Modelling Mountain Pine Beetle Abundance and Distribution in Novel Hosts and Changing Climate

by

Xiaoqi Xie

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Applied Mathematics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

The mountain pine beetle (*Dendroctonus ponderosae*, Hopkins 1902), an invasive bark beetle native to North America, has expanded its habitat from central British Columbia to Northern Alberta. This expansion poses an immediate threat to jack pine forests, which extend from Alberta to Nova Scotia. Past studies have revealed the beetle's successful attacks on jack pines, and laboratory research has indicated that jack pine possesses limited defensive capabilities. These findings have heightened concerns about the potential range expansion of the beetle and underscore the need to quantify potential risks in jack pine forests. Additionally, lodgepole pine, the primary host of the beetle, has been substantially infested in Alberta. The impacts of climate change could increase the frequency and severity of these infestations in lodgepole pine forests, leading to significant economic and ecological losses. Therefore, studying the effects of climate change on beetle's dynamics is imperative.

In this thesis, we addressed concerns regarding the susceptibility of jack pine and the potential impacts of climate-related covariates on infestations in lodgepole pine forests under climate change. In Chapter 2, we discussed the environmental and ecological covariates significant for mountain pine beetle population dynamics in Alberta and selected a suitable model using these covariates to examine these biological questions. Our results indicated that the hierarchical model outperforms others. Employing this model, Chapter 3 discussed the risk of infestations in jack pine and compared the reproduction rate of jack pine to that of lodgepole pine. The model revealed a low likelihood of infestations in jack pine forests, accompanied by a weaker reproduction rate. Chapter 4 introduced the influence of climate-related covariates

on the presence and abundance of infestations in lodgepole pine forests under climate change. We analyzed the potential influence of each covariate individually and their combined impact on the overall trends in the presence and abundance of infestations. The results underscored the significant role of degree days and overwinter survival rate, indicating an increase in the frequency and severity of infestations.

Preface

This thesis is an original work by Xiaoqi Xie. No part of this thesis has been previously published. This thesis was conducted with collaborations. M.A. Lewis was the supervisory author in all chapters. M.A. Lewis and Micah Brush were involved with concept formation and provided writing guidance for all chapters. Evan Johnson was involved with concept formation and provided writing guidance for Chapters 2 and 3. Chapter 3 was also collaborated with Catherine Cullingham and Jessica Duffy.

Acknowledgements

I would like to extend my heartfelt thanks to my supervisor, Dr. Mark Lewis, for providing such exceptional opportunities to model a real insect using real data of the highest quality. Mark has been incredibly kind and patient, always supporting his students with encouragement and positivity. I also extend my gratitude to Dr. Micah Brush for his patient guidance during my master's degree. I thank my parents for providing financial and emotional support throughout these two and a half years. This study has greatly enriched my understanding of biology. I am also grateful to my co-supervisor Dr. Hao Wang and my committee member Dr. Devin Goodsman, for their valuable feedback on my thesis. I am thankful for the scientific and financial support from TRIA-FoR. Additionally, I wish to express my appreciation to all past and present members of the Lewis Research Group for their expert advice, feedback, and support. Special thanks go to the Forestry Division, Government of Alberta, especially David Strauss and Caroline Whitehouse, for their assistance with our data. Lastly, I acknowledge the use of ChatGPT for assistance in checking grammar and vocabulary and providing comments on the structure of my work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Ecology of the mountain pine beetle

The mountain pine beetle (MPB), *Dendroctonus ponderosae* Hopkins, is widely recognized as the most destructive bark beetle in North America, killing more than 50% of lodgepole pines in British Columbia during the 2000s (Government of Canada, 2021). While the primary host of mountