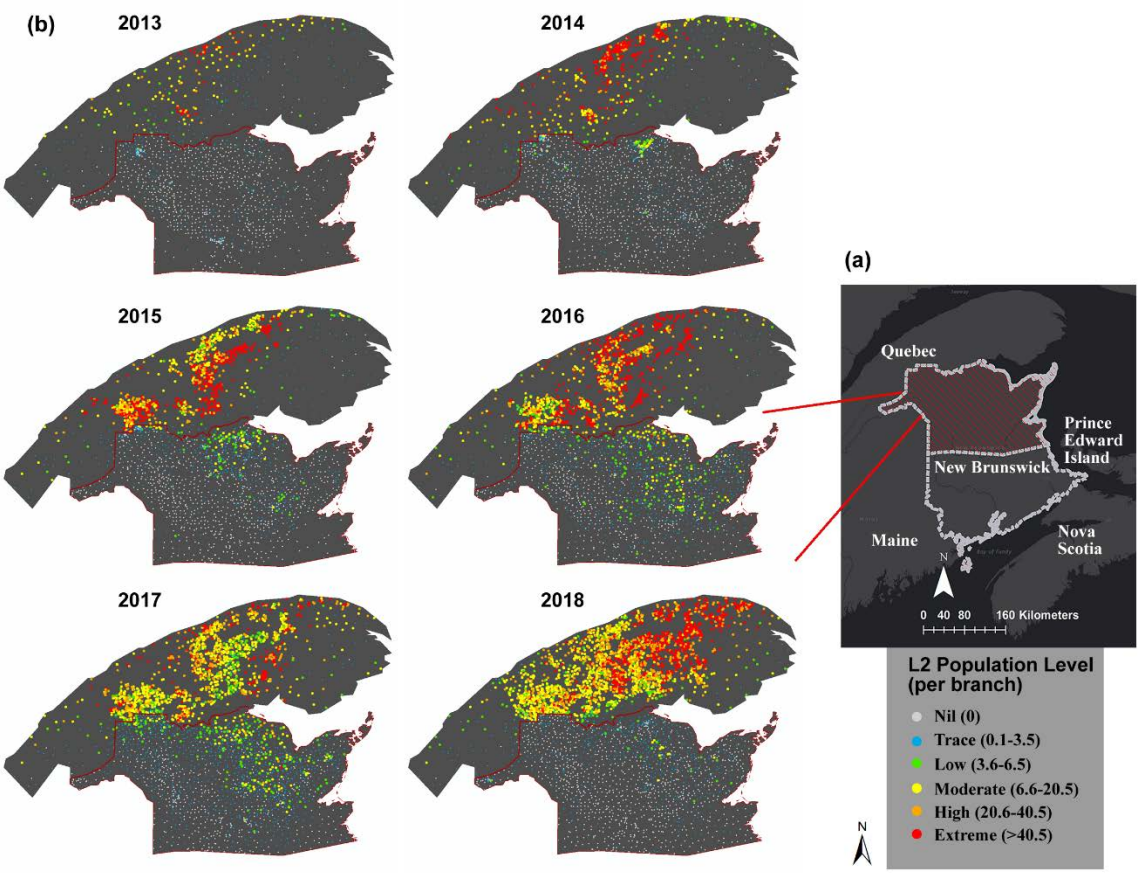

Figure 2.1**:** Location of the study area in northern New Brunswick (a) and the distribution of annual SBW L2 populations in sample points within the study area and the adjacent Gaspé-Bas St. Laurent, Québec, from 2013 to 2018 (b).

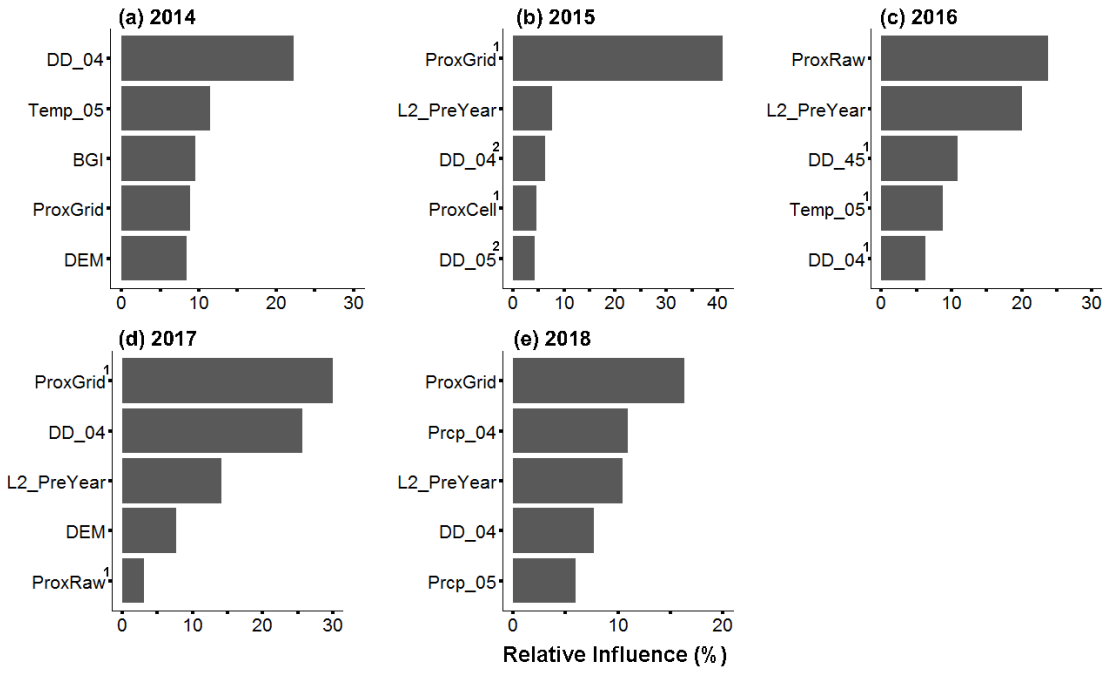

Figure 2.2: Relative influence (%) of the five most important predictor variables based on Gradient Boosting Machine analysis to predict SBW L2 population each year from 2014 to 2018 (a-e). Predictor variable abbreviations are described in Table 2.1, and predictors with the same superscripts were highly and positively correlated with each other (correlation coefficient $r \geq 0.7$).

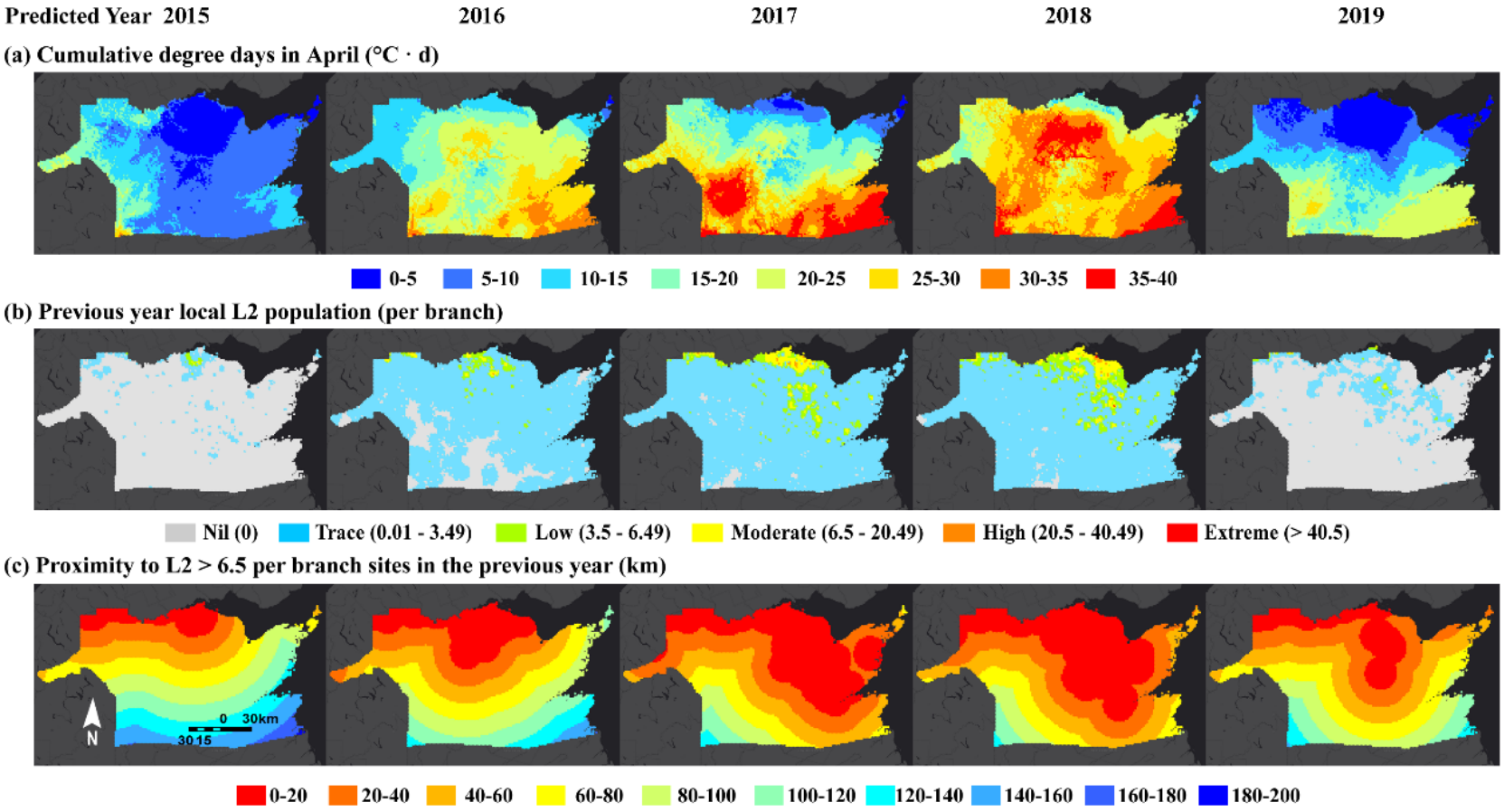

Figure 2.3: Maps of (a) cumulative degree days in April (°C·d), (b) previous local SBW L2 population (per branch), and (c) proximity to L2 > 6.5 per branch sites in the previous year (km) from 2015 to 2018. These three variables were generally the most important predictors and were used in the modeling process.

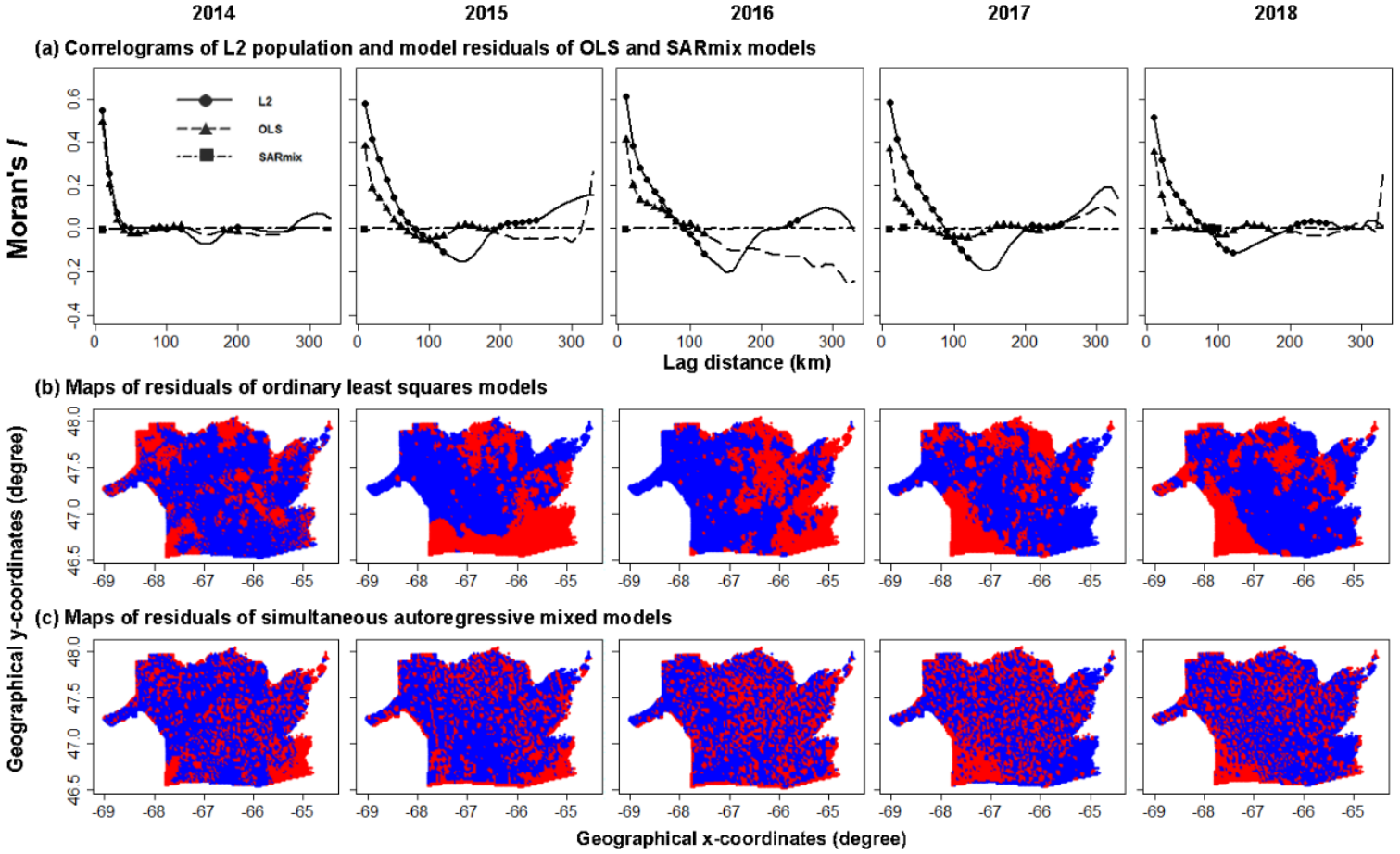

Figure 2.4: Spatial autocorrelation of SBW L2 population and model residuals of ordinary least squares (OLS) models and simultaneous autoregressive mixed (SAR_{mix}) models over 5 modeling years from 2014 to 2018, shown by (a) correlograms, where solid black symbols indicate significant coefficients after progressive Bonferroni corrections ($\alpha = 0.05$, 500 permutations), and spatial error maps of (b) OLS models and (c) SAR_{mix} models, where red area indicates positive residuals and blue area indicates negative residuals.

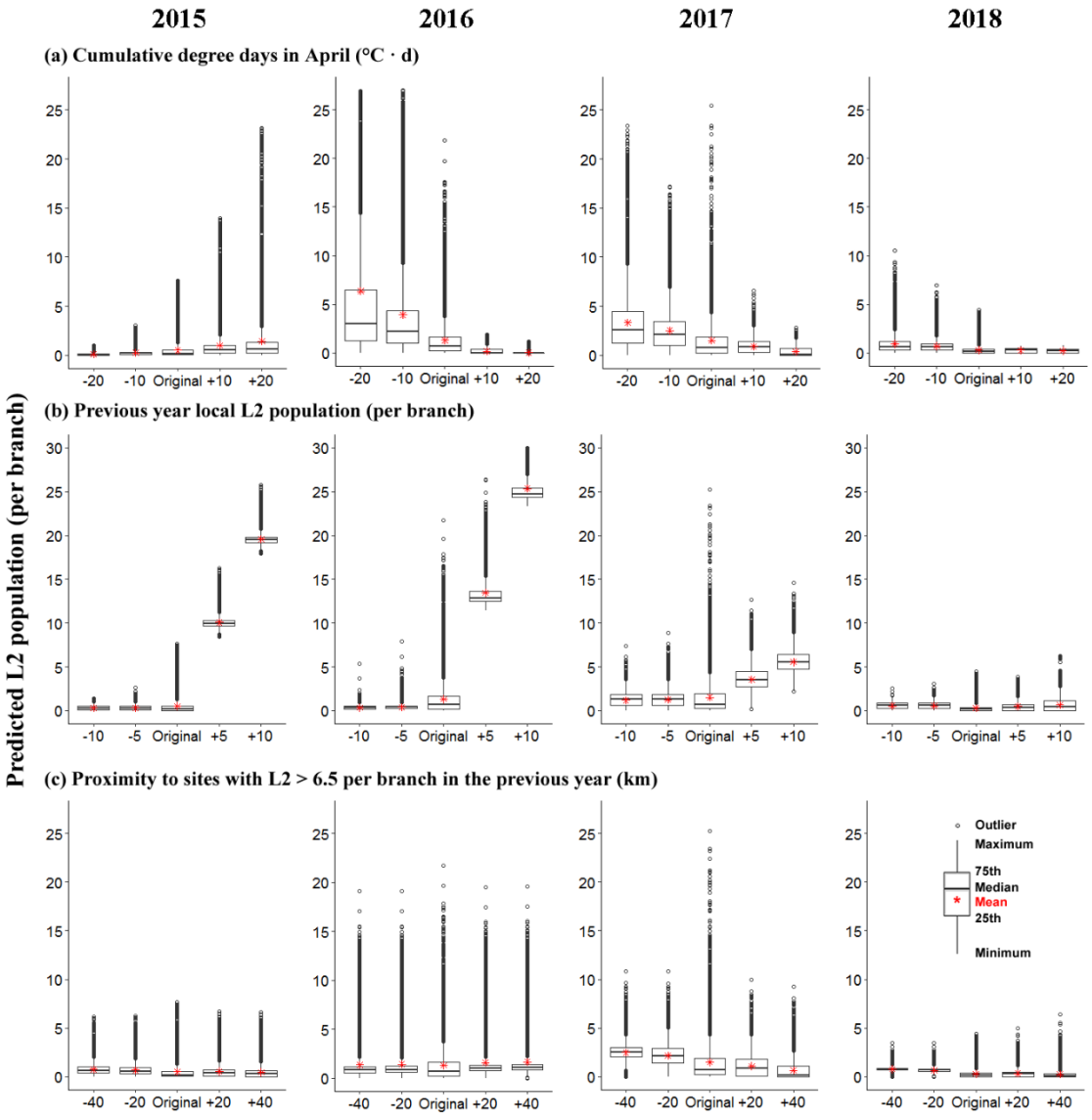

Figure 2.5: Sensitivity analyses of SBW L2 population (per branch) predicted by simultaneous autoregressive mixed models from 2015 to 2018 under three scenarios: (a) original values of cumulative degree days in April compared to original values -20, -10, +10, and +20 °C·d, (b) original values of previous local L2 population compared to original values -10, -5, +5, and +10 per branch, and (c) original values of proximity to L2 > 6.5 per branch sites in the previous year compared to original values -40, -20, +20, and +40 km. Previous local L2 population and proximity with values < 0 in (b) and (c) were set to zero before prediction.

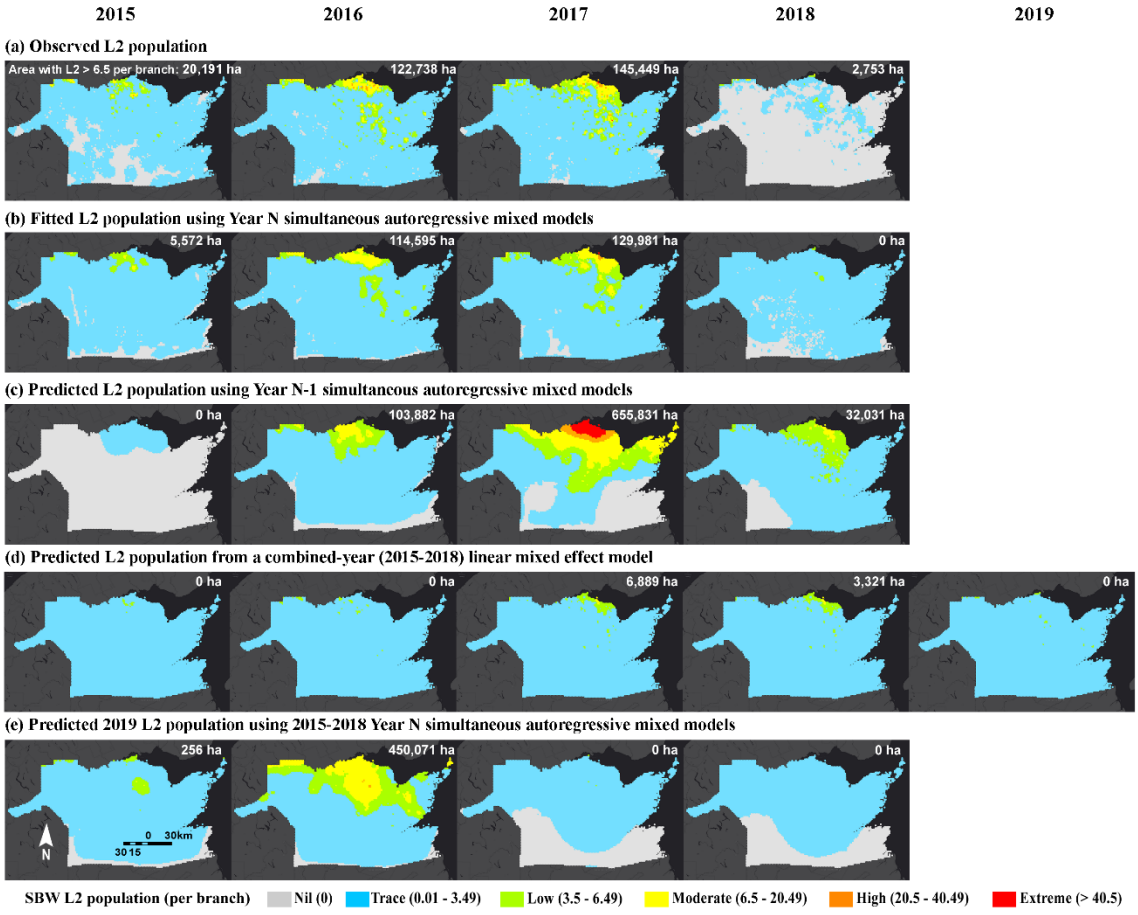

Figure 2.6: Comparison of the (a) observed L2 population spatial distribution, and the forecast spatial distribution, i.e., interpolated raster from L2 population point of (b) fitted values from Year N simultaneous autoregressive mixed models, (c) predicted values from Year N-1 simultaneous autoregressive mixed models, (d) predicted values from a combined-year (2015-2018) linear mixed effect model, and (e) 2019 L2 population predictions using annual Year N simultaneous autoregressive mixed models from 2015 to 2018. Areas (ha) with L2 > 6.5 per branch were estimated based on raster cells.

CHAPTER 3: SPATIAL-TEMPORAL PATTERNS OF SPRUCE BUDWORM DEFOLIATION WITHIN PLOTS IN QUÉBEC

Published in the journal *Forests* on March 6, 2019

Li, M.; MacLean, D.A.; Hennigar, C.R.; Ogilvie, J. 2019. Spatial-temporal patterns of spruce budworm defoliation within plots in Québec. Forests 10, 232.

3.1. Abstract

We investigated the spatial-temporal patterns of spruce budworm (*Choristoneura fumiferana* Clem.; SBW) defoliation within 57 plots over 5 years during the current SBW outbreak in Québec. Although spatial-temporal variability of SBW defoliation has been studied at several scales, the spatial dependence between individual defoliated trees within a plot has not been quantified, and effects of defoliation level of neighboring trees have not been addressed. We used spatial autocorrelation analyses to determine patterns of defoliation of trees (clustered, dispersed, or random) for plots and for individual trees. From 28% to 47% of plots had significantly clustered defoliation during the 5 years. Plots with clustered defoliation generally had higher mean defoliation per plot and higher deviation of defoliation. At the individual-tree-level, we determined 'hot spot trees' (highly defoliated trees surrounded by other highly defoliated trees) and 'cold spot trees' (lightly defoliated trees surrounded by other lightly defoliated trees) within each plot using local Getis-Ord *Gi** analysis. Results revealed that 11-27 plots had hot spot trees and 27% to 64% of them had mean defoliation < 25%, while plots with 75% to 100% defoliation had either cold spot trees or non-significant spots, which suggested that whether defoliation was high or low enough to be a hot or cold spot depended on the defoliation level of the entire plot. We fitted individual-tree balsam fir defoliation regression models as a function of plot and surrounding tree characteristics (using search radii of 3-5 m). The best model contained plot average balsam fir defoliation and subject tree basal area, and these two variables explained 80% of the variance, which was 2% to 5% higher than the variability explained by the neighboring tree defoliation, over the 3-5 m search radii tested. We concluded that plot-level defoliation and basal area were

adequate for modeling individual tree defoliation, and although clustering of defoliation was evident, larger plots were needed to determine the optimum neighborhood radius for predicting defoliation on an individual. Spatial autocorrelation analysis can serve as an objective way to quantify such ecological patterns.

Keywords: *Choristoneura fumiferana*; annual defoliation; spatial autocorrelation; spatial-temporal patterns; mixed effect models; intertree variance

3.2. Introduction

With advances in spatial analysis theory and methodologies, biologists and ecologists have become more interested in analyzing ecological processes from a spatial point of view. Spatial point pattern analyses can indicate how individuals locate with respect to each other, by testing the complete spatial randomness null hypothesis (Dale and Fortin 2014). Spatial autocorrelation analyses and indices can be applied to detect if the observed value of a variable at one site is dependent on values at neighboring sites (Sokal and Oden 1978; Cliff and Ord 1981). The first law of geography states "Everything is related to everything else, but near things are more related than those distant ones" (Tobler 1970). Spatial autocorrelation analyses are used for ecological processes that are distance-related, including speciation, extinction, dispersal, and synchronous population fluctuations (Augustin et al. 1998; Knapp et al. 2003). Quantitative indices of spatial autocorrelation identify and characterize distribution patterns of objects of interest on the ground, and help to bridge the gap between mechanisms and the behavior of the investigated phenomenon (Ryan et al. 2004; Foster et al. 2013).

Spruce budworm (*Choristoneura fumiferana* Clem.; SBW) outbreaks are a dominant, periodic natural disturbance in eastern Canada (Hardy et al. 1986). SBW larvae consume the new foliage of balsam fir (*Abies balsamea* L. Mill.) and spruce (*Picea* spp. A. Dietr.) trees (Hennigar et al. 2008), often for 10 or more years. This results in tree mortality typically ranging from about 40% to 85%, depending upon defoliation severity, tree species, and tree age (MacLean 1980; MacLean 1984). Baskerville and MacLean (1979) observed that SBW-caused tree mortality in immature balsam fir tended to have a strongly contagious spatial distribution. Mortality ranged from 18% to 80% of the total volume per

ha, and tended to occur in distinct patches spreading over time, with almost all trees in these patches killed (Baskerville and MacLean 1979). Variation in mortality from plot to plot was therefore related to the extent to which such patches occurred in each plot. Spatial variability of mortality has important implications for study of stand vulnerability to SBW attack. If within-stand variability is high and apparently not related to stand characteristics (Baskerville and MacLean 1979), then conventional large scale, between-stand sampling to establish characteristics of vulnerability could lead to statistically significant relationships between mortality and average stand conditions, even when mortality is not functionally related to average conditions (Baskerville and MacLean 1979).

SBW-caused tree mortality is strongly related to cumulative defoliation (Ostaff and MacLean 1989; Erdle and MacLean 1999; Chen et al. 2017), so one might assume that a clustered spatial distribution of dead trees reflects higher annual defoliation in those trees than in surrounding trees. No studies have determined spatial patterns of SBW defoliation among trees, so in this study, we used spatial autocorrelation analyses to determine patterns of tree locations (clustered, dispersed, or random) and of annual defoliation of trees for 5 years in 57 plots in Gaspé, Québec, Canada. In addition to implications on stand vulnerability to insect-caused mortality, spatial pattern of defoliation is important for projecting effects on stand growth reduction and mortality in distance-dependent or tree-list growth models (e.g., Lamb et al. 2018). Relationships between tree mortality and defoliation are non-linear, so use of average defoliation when some trees actually sustain higher levels of defoliation will underestimate impacts. For managing defoliated forests, knowledge of tree-to-tree variability of SBW defoliation can also benefit selection of the

scale and methods of biological insecticide spray treatment decisions to target moderate or high infestation.

Although the spatial interrelationships of SBW defoliation levels of trees within a plot are not well understood, inter-tree variance in defoliation has been observed to be greater than intra-tree variance or variance between plots in a stand (MacLean and Lidstone 1982). Defoliation levels are clearly a function of SBW population factors (Fleming and Van Frankenhuyzen 1992) including oviposition site selection, larval and moth dispersal, and larval survival, but also may be influenced by tree, stand, site, and topographical factors. At the landscape scale, studies determining "epicenters" of historical defoliation based on aerial survey data have concluded that host species or volume had only minor effects on defoliation (Hardy et al. 1983; Gray and MacKinnon 2006; Zhao et al. 2014). Defoliation among plots within stands was typically similar, with low variance (< 5%), especially when SBW populations were high, for both eastern (Ostaff and MacLean 1989) and western SBW (*Choristoneura occidentalis* Freeman; Alfaro et al. 1984). At finer scales, variation of defoliation of branches within a stand was less than that of shoots on a branch (MacLean and Lidstone 1982), because branches from one site tended to have similar defoliation, and distribution of shoot defoliation tended to follow skewed Poisson distributions at both light and severe defoliation levels. Different host species suffer different degrees of defoliation: white spruce (*Picea glauca* Moench Voss), red spruce (*Picea rubens* Sarg.), and black spruce (*Picea mariana* Mill. B.S.P.) had approximately 72%, 41%, and 28% as much defoliation as balsam fir (Hennigar et al. 2008). Non-host species also influence defoliation: balsam fir defoliation was lower when hardwood content increased, thought to be because of greater diversity and populations of SBW parasitoids (Su et al. 1996; Zhang

et al. 2018). In mixed fir-spruce stands, the density of balsam fir increased SBW defoliation severity in fir-dominated and fir-spruce mixed stands, while density of black spruce decreased the infestation in such stands (Bognounou et al. 2017).

There are several possible SBW population-related causes contributing to variable defoliation patterns. The frequency distribution of SBW larvae per tree is positively skewed, with a small proportion of trees hosting populations higher than the average (Morris and Mott 1963). This could result from unequal attraction of female moths to host trees, perhaps based on strength of their stimulus to the olfactory responses of the insect (Morris and Mott 1963), or different exposure to light (Morris 1955). In stands without severe defoliation, SBW egg populations were usually higher on taller and dominant trees that were well exposed to light, while in severely defoliated stands, greater defoliation of exposed trees makes them less attractive to female moths, and correlation between the egg population and tree height became negative (Morris 1955; Morris and Mott 1963). Larval SBW populations were greater in upper crowns because female moths tended to deposit eggs on the peripheral shoots of the crown (Morris 1955; Beckwith and Burnell 1982). Early instar larval dispersal may cause redistribution of larval populations within a stand, but mass dispersal movements only occur in severely infested stands where food supply becomes exhausted (Morris and Mott 1963); such larval dispersal is risky because of high resulting mortality (Miller 1958). Although no direct data exist on between-tree spatial distributions of SBW egg masses or larval populations, spatial clustering of other insects has been observed: both adult and larval cereal leaf beetle (*Oulema melanopus* L.) were in a clustered distribution within 0.4 ha areas in wheat fields (Reay-Jones 2010); and emerald

ash borer (*Agrilus planipennis*) larval populations declined with distance, with about 90% of larvae within 100 m of adults' emergence sites (Mercader et al. 2009).

In this study, we evaluated spatial patterns of 5 years of individual-tree SBW defoliation within 57 plots in Québec. Objectives were to: 1) determine whether spatial locations of tree stems within plots were clustered, dispersed, or random, as a baseline for comparing defoliation patterns; 2) quantify spatial aggregation of defoliation within plots, and detect if there was spatial dependence of individual tree defoliation; and 3) test whether the defoliation level of an individual balsam fir tree can be predicted using plot, tree, and neighboring tree defoliation and other characteristics. We evaluated two predictions: 1) defoliation of individual trees within a plot may be spatially clustered in some cases, i.e., highly defoliated trees will be surrounded by highly defoliated trees, and vice versa for lightly defoliated, as would result from preferential selection of oviposition sites by female moths; and 2) defoliation of an individual balsam fir tree within a plot can be predicted using defoliation level and species composition of surrounding neighborhood trees, combined with plot-level defoliation and species composition.

3.3. Methods

3.3.1. Study area and data collection

The study area was located in the central Gaspé Peninsula region of Québec, a balsam fir-white birch (*Betula papyrifera* Marsh.) mixed eco-district (Belanger et al. 1992). According to aerial forest surveys, light SBW defoliation began in 2012 or 2013 and increased to moderate or severe levels since 2014 (QMRNF 2014). The 57 circular (400 m^2) study plots were established in 2014, with 19 plots about 15 km northwest of Amqui

and 38 plots about 40 km southwest of Causapscal. These are a subset of plots studied by Donovan et al. (2018) and Zhang et al. (2018), and additional description of the sites is available there. Plot locations are shown in Figure 1 of (Donovan et al. 2018). However, we have omitted 18 plots from the 75 plots studied by Donovan et al. (2018) among which 15 plots were harvested from 2016 to 2018, and three plots missed some defoliation data in 2015. A portion of our studied plots were protected by aerial spraying of *Bacillus thuringiensis* biological insecticide to control defoliation each year: 28% of plots in 2014, 42% in 2015, 30% in 2016, 26% in 2017, and 38% in 2018. The plots included a total of 3693 trees, of which 3200 were balsam fir or spruce.

Plot center coordinates were recorded using survey-grade GPS. Attribute data collected for each tree with diameter at breast height (DBH) ≥ 10 cm included species, tree azimuth, distance from plot center, DBH, total tree height, height of the live crown base, crown widths in north, east, south, and west directions, and annual ocular estimates of current defoliation using binoculars. Foliage defoliation was classified into seven classes: 0%–10%, 11%–20%, 21%–40%, 41%–60%, 61%–80%, 81%–99%, and 100%, and the midpoint of each class was used for calculations. To calibrate observers and check accuracy, one mid-crown branch was sampled from each of 8-15 trees per host species per plot (sample sizes based on MacLean and MacKinnon 1998), and defoliation of each of 25 current-year shoots was assessed using the same classes as for ocular defoliation. Defoliation data were collected after SBW feeding ceased each year from 2014 to 2018. Criteria for plot selection and methods of data collection were described in more detail in (Donovan et al. 2018).

3.3.2. Point pattern analyses

We first evaluated whether tree stems within plots were clustered, dispersed, or randomly distributed using a point pattern analysis of average nearest neighborhoods (Ebdon 1985). Liu (2001) reviewed five methods for spatial pattern analysis, and indicated that when the purpose is to analyze the nearest neighbor, Pielou's statistics (Pielou 1959) and Clark and Evans' statistics (used in this study; Clark and Evans 1954) perform the best in avoiding Type II errors, although there was evidence showing that Pielou's statistics (Pielou 1959) have more power for detecting regular and Poisson cluster patterns. The Clark and Evans' statistics calculate the ratio of mean observed distance between all trees and their nearest neighbors, and expected mean nearest neighbor distance for a random arrangement (Clark and Evans 1954):

$$R = \frac{D_O}{D_E}, \quad (1)$$

where $D_O = \frac{\sum_{i=1}^{n} d_i}{n}$ is the observed mean distance between each tree and its nearest neighbor, in which d_i equals the distance between tree i and its nearest neighbor, and n is the number of trees in the plot; $D_E = \frac{0.5}{\sqrt{\frac{n}{A}}}$ is the expected mean distance between each tree and its nearest neighbor given in an ideal random pattern, in which A is the area of minimum enclosing rectangle around all features (set as plot area, 400 m^2, in this study). Generally, if the average distance is less than the average for a hypothetical random distribution ($R < 1$), the distribution is considered to be clustered; if $R > 1$, the distribution tends to be dispersed, and if $R = 1$ the distribution follows a random arrangement. The standard deviation, z-score was calculated as (Clark and Evans 1954):

$$z = \frac{D_0 - D_E}{SE}, \quad (2)$$

where $SE = \frac{0.26136}{\sqrt{\frac{n^2}{A}}}$ is the standard error of the mean, in which the constant is derived from the radius of a circle of unit area and the Poisson probability model. Average nearest neighborhood analyses were done separately on each of the 57 plots, on all host trees, by species (balsam fir, black spruce, and white spruce), and on hardwood trees. Analyses were only applied when number of trees analyzed was ≥ 5 per plot.

3.3.3. Spatial autocorrelation analyses

Two quantitative indices were used to detect and characterize spatial patterns of defoliation at plot and tree levels. First, global Moran's *I* coefficient (Moran 1950) was computed, to produce an overall measure of similarities or dissimilarities between neighboring trees within a plot. It helps to estimate the intensity of spatial dependence for the entire plot and summarizes it with a single value (Sokal et al. 1998; Dale and Fortin 2014). Global Moran's *I* was calculated as:

$$I = \frac{n}{S_0} \cdot \frac{\sum\sum W_{ij} Z_i Z_j}{\sum Z_i^2}, \quad (3)$$

where W_{ij} is the spatial weight between tree i and tree j using an inverse distance weighting method to conceptualize the spatial relationships between trees, such that numerical weights quantified the proximity of pair-wise observations, with closer trees having higher spatial weights (Sokal and Oden 1978). Z_i or Z_j is the deviation of defoliation for tree i or tree j from the mean of all trees in the plot, calculated as $x_i - \bar{x}$ or $x_j - \bar{x}$; $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}$ is the aggregate of all spatial weights and n is the number of trees in the plot. Global Moran's *I* ranges from -1 to 1; $I > 0$ or $I < 0$ corresponds to a

positive or negative spatial correlation, representing spatial clustering or spatial dispersal, respectively, and $I = 0$ means a random distribution (Cliff and Ord 1981). However, global Moran's I is difficult to interpret unless combined with statistical significance tests, and can only be interpreted within the context of the complete spatial randomness hypothesis. Z-score for global Moran's I was computed as (Moran 1950):

$$Z = \frac{I - E(I)}{\sqrt{V(I)}}, \tag{4}$$

where $E(I) = \frac{-1}{n-1}$ is the expectation of global Moran's I under the complete spatial randomness hypothesis; when the sample size tends to be infinite, $E(I)$ is zero. $V(I)$ is the variance, computed as $V(I) = E(I^2) - (E(I))^2$. The analyses omitted cases when all trees in the plot had the same level of defoliation.

Following the global Moran's I analysis, Getis-Ord Gi^* analysis, which is a local spatial autocorrelation statistic, was used to determine defoliation patterns of individual trees within each plot. Getis-Ord Gi^* analysis tests the spatial clustering of high or low values of the measured variable around location i, characterizing the internal clustering within each plot by determining the extent to which defoliation of a given tree is surrounded by a cluster of trees with either high or low values (Getis and Ord 1992; Ord and Getis 1995). Gi^* statistics proportionally compare the local sum for tree i and its neighbors to the sum of all trees, which is computed as (Getis and Ord 1992):

$$G_i^* = \frac{\sum_j W_{ij} x_j - \bar{X} \sum_j W_{ij}}{S \sqrt{\frac{n \sum_j W_{ij}^2 - (\sum_j W_{ij})^2}{n-1}}}, \tag{5}$$

where i is the subject tree, x_j is the defoliation value for one of the neighboring trees j, W_{ij} is the spatial weight between subject tree i and neighboring tree j determined by the fixed

distance band method (i.e., neighboring trees within 5 m were set with the same weight), $\bar{X}$ is the average defoliation value of the entire plot, *n* is the total number of trees in the plot, and *S* is the standard deviation of the entire plot, given by $S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2}$. *Gi** is calculated as sum of the differences between individual values and the mean of all individuals. Therefore, *Gi** is a standard normal distribution *z*-score. A statistically significant *Gi** results when difference between calculated local sum and expected local sum is too large to be a random chance. High and positive values of *Gi** indicate 'hot spot trees', severely defoliated trees that are surrounded by severely defoliated trees, whereas low and negative values indicate 'cold spot trees', lightly defoliated trees surrounded by lightly defoliated trees.

Both size and shape of the study plot affect the ability of the *Gi** statistics to accurately estimate the type and significance of a spatial pattern, because objects near the edge would have fewer neighbors than those in the middle of the plot. This is known as edge effect (Monserud and Ek 1974). If the plot does not contain the entire spatial process under study, as was the case in this research, it is important to consider edge effects to increase stability and power of statistics resulting from sampling (Sui and Hugill 2002). In this study, to avoid complicated edge-correcting procedures (Ripley 1979), we used an inner buffer zone in each plot to compensate for edge effects. A search radius of 5 m was used and a corresponding central circular subplot of 6 m was selected inside each plot. Trees in the border area were excluded from the Getis-Ord *Gi** analysis. All spatial autocorrelation analyses were done using ESRI ArcMap 10.4 (ESRI, Redlands, CA, USA).

Both global Moran's *I* and local Getis-Ord *Gi** spatial autocorrelation analyses were conducted separately on each of the 57 sample plots, and results of these and point pattern tree location analyses were presented as the number and percentage of plots that had significantly clustered results. With such a wide range of defoliation conditions purposefully sampled across plots and years (ranging from < 5% to > 90% current annual defoliation), it was highly unlikely that there would be a single clustered or not clustered answer for all plots.

3.3.4. Tree defoliation regression model

To determine whether defoliation levels of individual balsam fir can be predicted by plot or surrounding-tree characteristics, nonlinear mixed effect regression models were fitted, with plots nested within years as random effects. Current-year defoliation of balsam fir was the response variable. Plot-level predictors included average defoliation and % basal area of each host species and hardwoods per plot. Tree-level predictors included basal area of the subject balsam fir, defoliation of the subject balsam fir in the previous year, average defoliation of surrounding (3, 4, and 5 m search radii) trees (calculated for balsam fir, black spruce, white spruce, and for all host trees), total basal area of surrounding trees (balsam fir, hardwood, spruce trees, and for all host trees), and total basal area of surrounding trees with basal area higher than the subject balsam fir, representing relative social status of the subject tree (calculated for balsam fir, spruce, and for all host trees). Defoliation of the subject balsam fir in the previous year acted as a temporal variable, and those neighborhood-related predictors at the tree-level acted as spatial variables because they represented defoliation levels and species composition around the subject tree. We

also considered the possible impact of insecticide spraying by including a dummy variable representing the subject balsam fir being sprayed or not in the corresponding year. All subject trees were balsam fir located in the central 6 m radius subplots, and effects were tested using surrounding 3 m, 4 m, or 5 m search radii with at least one neighbor in the neighborhood.

Gradient boosting is a machine learning technique, which identifies the performance of decision trees by using gradients in the loss function (i.e., a measurement of the goodness of coefficients in the model) in a sequential fashion (Friedman 2001). Gradient Boosting Machine analysis (Ridgeway 2007) was used to determine the most important variables using the "caret" package (Kuhn 2008) within R (R Development Team 2018). One of the important features of Gradient Boosting Machine analysis is variable importance, which was summarized as the ranked variables based on their relative influence (importance) in training the model. Relative influence was computed based on how often a variable was selected for splitting, weighted by the squared improvement to the model in each split (Elith et al. 2008). Basal area of the subject tree was included in each tested model because it is related to foliage amount and to avoid having zero variance of fitted individual balsam fir defoliation values within a plot when it was predicted solely using plot-level variables. Except for the subject tree basal area, relatively important predictors were added into candidate models in sequence according to their ranks by Gradient Boosting Machine analysis. The contribution of newly-added predictors was determined by likelihood ratio tests, and based on this, whether the models should include the added predictors was evaluated. Highly correlated variables ($r \geq 0.7$) were avoided in the candidate models. Beta regression was used because the response variable, defoliation of the subject balsam fir,

was constrained between 0 and 1 (0-100%). A logit-link function was used to set up linear relationships between the response and predictor variables. Likelihood ratio tests were also used to determine whether the random effects from plots or years were significant, by comparing a mixed effect model and a fixed effect model as a null model. The significance of individual predictors in the models was assessed by *t* tests. The “nlme” and “gnls” functions in the "nlme" package (Lindstrom and Bates 1990) were used in fitting mixed effect models and fixed effect models, respectively.

Assumptions of normality and homoscedasticity of residuals were evaluated using residual plots. Goodness-of-fit of candidate models was assessed and compared by root mean square error (RMSE), adjusted R^2, and mean bias (predicted-observed).

3.4. Results

3.4.1. Stand and plot characteristics

Characteristics of the 57 plots in the sampled stands ranged from 13.5-18.0 m height, 14.9-22.2 cm DBH, 1075-1919 stem ha^{-1} density, and 34.9-47.7 m^2 ha^{-1} basal area (Table 3.1). Basal area of each host species in the stands was 6-98% balsam fir, 0-70% black spruce, and 0-65% white spruce (Table 3.1). Plot species composition averaged 64% balsam fir, 19% black spruce, 10% white spruce, 2% other softwood species, and 5% hardwood species (Table 3.1). Annual defoliation in the 5 years from 2014 to 2018 averaged 34%, 51%, 28%, 38%, and 32% by ocular estimation, and 45%, 52%, 34%, 54%, and 38% by shoot defoliation estimation on sampled branches. Plot locations were specifically selected in 2014 to represent the full range of defoliation from < 10% to 90-100%, and insecticide sprayed plots were the only way to obtain low defoliation levels.

3.4.2. Spatial patterns of tree stems within plots

Average nearest neighborhood analyses for all host species within each plot showed that 35 plots (61%) had randomly distributed host trees, and the remaining 22 plots (39%) had significantly dispersed host trees ($\alpha = 0.05$; Table 3.2). No plots had significantly clustered host tree stem locations. The average nearest neighborhood analysis for each host species and hardwoods also showed that balsam fir, black spruce, white spruce, and hardwoods were distributed randomly or dispersed in most plots, and few (0-5%) plots had significantly clustered tree stems by species (Table 3.2). Generally, the null hypothesis that there is no difference between observed and random nearest neighbor values was not rejected for almost all plots, either by species or for all host species combined. The results of these average nearest neighborhood analyses set the baseline tree spatial distributions for the following spatial autocorrelation analyses of defoliation levels.

3.4.3. Spatial patterns of current year defoliation for plots

Global Moran's *I* analyses results for all host species showed that each year from 2014 to 2018, 47%, 28%, 35%, 30%, and 33% of plots, respectively, had significantly clustered intertree defoliation patterns ($\alpha = 0.05$; Table 3.3). This indicates that in these plots, defoliated trees had a strong tendency to be located closer to trees with similar defoliation values. These results therefore differed substantially from that for the spatial locations of trees, in which no plots had clustered tree spatial distributions (Table 3.2). In roughly one-third to one-half of the plot-years, defoliation of trees was clustered. A total of 45 plots (79%) exhibited clustered defoliation in at least 1 of the 5 sampled years. For balsam fir, 11-33% of plots had clustered defoliation distribution, somewhat fewer than for all host

species combined (Table 3.3). With relatively few spruce trees present, only 2-5% and 0-4% of plots had clustered defoliation on black spruce and white spruce (Tables 3.1, 3.3). A dispersed defoliation pattern occurred only for black spruce in one plot, with no dispersed defoliation patterns for all host trees or by species.

We used box plots to compare the distributions of total basal area, average annual plot defoliation, and standard deviation of individual tree defoliation between plots with clustered and non-clustered defoliation on all host trees (Figure 3.1). Plots were divided into 25% balsam fir plot basal area classes (x-axis) in Figure 3.1 because the analyzed plots varied widely in species composition. With low balsam fir content (< 25% basal area), 18 plot-years with non-clustered defoliation had significantly higher basal area than 12 plot-years with clustered defoliation (mean 46 versus 41 m^2 ha^{-1}; Figure 3.1a). Overall, annual plot defoliation levels increased with the proportion of balsam fir in plots (Figure 3.1b). Plots with clustered defoliation had higher defoliation in all balsam fir % basal area classes than plots with non-clustered defoliation, significantly different for 50-75% basal area fir plots (43% versus 33% defoliation in clustered and non-clustered plots; Figure 3.1b). Plots with clustered defoliation consistently had higher standard deviations of tree defoliation than plots with non-clustered defoliation, with significant differences of 9.1% and 6.4% defoliation for plots with 0-25% and 50-75% balsam fir (Figure 3.1c).

3.4.4. Hot spot and cold spot trees within plots

In Figure 3.2, we integrated results of the global Moran's *I* and Getis-Ord *Gi** statistics for all analyzed plots in all years. We ordered the stand_plot number by annual plot

defoliation each year, in order to visually convey that although defoliation was clustered in some plots (shown by * above each bar), how and whether hot or cold spots could be detected showed a tendency to be related to defoliation level. Hot spot trees (red bars in Figure 3.2) tended to be in lightly defoliated plots, and cold spot trees (blue bars) tended to be in highly defoliated plots. Across the 5 sampled years, minimum 42 and maximum 80 hot spot trees (highly defoliated trees surrounded by other highly defoliated trees), and minimum 44 and maximum 59 cold spot trees (lightly defoliated trees surrounded by other lightly defoliated trees) were detected within the 6m-radius subplots (Figure 3.2). Plots with hot or cold spot trees did not necessarily have clustered defoliation (* in Figure 3.2), although plots with significantly clustered defoliation did tend to have hot or cold spot trees: 25-45%, 10-31%, 13-35%, and 10-36% of plots with clustered defoliation had hot spot trees, cold spot trees, both hot and cold spot trees, and non-significant spots, respectively, across the 5 years (Figure 3.2).

There was a tendency for highly defoliated plots to have more cold spot trees, and lightly defoliated plots to have more hot spot trees (Figure 3.2). Over the 5 sampled years, plots with 0-25% defoliation had 0-4 cold spots (0.3 on average) and 0-8 hot spots (1.0 on average), while plots with 75-100% defoliation had 0-12 cold spots (1.7 on average) and virtually no hot spots. Over the 5 sampled years, 27, 11, 26, 18, and 26 plots had hot spot trees (including those with both hot and cold spot trees), while 13 (48%), 7 (64%), 16 (62%), 7 (39%), and 7 (27%) of them were 0-25% defoliated (Figure 3.3). Plots with 75-100% defoliation had either cold spot trees or non-significant spots within them, except one plot in 2017 with > 75% defoliation that had one hot spot tree (Figure 3.3d). That plot

experienced 75% annual defoliation, with 100% balsam fir composition, and had only one balsam fir detected as a hot spot tree in 2017.

Stem maps of three example plots were selected to represent generally low, moderate, and high defoliation that had hot and cold spot trees (Figures 3.4, 3.5). Plot 1_01 had low defoliation (5-17%) in all years from 2014-2018 (Figure 3.4), and had 1-5 hot spot trees in 4 years, but none in 2017 (Figure 3.5). Interestingly, locations of hot spot trees were the same for only three of the 11 trees, in 2 years (2015-2016); otherwise they differed. Mean defoliation of trees in plot 12_02 ranged from 29-74% over the 5 years, and covered defoliation classes from 0-20% to 81-100% (Figure 3.4). There were no hot spot trees and only 6 cold spot trees in 2 years in this plot (Figure 3.5). Plot 3_01 had very high defoliation (85%) in 3 years, and moderate (31% and 52%) defoliation in 2014 and 2018 (Figure 3.4), resulting in many cold spot trees in all years and 2 hot spot trees in the first year only (Figure 3.5). It is noteworthy that at moderate defoliation levels, variation among trees was high: with 31% mean defoliation in 2018 in plot 3_01, trees with all five 20% defoliation classes occurred in the plot (Figure 3.4).

3.4.5. Prediction of subject tree balsam fir defoliation using regression models

Results of Gradient Boosting Machine analysis consistently showed that plot average annual defoliation of balsam fir was the most important predictor for all three neighborhood search radii, with relative influence of 73%, 57%, and 48% for 3 m, 4 m, and 5 m search radii, respectively (Figure 3.6). Given the variability in defoliation among trees within a plot (e.g., Figure 3.4), it was surprising to us that plot average defoliation was superior to local within-plot defoliation for the three radii circles in predicting defoliation of a single

tree. Average annual defoliation of surrounding balsam fir had the highest relative influence of all tree-level predictors, at 15%, 30%, and 41% for 3 m, 4 m, and 5 m search radii (Figure 3.6). Table 3.4 lists all plot- and tree-level predictors investigated using Gradient Boosting Machine analysis. Correlation analysis of the predictor variables suggested that plot average balsam fir defoliation (i.e., mean of all trees in the plot) was highly correlated ($r > 0.9$) with average defoliation of balsam fir within its neighborhood (i.e., within 3 m, 4 m, or 5 m circles). Therefore, average defoliation of balsam fir at the plot- and tree-level were not both included in the same candidate model. Basal area, representing subject tree size, was included in each model. Total basal area of host trees having higher basal area than the subject tree within 3 m and 5 m, and total basal area of host trees within 4 m, which ranked as the sixth most important predictors, were also tested in candidate models. Spraying and defoliation in the previous year variables had little influence (0.3% at most) in the Gradient Boosting Machine analysis. The remaining predictors, including spray and defoliation in the previous year, were dropped in the following process.

Plots nested within years was added as a random effect variable, but likelihood ratio tests suggested that it had little influence, in comparing models with mixed effects versus fixed effects as a null model ($p = 0.9$). This meant that the variance associated with plot-nested-in-year groups could occur by chance. Therefore, the random effects were removed from candidate models, and fit statistics of models with only fixed effects were reported (Table 3.5).

Results of models to predict subject tree defoliation as a function of plot and neighboring tree variables showed that Model 1 based on plot average balsam fir

defoliation and subject tree basal area explained 80% of the total variance in the response variable, with RMSE of 14.1% and bias of 2.8% (Table 3.5). Variance explained by plot average balsam fir defoliation was slightly higher than that explained by neighboring average balsam fir defoliation within 4 m or 5 m (79% and 78%), which were higher than 3 m (75%; Table 3.5). Coefficients ($p < 0.01$) in these four candidate models (Models 1, 2, 4, and 6) were all positive, indicating that higher subject balsam fir defoliation occurred with higher subject tree basal area, higher plot average defoliation, and higher neighboring balsam fir average defoliation. The other two predictors, total basal area of host trees having higher basal area than the subject tree within 3 m (Model 3) and total basal area of host trees within 4 m (Model 5) were significant ($p < 0.05$), and had slightly better R^2, RMSE, and bias, compared to Model 1. This suggested that including neighboring host tree basal area can slightly improve performance compared to a model that included only plot average balsam fir defoliation.

3.5. Discussion

3.5.1. Is defoliation of individual trees clustered?

Generally, defoliation was clustered in some plots and some years, depending on defoliation levels. Plots with clustered defoliation consistently had higher annual defoliation than plots with non-clustered defoliation. Spatial locations of fir and spruce trees within plots were randomly distributed in 61% of plots and dispersed in the remaining 39% of plots; none were clustered. In contrast, an average of 35% of plots had significantly clustered defoliation distributions of balsam fir and spruce, with amount clustered varying

from 28-47% among years. Plots with clustered defoliation had larger standard deviation of defoliation among trees, reflecting higher variability in defoliation within the plots.

The added value of this study is the first demonstration that there can be spatially clustered defoliation patterns among trees, even though the baseline tree locations are not spatially clustered. Although SBW populations on each tree were not measured, higher defoliation must result from higher SBW populations or higher larval survival. Clustered defoliation within plots probably results from SBW population processes, primarily moth oviposition selection and larval dispersal. SBW egg populations are usually higher on taller and dominant trees that are well exposed to light, but severe defoliation may make such trees less attractive to female moths, and the correlation between egg population and tree height becomes negative (Morris 1955; Morris and Mott 1963). Larval dispersal leads to tree-to-tree redistribution of SBW, which can result in reduced tree-to-tree correlation (i.e., more even SBW pressure), because such dispersal occurs rather randomly in direction especially at a small scale (Beckwith and Burnell 1982). Not much air movement is needed for larval dispersal due to their light body weight (Batzer 1973). However, a large proportion of the SBW population must be engaged in dispersal processes to contribute to a noticeable change in intertree population distribution, and such mass movements only occur in highly defoliated stands with high competition among feeding larvae (Morris and Mott 1963). Also, SBW larvae tend to avoid risky dispersal behavior because of the high probability of mortality during dispersal (Miller 1958). Given that current annual defoliation exceeded 80% in only 8.5% of our 260 plot-years sampled (and most of those were in 2015), larval dispersal probably had minor effects on intertree defoliation distribution in this study.

3.5.2. Interpretation of local hot and cold spot trees

Conceptually, Getis-Ord *Gi** is the ratio of average local defoliation over average global (plot in our case) defoliation. It indicates hot spot (highly defoliated trees surrounded by other highly defoliated) and cold spot (lightly defoliated trees surrounded by other lightly defoliated) trees. Over all plots and years, 28% of plot-years had hot spot trees, 18% had cold spot trees, 12% had both, and 42% had none; in other words, there was significant spatial variability in defoliation in 58% of cases. Whether defoliation is high enough to be a hot spot or low enough to be a cold spot depends on the general defoliation level of the plot. SBW outbreak spatial-temporal dynamics typically begin with patchy low-level but increasing defoliation for several years, observable on individual branches initially and then on trees. It typically takes several years for SBW populations to build sufficiently to result in widespread severe defoliation (> 70% of current year foliage on most trees) across stands and regions. Defoliation is most variable among trees and plots at moderate (30-70% of current year foliage) levels; when low or very high, defoliation tends to be consistent (MacLean and Lidstone 1982). Plots with severe defoliation tended to have cold spot trees, and plots with light defoliation tended to have hot spot trees. This is because a target highly defoliated tree surrounded by other highly defoliated trees will have a higher value of *Gi** (ratio of average local defoliation over average global defoliation) if it is located in a plot with overall low defoliation, while it is more likely to be non-significant if it is located in a highly defoliated plot. A similar principle applies for cold spot trees.

Highly significant global spatial autocorrelation can lead to overly liberal local spatial autocorrelation, which means significance tests of the local-spatial-autocorrelation coefficients can reject the null hypothesis excessively when a global spatial autocorrelation

occurs (Sokal et al. 1998). Although the local spatial statistic used in this study, Getis-Ord *Gi**, does not need extra significance tests, values are still sensitive to the overall spatial structure of defoliation in plots (Dale and Fortin 2014). Other statistics have been proposed to address such issues (Kabos and Csillag 2002; Boots 2003), but a local statistic avoiding influence from or accounting for global spatial autocorrelation is still needed (Getis 2008). As suggested by Sokal et al. (1998), the major application of local spatial autocorrelation tests should be exploratory instead of significance testing. Our main conclusion here is that in about one-third to one-half of plots over 5 years, clustered defoliation patterns occurred, at either higher (hot spot) or lower (cold spot) levels than the plot average. Therefore the spatially contagious SBW-caused tree mortality observed by Baskerville and MacLean (1979) within a uniform immature balsam fir stand, occurring at a scale smaller than the 400 m^2 plots, may have resulted from higher defoliation in some trees. Over the longer term, mortality creating such 'holes' in stands is also probably exacerbated by windthrow disturbance (Taylor and MacLean 2007, 2009).

3.5.3. Prediction of subject balsam fir defoliation

Average defoliation for the full plot combined with subject tree basal area explained 80% of the variability of subject balsam fir defoliation, which was 2-5% higher than the variability explained by the neighboring tree defoliation, over the 3 m, 4 m, and 5 m search radii tested. However, the relative influence of neighboring tree defoliation on subject tree defoliation increased as search radius increased, from 15% to 30% to 41% as search radius increased from 3 m to 4 m to 5 m. At a 5 m search radius, the relative influence of neighboring tree defoliation was nearly as high as plot average defoliation. This suggested

that a neighborhood search radius larger than 5 m, and therefore larger overall plot size, was needed, and it might be superior to average plot defoliation for individual-tree-defoliation prediction.

There was wide variation in defoliation among trees in the sampled plots, and although plot mean defoliation was the strongest predictor of subject tree defoliation, inferring similar average defoliation levels for all trees in a plot creates a problem when inputting defoliation values into stand growth models to predict effects of defoliation on growth and mortality. Model driving relationships between growth reduction, mortality, and defoliation are non-linear, such that trees with the highest defoliation will sustain substantially higher rates of growth reduction and mortality than the mean. Recognizing that a portion of trees are in hot or cold spots, with clusters of higher or lower defoliation, implies that model predictions will be more accurate using input distributions of defoliation for tree-list models and spatial defoliation distributions for distance-dependent growth models.

For these primarily balsam fir plots, species composition at either the plot-level or tree-level was not a meaningful predictor of individual balsam fir defoliation. The stands sampled had 51-98% fir in all but four plots, which had 77-92% black and white spruce. The four spruce plots exhibited hot and cold spot trees in 75% of the 20 plot-years sampled. Although hardwood content has been shown to influence SBW defoliation (Su et al. 1996; Zhang et al. 2018), hardwoods comprised only an average of 6% of basal area in these plots (range 1-20%). This gave too few samples to test neighborhood effects of hardwoods on defoliation. Although insecticide spraying was an important predictor when forecasting plot-level annual defoliation (Donovan et al. 2018), it had little impact on tree-level annual

defoliation in this study, because the effects of spraying were already reflected by the plot mean defoliation variable. Defoliation of subject trees in the previous year also showed little influence, suggesting that individual-tree defoliation can vary considerably from year to year, as a function of SBW population level each year and insecticide treatments.

3.6. Conclusions

We evaluated neighborhood effects on individual-tree SBW defoliation, including the detection of defoliation clustering by spatial autocorrelation analysis, and the prediction of subject balsam fir defoliation by mixed effect regression. Key messages from the results include:

1. Including all host species, 47%, 28%, 35%, 30%, and 33% of plots showed significantly clustered defoliation patterns from 2014 to 2018. Plots with clustered defoliation tended to have higher and less uniform defoliation among trees. Results suggested that spatial defoliation patterns resulted from uneven SBW pressure on trees, perhaps from oviposition site selection.
2. Plots with severe defoliation generally tended to exhibit cold spot trees, and plots with light defoliation tended to have hot spot trees, because whether defoliation was high or low enough to be a hot or cold spot depended on the defoliation level of the entire plot.
3. Plot-level average defoliation combined with subject tree basal area explained 80% of the variability of subject balsam fir defoliation, which was 2-5% higher than variability explained by the neighboring tree defoliation.
4. Spatial variability of defoliation decreased with larger radius neighborhoods from 3 m to 5 m, suggesting that a neighborhood search radius larger than 5 m (and thus plot

sizes larger than 400 m^2 (11.3 m radius) to deal with edge effects) may provide better predictions of subject balsam fir defoliation.

5. For these primarily balsam fir plots, species composition at both plot- and tree-levels were not significant predictors of individual balsam fir defoliation.

Spatial autocorrelation analysis is a useful means to describe and quantify spatial-temporal, ecological patterns of insect defoliation. For managing defoliated forests, knowledge of tree-to-tree variability of SBW defoliation can benefit selection of the scale and methods of biological insecticide spray treatment decisions to target moderate or high infestation. Our results showed that in spite of within-plot variability, there was far more plot-to-plot variability in defoliation. It indicated that a large number of plots is preferred rather than fewer larger plots for sampling and modeling. On the other hand, understanding spatial variability among defoliation levels of trees can help improve defoliation inputs into distance-dependent or tree-list stand growth models (e.g., Lamb et al. 2018). If only the averaged plot defoliation is used as inputs to growth and yield models for all trees in a stand, it will underestimate growth reduction, and especially mortality for those trees suffering from higher defoliation levels (i.e., clusters). Adopting a distribution of defoliation levels of trees as model inputs rather than merely the plot average values would likely result in more accurate forecasts of SBW impacts.

3.7. Acknowledgments

This research was funded by the Atlantic Innovation Fund project “Spruce Budworm Early Intervention Strategy,” grant number 203544 to D.A.M. The Atlantic Innovation Fund was funded by the Atlantic Canada Opportunities Agency. We acknowledge project

support from the Healthy Forest Partnership. We appreciate the excellent work of all field assistants involved in the data collection: Shawn Donovan, Sean Lamb, Rebecca Landry, Bo Zhang, Maggie Brewer, Jessica Cormier, Olivia Doran, David Alton, and Kerstin Trainor.

3.8. References

Alfaro, R.I., Shore, T.L., and Wegwitz, E. 1984. Defoliation and mortality caused by western spruce budworm: variability in a Douglas-fir stand. J. Entomol. Soc. B.C. **81**: 33-38.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. 1998. The role of simulation in modelling spatially correlated data. Environmetrics **9**: 175-196.

Baskerville, G.L. and MacLean, D.A. 1979. Budworm-caused mortality and 20-year recovery in immature balsam fir stands. Can. For. Serv., Fredericton, NB, Canada. Inf. Rep. M-X-102.

Batzer, H.O. 1973. Net effect of spruce budworm defoliation on mortality and growth of balsam fir. J. For. **71**: 34-37.

Beckwith, R.C. and Burnell, D.G. 1982. Spring larval dispersal of the western spruce budworm (Lepidoptera: Tortricidae) in north-central Washington. Environ. Entomol. **11**: 828-832.

Belanger, L., Bergeron, Y., and Camire, C. 1992. Ecological land survey in Québec. For. Chron. **68**(1): 42-52.

Bognounou, F., Grandpre, L.D., Pureswaran, D.S., and Kneeshaw, D. 2017. Temporal variation in plant neighborhood effects on the defoliation of primary and secondary hosts by an insect pest. Ecosphere **8**(3): e01759.

Boots, B. 2003. Developing local measures of spatial association for categorical data. J. Geogr. Syst. **5**: 139-160.

Chen, C., Weiskittel, A., Bataineh, M., and MacLean, D.A. 2017. Even low levels of spruce budworm defoliation affect mortality and ingrowth but net growth is more driven by competition. Can. J. For. Res. **47**(11): 1546-1556.

Clark, P.J. and Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology **35**(4): 445-453.

Cliff, A.D. and Ord, J.K. 1981. Spatial processes: models & applications. Pion Ltd., London, UK.

Dale, M.R. and Fortin, M.J. 2014. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Donovan, S.D., MacLean, D.A., Kershaw, J.A., and Lavigne, M.B. 2018. Quantification of forest canopy changes caused by spruce budworm defoliation using digital hemispherical imagery. Agric. For. Meteorol. **262**: 89-99.

Ebdon, D. 1985. Statistics in geography: a practical approach. Basil Blackwell Ltd., New York, NY, US.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. **77**: 802-813.

Erdle, T.A. and MacLean, D.A. 1999. Stand growth model calibration for use in forest pest impact assessment. For. Chron. **75**: 141-152.

Fleming, R.A. and Van Frankenhuyzen, K. 1992. Forecasting the efficacy of operational *Bacillus thuringiensis* Berliner applications against spruce budworm, *Choristoneura fumiferana* Clemens (Lepidoptera: Tortricidae), using dose ingestion data: initial models. Can. Entomol. **124**(6): 1101-1113.

Foster, J.R., Townsend, P.A., and Mladenoff, D.J. 2013. Spatial dynamics of a gypsy moth defoliation outbreak and dependence on habitat characteristics. Landsc. Ecol. **28**: 1307-1320.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5): 1189-1232.

Getis, A. 2008. A history of the concept of spatial autocorrelation: a geographer's perspective. Geogr. Anal. **40**: 297-309.

Getis, A. and Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. **24**(3): 189-206.

Gray, D.R. and MacKinnon, W.E. 2006. Outbreak patterns of the spruce budworm and their impacts in Canada. For. Chron. **82**(4): 550-561.

Hardy, Y., Mainville, M., and Schmitt, D.M. 1986. Spruce budworms handbook: an atlas of spruce budworm defoliation in eastern North America, 1938-1980. U.S. Department of Agriculture, Forest Service, Cooperative State Research Service, Washington, DC, USA.

Hardy, Y.J., Lafond, A., and Hamel, L. 1983. The epidemiology of the current spruce budworm outbreak in Québec. For. Sci. **29**(4): 715-725.

Hennigar, C.R., MacLean, D.A., Quiring, D.T., and Kershaw, J.A. 2008. Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. For. Sci. **54**(2): 158-166.

Kabos, S. and Csillag, F. 2002. The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. Comput. Geosci. **28**: 901-910.

Knapp, R.A., Matthews, K.R., Preisler, H.K., and Jellison, R. 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. Ecol. Appl. **13**: 1069-1082.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**(5): 1-26.

Lamb, S.M., MacLean, D.A., Hennigar, C.R., and Pitt, D.G. 2018. Forecasting forest inventory using imputed tree lists for LiDAR grid cells and a tree-list growth model. Forests **9**: 167.

Lindstrom, M.J. and Bates, D.M. 1990. Nonlinear mixed effects models for repeated measures data. Biometrics **46**(3): 673-687.

Liu, C. 2001. A comparison of five distance-based methods for spatial pattern analysis. J. Veg. Sci. **12**(3): 411-416.

MacLean, D.A. 1980. Vulnerability of fir-spruce stands during uncontrolled spruce budworm outbreaks: a review and discussion. For. Chron. **56**: 213-221.

MacLean, D.A. 1984. Effects of spruce budworm outbreaks on the productivity and stability of balsam fir forests. For. Chron. **60**: 273-279.

MacLean, D.A. and Lidstone, R.G. 1982. Defoliation by spruce budworm: estimation by ocular and shoot-count methods and variability among branches, trees, and stands. Can. J. For. Res. **12**(3): 582-594.

MacLean, D.A. and MacKinnon, W.E. 1998. Sample sizes required to estimate defoliation of spruce budworm. North. J. Appl. For. **15**(3): 135-140.

Mercader, R.J., Siegert, N.W., Liebhold, A.M., and Mccullough, D.G. 2009. Dispersal of the emerald ash borer, Agrilus planipennis , in newly-colonized sites. Agric. For. Entomol. **11**: 421-424.

Miller, C.A. 1958. The measurement of spruce budworm populations and mortality during the first and second larval instars. Can. J. Zool. **36**: 409-422.

Monserud, R.A. and Ek, A.R. 1974. Plot edge bias in forest stand growth simulation models. Can. J. For. Res. **4**(4): 419-423.

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. Biometrika **37**: 17-23.

Morris, R.F. 1955. The development of sampling techniques for forest insect defoliators, with particular reference to the spruce budworm. Can. J. Zool. **33**(4): 226-294.

Morris, R.F. and Mott, D.G. 1963. Dispersal and the spruce budworm. *In* The dynamics of epidemic spruce budworm populations. *Edited by* R.F. Morris. Memoirs of the Entomological Society of Canada, Ottawa, ON, Canada. pp. 180-189.

Ord, J.K. and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. **27**(4): 286-305.

Ostaff, D.P. and MacLean, D.A. 1989. Spruce budworm populations, defoliation, and changes in stand condition during an uncontrolled spruce budworm outbreak on Cape Breton Island, Nova Scotia. Can. J. For. Res. **19**(9): 1077-1086.

Pielou, E.C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. J. Ecol. **47**(3): 607-613.

QMRNF. 2014. Aires Infestées par la Tordeuse des Bourgeons de l'épinette au Québec en 2014. Québec Ministère des Ressources Naturelles et de la Faune. Available from http://www.mffp.gouv.qc.ca/publications/forets/fimaq/insectes/tordeuse/TBE_2014_P.pdf [accessed 30 October 2017].

R Development Core Team. 2018. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Reay-Jones, F.P.F. 2010. Spatial distribution of the cereal leaf beetle (Coleoptera: Chrysomelidae) in wheat. Environ. Entomol. **39**: 1943-1952.

Ridgeway, G. 2007. Generalized boosted models: a guide to the GBM package. R Package Vignette.

Ripley, B.D. 1979. Tests of "randomness" for spatial point patterns. J. Royal Stat. Soc. **41**(3): 368-374.

Ryan, P.A., Lyons, S.A., Alsemgeest, D., Thomas, P., and Kay, B.H. 2004. Spatial statistical analysis of adult mosquito (Diptera: Culicidae) counts: an example using light trap data, in Redland Shire, southeastern Queensland, Australia. J. Med. Entomol. **41**(6): 1143-1156.

Sokal, R.R. and Oden, N.L. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10**: 199-228.

Sokal, R.R.S., Oden, N.L., and Thomson, B.A. 1998. Local spatial autocorrelation in biological variables. Biol. J. Linn. Soc. **65**: 41-62.

Su, Q., MacLean, D.A., and Needham, T.D. 1996. The influence of hardwood content on balsam fir defoliation by spruce budworm. Can. J. For. Res. **26**(9): 1620-1628.

Sui, D.Z. and Hugill, P.J. 2002. A GIS-based spatial analysis on neighborhood effects and voter turn-out: a case study in College Station, Texas. Political Geogr. **21**: 159-173.

Taylor, S.L. and MacLean, D.A. 2007. Spatiotemporal patterns of mortality in declining balsam fir and spruce stands. For. Ecol. Manage. **253**: 188-201.

Taylor, S.L. and MacLean, D.A. 2009. Legacy of insect defoliators: increased wind-related mortality two decades after a spruce budworm outbreak. For. Sci. **55**(3): 256-267.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **46**: 234-240.

Zhang, B., MacLean, D.A., Johns, R.C., and Eveleigh, E.S. 2018. Effects of hardwood content on balsam fir defoliation during the building phase of a spruce budworm outbreak. Forests **9**: 530.

Zhao, K., MacLean, D.A., and Hennigar, C.R. 2014. Spatial variability of spruce budworm defoliation at different scales. For. Ecol. Manage. **328**: 10-19.

Table 3.1: Summary of mean ($\overline{X}$) and standard deviation (σ) characteristics per stand of 57 plots located in 19 stands near Amqui and Causapscal, Québec, Canada.

Stand no.	No. plots[1]	Density (stem ha^{-1})		DBH[2] (cm)		Height (m)		Basal area (m^2 ha^{-1})		Species composition[3] (% basal area)				
		$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ	BF	BS	WS	HW	OSW
1	4	1919	289	15.0	1.1	13.7	0.9	38.8	3.8	54	30	4	6	7
2	5	1775	417	17.0	1.4	16.5	1.0	42.8	4.3	51	42	3	3	
3	2	1913	636	16.0	1.4	16.0	0.9	41.8	6.7	83	1	11	5	
4	1	1200		20.1		18.0		39.8		6	27	65	2	
5	2	1825	159	15.8	0.1	16.1	0.2	38.1	1.7	98	1		1	
6	5	1600	329	16.5	2.2	14.8	2.0	38.7	4.0	77	7	4	7	7
7	1	1950		14.9		14.8		38.3		78			20	2
8	2	1800	159	15.8	0.4	15.5	0.6	38.1	1.7	70	20	5	6	
9	3	1808	426	16.7	2.6	16.2	1.3	44.2	4.2	86		2	11	1
10	3	1475	275	17.2	0.9	15.9	0.5	36.3	2.7	22	70	7	1	
11	2	1238	106	18.1	1.0	16.5	0.7	34.9	1.8	54	41	2	2	
12	5	1775	191	17.0	0.8	17.3	1.0	43.5	5.7	50	38	9	2	
13	5	1250	317	20.0	1.7	18.4	1.1	42.2	5.5	57		41	2	
14	4	1844	236	16.5	0.8	16.2	0.6	42.4	4.5	90			10	
15	5	1730	224	16.7	0.4	15.2	0.7	40.3	5.3	69	28	2	1	
16	2	1113	177	22.2	1.1	19.2	0.2	47.7	2.7	55		43	2	
17	2	1075	159	20.3	0.9	16.1	0.2	40.0	2.0	71		11	6	15
18	2	1325	265	17.5	0.3	13.5	0.6	44.0	4.5	80			20	
19	2	1763	636	17.3	2.9	14.2	3.0	51.3	8.5	59		11	14	18

[1]We originally (in 2014) established at least three plots per stand (Donovan et al. 2018), but 15 of the original 75 plots were harvested from 2016 to 2018, and three plots had missing defoliation data. Analyses in this paper are all within plots, and stands were used here only to summarize general characteristics.

[2]DBH = diameter at breast height.

[3]Species abbreviations: BF = balsam fir; BS = black spruce; WS = white spruce; HW = hardwood species, including balsam poplar (*Populus balsamifera*), American mountain ash (*Sorbus americana*), trembling aspen (*Populus tremuloides*), willow (*Salix* spp.), white birch, yellow birch (*Betula alleghaniensis*), red maple (*Acer rubrum*), mountain maple (*Acer spicatum*), and striped maple (*Acer pensylvanicum*); OSW = non-host softwood species, including eastern white cedar (*Thuja occidentalis*), eastern white pine (*Pinus strobus*), and eastern larch (*Larix laricina*).

Table 3.2: Number and percentage of plots with clustered, dispersed, or random tree stem locations based on average nearest neighbor analyses ($\alpha = 0.05$), for balsam fir, black spruce, white spruce, hardwoods, and all host species per plot.

	Balsam fir	**Black spruce**	**White spruce**	**Hardwoods**	**All host species**
Clustered	0	1 (2%)	0	3 (5%)	0
Dispersed	21 (37%)	3 (5%)	1 (2%)	2 (4%)	22 (39%)
Random	35 (61%)	20 (35%)	12 (21%)	25 (44%)	35 (61%)

Table 3.3: Number and percentage of plots with significantly clustered patterns of defoliation of trees, based on global Moran's *I* analyses among years, for balsam fir, black spruce, white spruce, and all host species in each plot ($\alpha = 0.05$, search radius = 5 m).

Year	Balsam fir	Black spruce	White spruce	All host species
2014	19 (33%)	1 (2%)	2 (4%)	27 (47%)
2015	11 (19%)	3 (5%)	0	16 (28%)
2016	13 (23%)	1 (2%)	0	20 (35%)
2017	6 (11%)	2 (4%)	0	17 (30%)
2018	15 (26%)	2 (4%)	0	19 (33%)

Table 3.4: Abbreviations and description of predictor variables at both plot and tree levels included in Gradient Boosting Machine analysis to determine their relative importance in predicting annual defoliation of a subject balsam fir tree.

Predictor variables	Description
Plot level	
PlotAvgDefol	Average annual defoliation of all host species per plot (%)
PlotAvgBFDefol	Average annual defoliation of balsam fir per plot (%)
PlotBFBA	% basal area of balsam fir
PlotBSBA	% basal area of black spruce
PlotWSBA	% basal area of white spruce
PlotHWBA	% basal area of hardwoods
Spray	Dummy variable: whether the plot was sprayed by insecticide (1) in corresponding given year or not (0)
Tree level[1]	
BA	Basal area of the subject balsam fir (m^2 ha^{-1})
PreYearDefol	Annual defoliation of subject balsam fir in previous year (%)
NeiAvgDefol	Average annual defoliation of neighboring host trees (%)
NeiAvgBFDefol	Average annual defoliation of neighboring balsam fir (%)
NeiAvgBSDefol	Average annual defoliation of neighboring black spruce (%)
NeiAvgWSDefol	Average annual defoliation of neighboring white spruce (%)
NeiHostBA	Total basal area of neighboring host trees (m^2 ha^{-1})
NeiBFBA	Total basal area of neighboring balsam fir (m^2 ha^{-1})
NeiSPBA	Total basal area of neighboring spruce trees (m^2 ha^{-1})
NeiHWBA	Total basal area of neighboring hardwoods (m^2 ha^{-1})
NeiHBA	Total basal area of all trees with basal area greater than the subject balsam fir in the neighborhood (m^2 ha^{-1})
NeiHostHBA	Total basal area of host trees with basal area greater than the subject balsam fir in the neighborhood (m^2 ha^{-1})
NeiSPHBA	Total basal area of spruce trees with basal area greater than the subject balsam fir in the neighborhood (m^2 ha^{-1})
NeiBFHBA	Total basal area of balsam fir with basal area greater than the subject balsam fir in the neighborhood (m^2 ha^{-1})

[1]Search radii of 3, 4, and 5 m were used for balsam fir trees in a circular subplot of 6 m inside each plot.

Table 3.5: Adjusted R^2, root mean squared error (RMSE), and mean bias of predictions of individual balsam fir defoliation (%) by candidate models with neighborhood tree search radius equal to 3 m, 4 m, and 5 m. Predictor variable abbreviations are described in Table 3.4.

Candidate models	**Predictors**	**Fit statistics[1]**		
		Adjusted R^2	RMSE	Bias
Model 1	PlotAvgBFDefol + BA	0.8001	0.1411	0.0028
	Search radius = 3 m			
Model 2	NeiAvgBFDefol + BA	0.7539	0.1566	0.0017
Model 3	PlotAvgBFDefol + BA + NeiHostHBA	0.8015	0.1406	0.0025
	Search radius = 4 m			
Model 4	NeiAvgBFDefol + BA	0.7823	0.1473	0.0019
Model 5	PlotAvgBFDefol + BA + NeiHostBA	0.8007	0.1409	0.0027
	Search radius = 5 m			
Model 6	NeiAvgBFDefol + BA	0.7889	0.1450	0.0025

[1]The fit statistics were tested for fixed effect models without random effect terms, which in previous model runs had little contribution to models by likelihood ratio tests ($p = 0.9$).

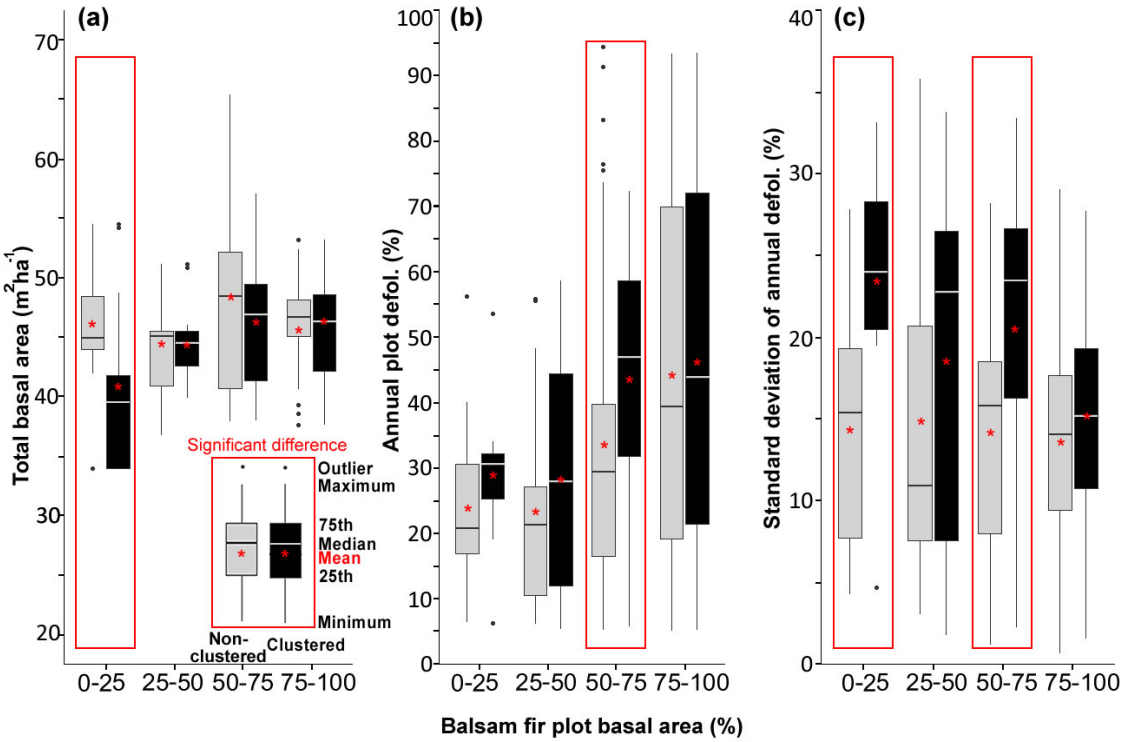

Figure 3.1: Comparison of plots with and without significant clustering of defoliation (based on results of global Moran's *I* analyses (α =0.05) for all host trees) by 25% balsam fir plot basal area classes for (a) total basal area in the plot, (b) average current year defoliation, and (c) standard deviation of individual tree defoliation within plots.

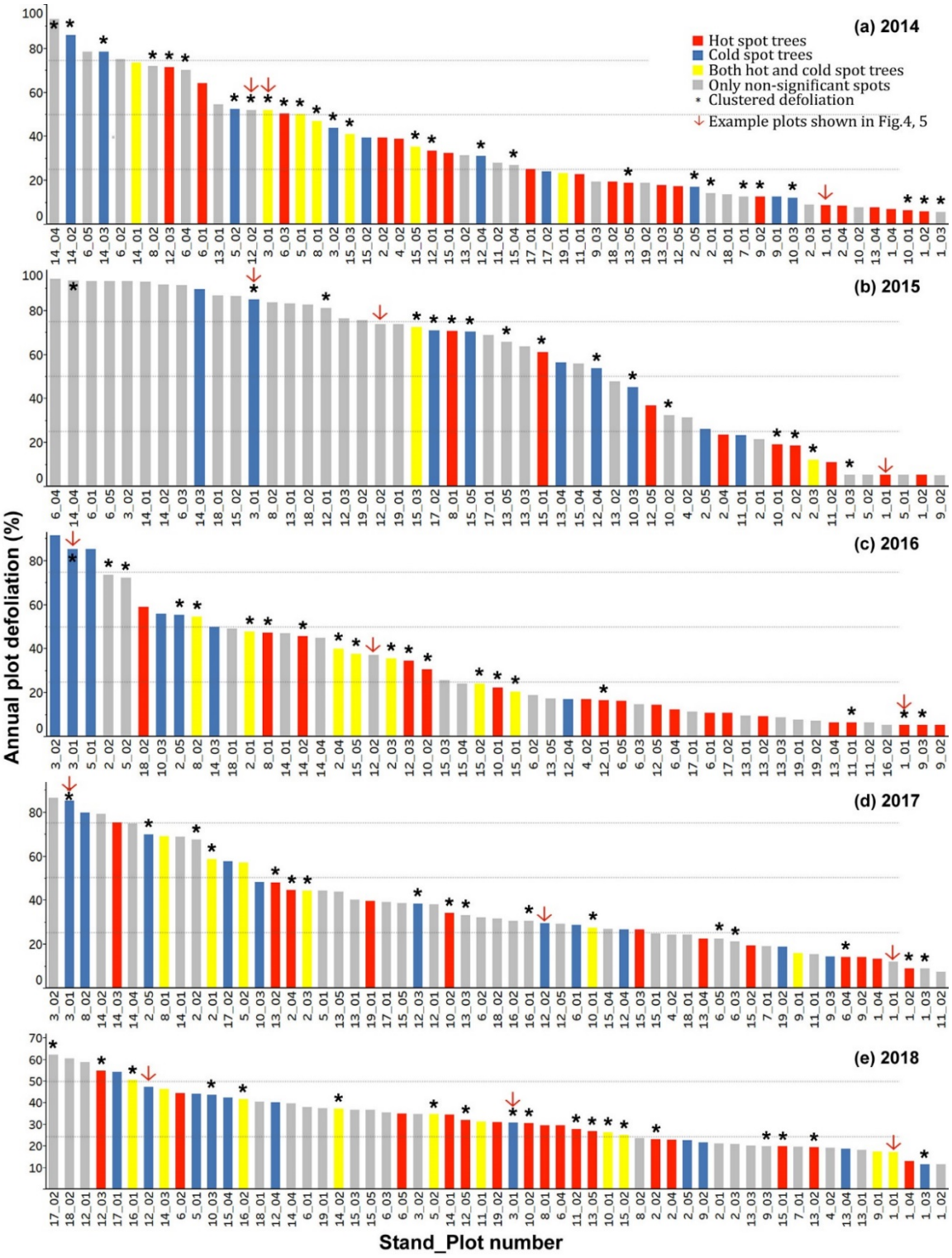

Figure 3.2: Average current year defoliation of plots, ordered from highest to lowest defoliation each year from 2014 to 2018, showing plots with significantly clustered (α =0.05) defoliation (*), with hot spot trees (red), cold spot trees (blue), both hot and cold spot trees (yellow), and only non-significant trees (grey), based on results of Getis-Ord *Gi** analyses (α =0.05).

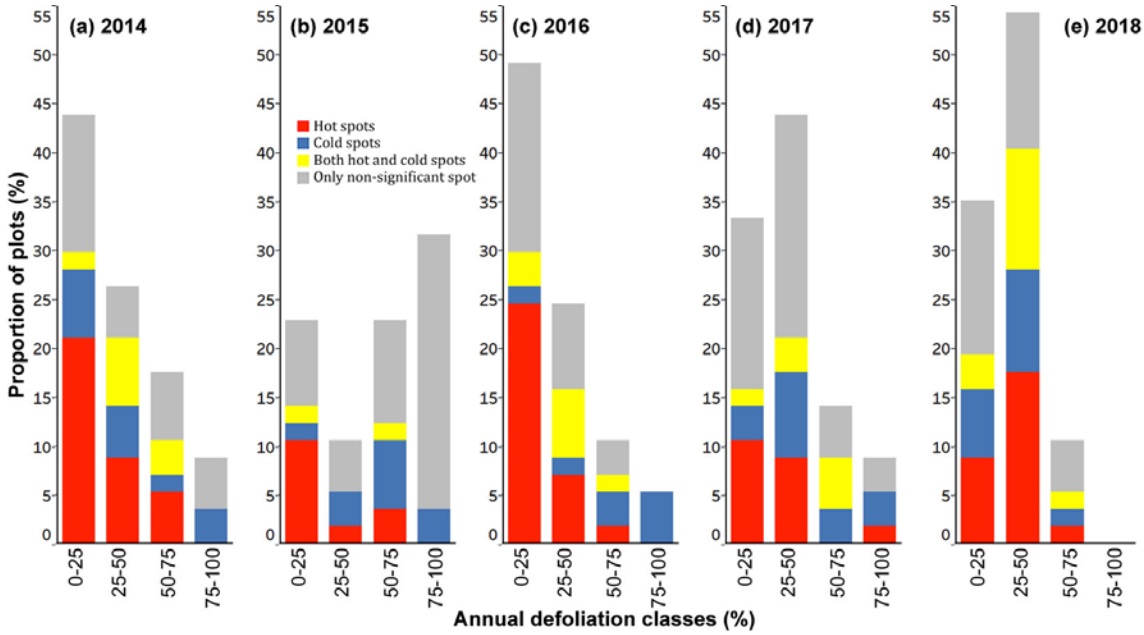

Figure 3.3: Proportion of plots with hot spot trees, cold spot trees, both hot and cold spot trees, and only non-significant trees (based on results of Getis-Ord *Gi** analyses (α =0.05)) by 25% annual defoliation classes from 2014 to 2018.

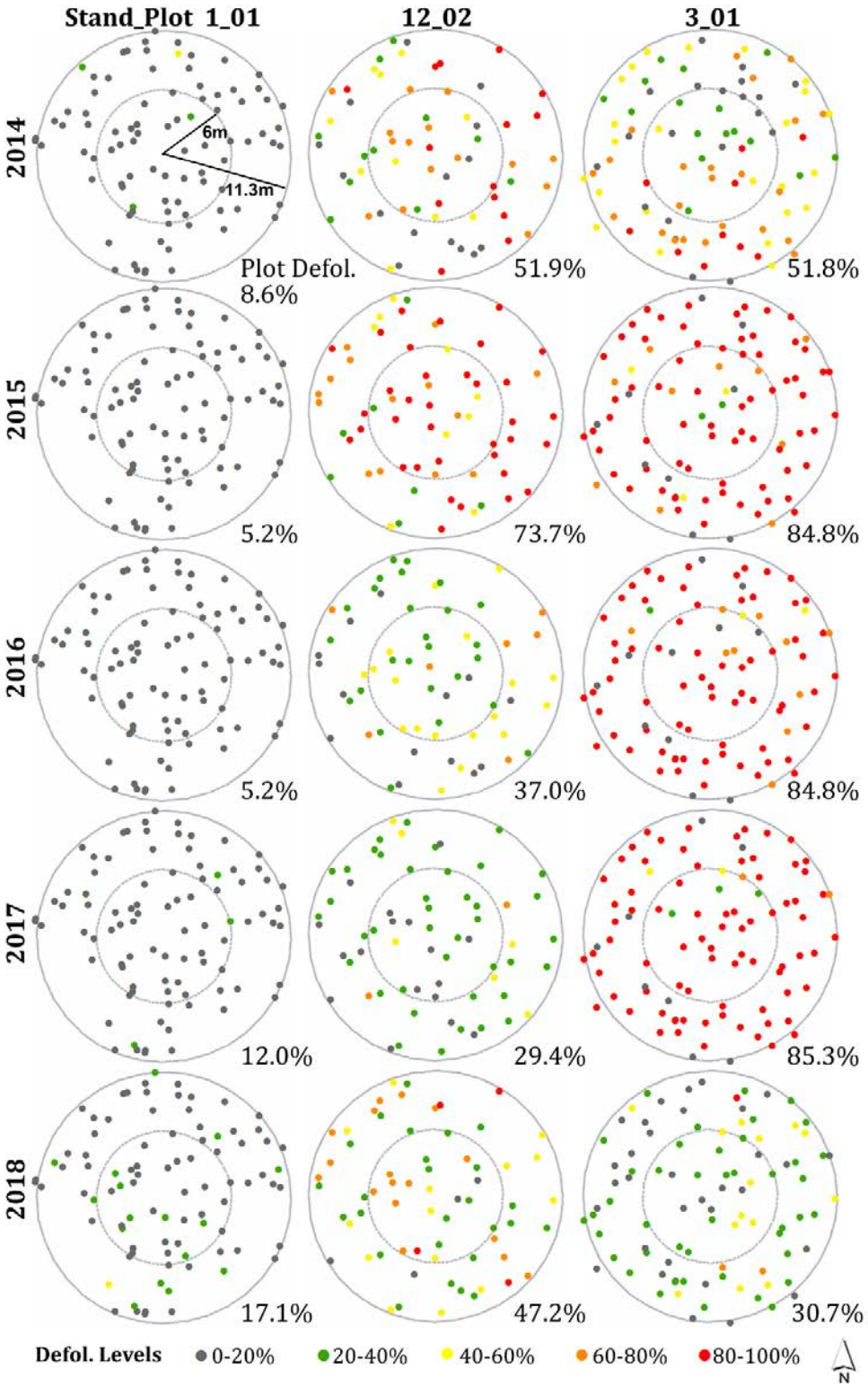

Figure 3.4: Stem maps of tree locations (diameter at breast height (DBH) ≥ 10 cm) of three example plots for 5 years, showing spatial distribution of defoliation. The three example plots were selected to represent generally low (1_01), moderate (12_02), and high (3_01) defoliation levels that contained hot spot and cold spot trees (shown in Figure 3.5).

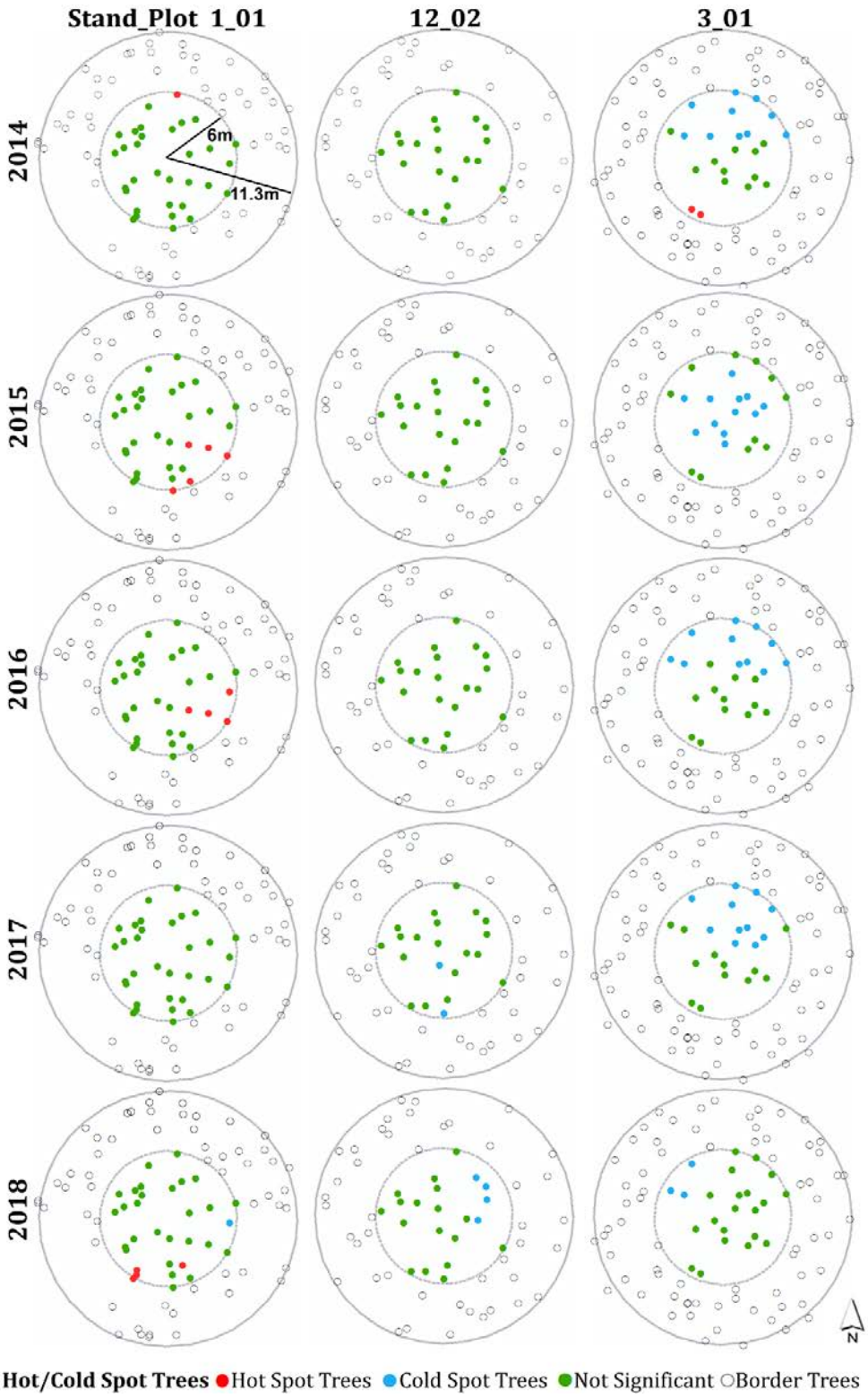

Figure 3.5: Stem maps of tree locations shown for the inner 6 m center of three example plots (same as in Figure 3.4) for 5 years, showing spatial distribution of hot spot trees, cold spot trees, and non-significant trees (based on results of Getis-Ord *Gi** analyses; $\alpha = 0.05$).

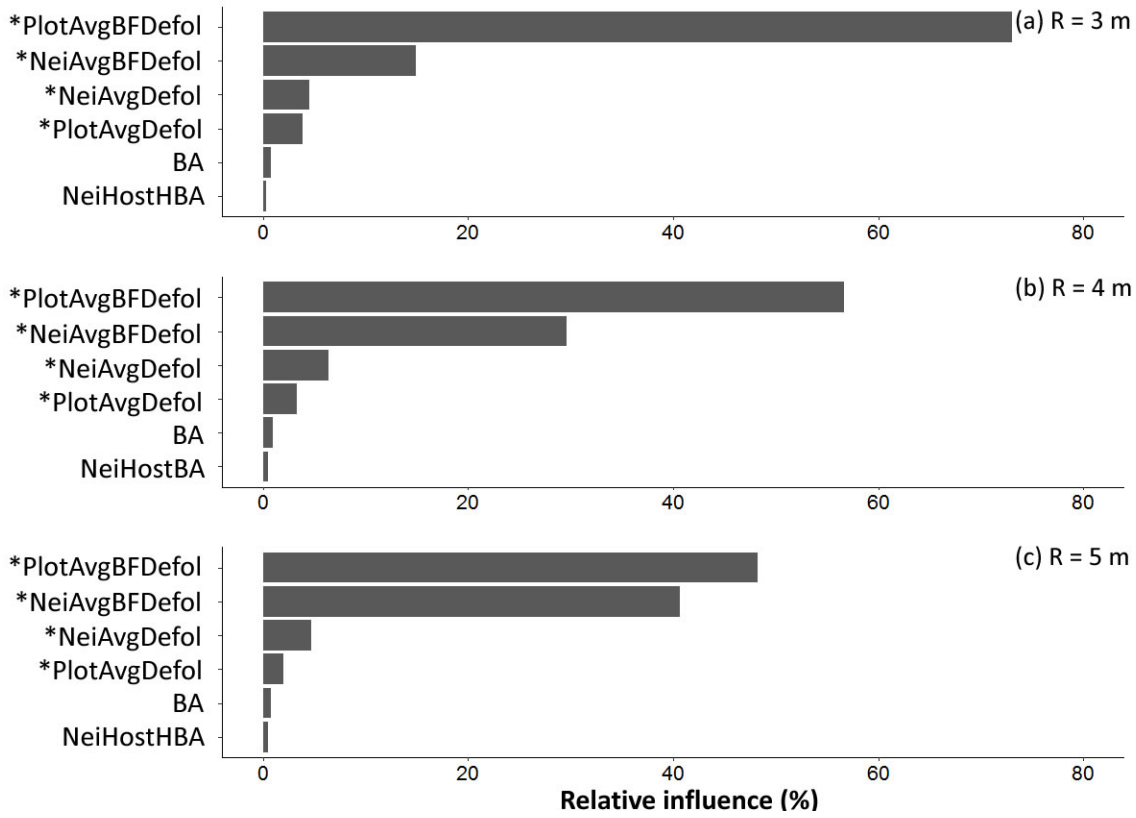

Figure 3.6: Relative influence (%) of the six most important predictor variables based on Gradient Boosting Machine analysis to predict current year defoliation of individual balsam fir trees (%) with neighborhood tree search radius (R) of (a) 3 m, (b) 4 m, or (c) 5 m. Predictor variable abbreviations are described in Table 3.4, and predictors marked with * were highly correlated with each other (correlation coefficient $r \geq 0.7$).

CHAPTER 4: DISCUSSION AND CONCLUSIONS

4.1. Introduction

The overall objectives of this thesis were to determine potential effects of previous spruce budworm (*Choristoneura fumiferana* Clem.; SBW) outbreak conditions, insecticide spray treatment, forest and environmental factors on SBW populations during the outbreak initiation in northern New Brunswick, and to evaluate spatial patterns of SBW defoliation of individual trees within plots in Gaspé, Québec. The current SBW epidemic began about 2006 in Québec, and expanded to northern New Brunswick, adjacent to the Gaspé, Québec area, with increasing SBW second instar larval (L2) populations being observed especially since 2014. Although SBW populations experienced over 90% unexpected reductions in 2018, the decline may be a temporary annual reduction in an overall trend of population increases (MacLean et al. 2019). The improved understanding of influencing factors of SBW population during the outbreak initiation benefits sampling strategy and treatment allocations in the 'early intervention strategy' (EIS) research project, a proactive population control strategy to suppress rising SBW populations before major defoliation occurs. Thus, factors potentially impacting SBW populations in northern New Brunswick were determined from 2014 to 2018 in Chapter 2.

In previous research, Baskerville and MacLean (1979) observed that SBW-caused tree mortality of balsam fir (*Abies balsamea* L. Mill.) had a strongly contagious spatial distribution, which suggests that clustered severe defoliation patterns could be the underlying reason. In fact, the spatial dependence among individual defoliated trees within a plot has not been quantified, and effects of defoliation level of neighboring trees have not been addressed. Hence, spatial patterns of 5 years of individual-tree SBW defoliation within 57 plots in Québec were evaluated from 2014 to 2018 in Chapter 3. The thesis as a

whole also attempted to apply spatial analysis techniques in the forest ecology context, bridging the gap between ecological mechanisms and the behavior of the investigated ecological phenomena.

The objectives of this chapter are to: 1) summarize the main results from the previous two chapters; 2) identify the study limitations and critique the methods used in two projects; 3) relate the previous two chapters, and discuss the management implications of the thesis results.

4.2. Summary of Results

4.2.1. Previous year SBW population level, proximity to high population sites, and spring climate predict SBW population changes in the following year

The possible influencing factors on SBW population levels involved in this analysis included: local SBW population in the previous year, proximity to high population sites in the previous year, insecticide spray history, forest composition, spring and summer climate conditions, topographic characteristics, and site quality. SBW L2 data in Gaspé-Bas St. Laurent region were incorporated in the proximity calculation, although only those L2 points within the northern New Brunswick study area were used in the statistical analyses.

Based on the Gradient Boosting Machine analysis, local previous-year L2 population levels, proximity to sites with high population in the previous year, and early spring climate were consistently the most important factors for predicting current year SBW population levels from 2015 to 2018. In 2014, early spring climate and biomass growth index were the top influencing factors. Simultaneous autoregressive mixed (SAR_{mix}) models were used to address spatial autocorrelation in L2 data when forecasting L2 populations, with local

previous-year SBW populations, proximity to sites with high SBW population in the previous year, and cumulative degree days in April as three predictors. The SAR_{mix} models explained 68% to 79% of the variance in the L2 population, which was higher than that explained by the ordinary least square (OLS) models (6% to 35%) over the years from 2015 to 2018. Additionally, results showed that SAR_{mix} models were able to address the spatial autocorrelation in model residuals, which indicated the suitability of SAR_{mix} models for the ecology data analyzed in this study. To aggregate multiple years' data and generate a generalized model form, the linear mixed effects (LME) model was fit combining 2015-2018 data with year as a random effect. The LME model explained 53% of the variance in L2 population, versus only 22% of the variance explained by the OLS model.

Results of sensitivity analysis of the predictor variables indicated that, in general, locations with or closer to sites with high SBW population levels detected in the previous year tended to have high populations in the current year. In 2016 and 2017, higher L2 populations were associated with lower cumulative degree days in April. Previous year outbreak conditions such as local population level and proximity to high population locations were the most important factors for predicting SBW L2 population in the subsequent year, and the cumulative degree days in April helped to characterize the upper and lower population bounds in the predicted year.

4.2.2. Spatial-temporal patterns of spruce budworm defoliation within plots in Québec

With the application of spatial autocorrelation analysis, I investigated the spatial-temporal patterns of SBW defoliation within 57 plots over 5 years during the current SBW

outbreak in Québec. Even though the baseline tree locations were not spatially clustered, defoliation was found to be clustered in 28% to 47% of plot-year samples, depending on the plot defoliation levels. Plots with clustered defoliation tended to have higher and less uniform defoliation among trees. The exhibited spatial defoliation patterns resulted from uneven SBW pressure on trees, which probably was due to oviposition site selection by female moths. Hot spot trees and cold spot trees were detected in plot-year samples, and whether the defoliation was high enough to be a hot spot tree or low enough to be a cold spot tree depended on the general defoliation level of the whole plot. This might be one of the reasons for the spatially contagious SBW-caused tree mortality observed by Baskerville and MacLean (1979), considering that tree mortality is strongly associated with cumulative defoliation (Erdle and MacLean 1999; Chen et al. 2017). A group of trees suffering more severe defoliation than other trees in the stand, which are clustered in space, may result in clustering patterns of mortality, i.e., “holes”, within the stand. Additionally, in the long run, windthrow disturbance should also be an important factor contributing to the creation of “holes” in the stands (Taylor and MacLean 2007, 2009).

Nonlinear mixed effect regression models were fitted to determine whether defoliation levels of individual balsam fir can be predicted by plot or surrounding-tree characteristics. Plots nested within years firstly served as random effects, then were removed in the reduced models because the variance associated with plot-nested-in-year groups could occur by chance, indicated by the likelihood ratio tests. Results of the reduced models with fixed effects indicated that plot-level average defoliation, as the top predictor in forecasting individual balsam fir defoliation, explained 80% of the variance in the response, and it was higher than the variance explained by the neighboring tree

defoliation with different search radii as predictors. The result that plot average defoliation was superior to neighboring tree defoliation in predicting subject fir defoliation was surprising, given the variability in defoliation among trees within a plot. However, average annual defoliation of surrounding balsam fir was the most important among all tree-level predictors in forecasting subject fir defoliation, and explained 75%, 78%, and 79% variance in the response when the search radius increased from 3 m, to 4 m, to 5 m. The relative influence of neighboring tree defoliation was nearly as high as plot average defoliation with a 5 m search radius. This evidence suggested that a neighborhood search radius greater than 5 m, and hence larger overall plot size, could be needed to better predict individual balsam fir defoliation, and this could be superior to average plot defoliation for the individual balsam fir defoliation forecasts.

The species composition at either the plot level or tree level was not a key factor in the defoliation prediction for those primarily balsam fir plots. Hardwoods comprised merely 6% of the basal area on average in the sample plots, which did not give enough samples to test neighborhood impacts on subject tree defoliation, although the hardwood content has been shown to affect SBW defoliation (Zhang et al. 2018). Insecticide spray treatment was noted as an important explanatory factor of plot-level defoliation (Donovan et al. 2018), but it did not show noticeable influence in this study, probably because the spray effects are already reflected in the plot average defoliation variable. The insecticide treatment together with the SBW population level led to a considerable variation of individual-tree defoliation from year to year, which explained the potential reason that previous-year defoliation showed little influence on current-year defoliation.

4.3. Critique and Study Limitations

4.3.1. Study phase of the outbreak in New Brunswick

The current SBW outbreak in New Brunswick started in 2014, and since then moderate and higher levels of SBW L2 populations have been observed in the northern region of the province. Our study focused on the outbreak initiation phase, from 2014 to 2018. Due to the fact that the current SBW outbreak in New Brunswick is expanding from the north, adjacent Gaspé-Bas St. Laurent region, the most northern region in New Brunswick, suffered from the epidemic earlier than the other southern regions in the province. Hence, we observed that, in terms of early spring degree days, higher population levels tended to occur in the region with a colder climate. As the SBW outbreak develops further, it may continue to expand southward in New Brunswick, and the relationship between insect population and spring climate may well be weakened with the expanding process of the outbreak. In short, the relationships between environmental factors and insect population may vary during different periods of an outbreak and different locations. The conclusions and model prediction may not be applicable in the following outbreak phases. Moreover, SBW population levels experienced over 90% reductions in 2018. The causes behind the unexpected reduction still remain unknown, and cannot be answered by this study, although the decline was thought to possibly be a temporary annual reduction in a trend of population increases (MacLean et al. 2019).

4.3.2. Plot sampling in Québec

The 57 circular 400 m^2 study plots in Québec used in Chapter 3 were a subset of plots studied by Donovan et al. (2018) and Zhang et al. (2018). These plots were

established at least 75 m apart, with plot groupings ranging from 1 to 6 per stand at the beginning. However, 15 of the original 75 plots were harvested from 2016 to 2018, and three plots had missing defoliation data so were excluded. Plot selection criteria included forest composition (i.e., dominated by SBW host species), and the full range of current defoliation present in 2014, which was the year of plot establishment (Donovan et al. 2018). Ten plots in each 10% current defoliation class from 0 to 100% were originally planned to be established for a balanced sampling design (Donovan et al. 2018), but not all were available. According to the modeling results in Chapter 3, plot sizes larger than 400 m^2 may provide better predictions of subject balsam fir defoliation, given that spatial variability of defoliation decreased with larger radius neighborhoods from 3 m to 5 m. On the other hand, considering that there was far more between-plot variability than within-plot variability in defoliation, a large number of plots is preferred than fewer larger plots.

For the purpose of detecting within-plots defoliation patterns, the relative locations of each plot did not impact the study results. However, if scientists would like to investigate the stand-level or regional defoliation patterns, plot locations do matter. In these cases, the distances between sample plots (i.e., spatial lag distances) should be smaller than the expected defoliation patch size, in other words, there should be several sample plots within each stand/regional scale defoliation patch (Dale and Fortin 2014). Furthermore, to detect stand-level or regional defoliation patterns, the plot size should be large enough to contain a reasonable number of individual trees while small enough to avoid a large within-plot variability. However, it is common in nature that researchers would not have prior knowledge of the investigated spatial pattern sizes. Thus, multiple attempts using various spatial lag distances and sample unit sizes are recommended

(Fortin et al. 1989). Field data collection is undoubtedly time-consuming, so remotely sensed data gathering could be preferred to study defoliation patterns at a larger scale, if the related image interpretation techniques are accurate.

4.3.3. Spatial modeling methods

Alternative spatial modeling techniques could have been used to predict SBW L2 populations in Chapter 2. This study tried to use both SAR models and autocovariate models to predict SBW L2 populations, and it was evident that both models had better performance than classic OLS models. However, one year's autocovariate model form cannot be applied on another year's dataset to predict a given year's insect population, because the autocovariate term used in the model needs to be based on the observed population in the predicting year, which would remain unknown until the observation is available.

In addition to SAR models, there are also other model forms that have the capability of "spatial filtering," accounting for the spatial structure of the modeling data. For example, conditional autoregressive (CAR) models, very similar to SAR models, are also used by ecologists, where the probability of a particular location taking a particular value is conditional upon the neighboring values (Dormann et al. 2007). CAR models only consider the first order neighboring effects, whereas the SAR model allows for higher order neighboring effects (Keitt et al. 2002). Moran's eigenvector maps (MEMs) could be used as spatial predictors, especially when investigating regional-scale or multi-scale spatial structures, although biased parameter estimation and inflated Type I error rates were found when using MEMs (Beale et al. 2010). For those spatial structures having

both spatial autocorrelation and heterogeneity, i.e., tending to be classified into different regions, geographically weighted regression (GWR) might be advisable (Dale and Fortin 2014). It fits as many regressions as there are sampled locations using neighboring samples (Fotheringham et al. 2002). Other than the above models intending to balance the impact from spatial autocorrelation, wavelet revised models managed to remove spatial trends from the model errors by replacing the response data and covariates with the alternative values (Carl and Kühn 2008). Nonetheless, removing large-scale spatial trends in the response can lead to poor model performance (Beale et al. 2010).

Among all these spatial modeling approaches, researchers should select wisely depending on the study purpose, data characteristics, sample sizes, model types, and sources of spatial autocorrelation. Performance of different spatial models has been estimated and compared by many studies, which gives ecologists evidence regarding model selection in different study conditions (Dormann et al. 2007; Kissling and Carl 2008; Bini et al. 2009; Beale et al. 2010).

4.4. Management Implications

The current SBW outbreak initiation in northern New Brunswick was a good chance to explore what factors contribute to the development of an outbreak. Understanding the spatial patterns and mechanisms can help scientists to forecast insect outbreak situations in the future, and help forest managers to make better decisions. In the SBW case, identifying subsequent-year high SBW population areas benefits insecticide treatment strategies and facilitates effective sampling design. For example, compared to logistics-based approaches, strategies of number and location of insect sampling sites will be more

efficient and reasonable to base on bionomics and spatial distribution of the target insects, SBW in our case, to reduce the number of sampling sites while still obtaining reasonable L2 density estimates for a given area. From an early intervention strategy or foliage protection viewpoint, accurately delimiting the occurrence of moderate or higher SBW populations or defoliation is key to small area target-specific insecticide spray application.

Incorporating the clusters of higher or lower defoliation in plots or stands, namely the distributions of defoliation, will lead to more accurate distance-dependent or tree-list growth model (e.g., Lamb et al. 2018) predictions in terms of growth reduction or mortality. Relationships between growth loss or mortality and defoliation values are often non-linear, so that trees defoliated severely will potentially suffer from severe growth loss or mortality. In other words, using the average defoliation as the model inputs when there are trees sustaining more or less severe defoliation will lead to the underestimation or overestimation of the effects, embodied as the predicted rate of growth or mortality. However, given a known distribution of individual-tree defoliation, the forecasts of such effects from tree defoliation will be of greater value. Additionally, for managing defoliated forests, knowledge of tree-to-tree variability of SBW defoliation can also benefit the selection of scales and methods of biological insecticide spray treatment decisions targeting moderate or higher infestation.

This thesis also showed that spatial homogeneity and heterogeneity are essential components in ecological studies. Knowledge of spatial structures benefits relating observed spatial patterns to the ecological hypothesis, and providing insights regarding new ecological hypotheses to be tested. It can also serve as the prior knowledge for

follow-up investigations of spatial patterns. A better understanding of multi-scale spatial structure is important to comprehend the complexity of ecological processes.

4.5. Conclusions

In this thesis, I used several spatial analysis approaches to examine the spatial patterns and factors influencing SBW infestations in eastern Canada forests, based on 6 years of insect population data (2013-2018) and 5 years of tree defoliation data (2014-2018).

First, patterns and factors influencing SBW populations during the outbreak initiation phase in northern New Brunswick were examined. It was concluded that previous-year outbreak conditions, including local population level and proximity to outbreak "hot spots," were important predictor variables for subsequent-year SBW population forecasts. Locations with higher previous-year insect populations and that were closer to the outbreak "hot spots" had more possibility to have higher SBW populations in subsequent years. Spring climate (April degree-days) was the top variable among all environmental variables tested, and it can help to determine the population ranges in the coming years.

Second, spatial patterns of defoliation were detected and quantified within 57 plots in the central Gaspé Peninsula region of Québec. Plots with clustered defoliation tended to have higher and less uniform defoliation among trees, which would reflect uneven SBW pressure on trees, perhaps due to oviposition site selection. It was suggested that the spatially contagious SBW-caused tree mortality observed by Baskerville and MacLean (1979) within a balsam fir stand may have resulted from higher defoliation in a portion of

trees. Within the established 400 m^2 plots, plot-level defoliation and basal area were adequate for modeling individual tree defoliation. Larger plots were needed to determine the optimum neighborhood radius for individual-tree defoliation forecasts.

Spatial analysis techniques used in this thesis included spatial pattern analysis, spatial autocorrelation analysis, spatial overlay analysis, and spatial autoregressive modeling. It was evident that spatial autoregressive modeling can lower the spatial dependence among model residuals and improve the model performance. Spatial autocorrelation analysis can serve as an objective way to quantify ecological patterns.

4.6. References

Baskerville, G.L. and MacLean, D.A. 1979. Budworm-caused mortality and 20-year recovery in immature balsam fir stands. Can. For. Serv., Fredericton, NB, Canada. Inf. Rep. M-X-102.

Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J., and Elston, D.A. 2010. Regression analysis of spatial data. Ecol. Lett. **13**: 246-264.

Bini, L.M., Diniz-filho, J.A.F., Rangel, T.F.L.V.B., Akre, S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araújo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Böhning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Jr, P.D.M., David, S., Ferrer-castán, D., Field, R., Filloy, J., Fleishman, E., Jose, F., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araijo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-gaese, K., Borges, P.A.V., Castro-parga, I., Chey, V.K., Chown, S.L., Marco, P.D., Dobkin, D.S., Ferrer-castain, D., Field, R., Filloy, J., Fleishman, E., Gmez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D.,

Kitching, J., Len-cortes, J.L., Lobo, J.M., Montoya, D., Morales-castilla, I., Moreno, J.C., Oberdorff, T., Olalla-t, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R., and Hawkins, B.A. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography **32(**2): 193-204.

Carl, G. and Kühn, I. 2008. Analyzing spatial ecological data using linear regression and wavelet analysis. Stoch. Environ. Res. Risk. Assess. **22**(3): 315-324.

Chen, C., Weiskittel, A., Bataineh, M., and MacLean, D.A. 2017. Even low levels of spruce budworm defoliation affect mortality and ingrowth but net growth is more driven by competition. Can. J. For. Res. **47**(11): 1546-1556.

Dale, M.R. and Fortin, M.J. 2014. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, UK.

Donovan, S.D., MacLean, D.A., Kershaw, J.A., and Lavigne, M.B. 2018. Quantification of forest canopy changes caused by spruce budworm defoliation using digital hemispherical imagery. Agric. For. Meteorol. **262**: 89-99.

Dormann, C.F., Mcpherson, J.M., Arau, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Ohlemu, R., Peres-neto, P.R., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**: 609-628.

Erdle, T.A. and MacLean, D.A. 1999. Stand growth model calibration for use in forest pest impact assessment. For. Chron. **75**: 141-152.

Fortin, M.J., Drapeau, P., and Legendre, P. 1989. Spatial autocorrelation and sampling design in plant ecology. Vegetatio **83**: 209-222.

Fotheringham, A.S., Brunsdon, C., and Charlton, M. 2002. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, Incorporated, Chichester, West Sussex, England.

Keitt, T.H., Bjørnstad, O.N., Dixon, P.M., and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography **25**: 616-625.

Kissling, W.D. and Carl, G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecol. Biogeogr. **17**: 59-71.

Lamb, S.M., MacLean, D.A., Hennigar, C.R., and Pitt, D.G. 2018. Forecasting forest inventory using imputed tree lists for LiDAR grid cells and a tree-list growth model. Forests **9**: 167.

MacLean, D.A., Amirault, P., Amos-Binks, L., Carleton, D., Hennigar, C., Johns, R., and Régnière, J. 2019. Positive results of an early intervention strategy to suppress a spruce budworm outbreak after five years of trials. Forests **10**(5): 448.

Taylor, S.L. and MacLean, D.A. 2007. Spatiotemporal patterns of mortality in declining balsam fir and spruce stands. For. Ecol. Manage. **253**: 188-201.

Taylor, S.L. and MacLean, D.A. 2009. Legacy of insect defoliators: increased wind-related mortality two decades after a spruce budworm outbreak. For. Sci. **55**(3): 256-267.

Zhang, B., MacLean, D.A., Johns, R.C., and Eveleigh, E.S. 2018. Effects of hardwood content on balsam fir defoliation during the building phase of a spruce budworm outbreak. Forests **9**: 530.

Curriculum Vitae

Candidate's full name: Mingke Li

Universities attended:
Nanjing Forestry University (2013-2016), Bachelor of Geographic Information Science
University of New Brunswick (2016-2017), 4th year exchange student involved in '3+1+1' program
University of New Brunswick (2017-2018), Master of Forestry
University of New Brunswick (2018-2019), Master of Science in Forestry, transferred

Articles published in or submitted to refereed journals:
Li, M.; MacLean, D.A.; Hennigar, C.R.; Ogilvie, J. 2019. Spatial-temporal patterns of spruce budworm defoliation within plots in Québec. Forests 10, 232.
Li, M.; MacLean, D.A.; Hennigar, C.R.; Ogilvie, J. Previous year population level, proximity to high population sites, and spring climate predict SBW population changes in the following year. Planned for submission to For. Ecol. Manage.

Selected non-refereed contributions (posters, presentations, tech. Rep.):
Li, M.; MacLean, D.A.; Hennigar, C.; Ogilvie, J. 2018. Spatial-temporal patterns of spruce budworm defoliation within measured plots in Québec. The 9th Bi-Annual Eastern Canada - USA Forest Science Conference, Fredericton, NB, Canada, Oct. 2018.
Li, M.; MacLean, D.A. 2018. GIS analyses of factors influencing spruce budworm outbreak initiation in northern New Brunswick. SERG International Workshop, Edmonton, AB, Canada, Feb. 2018.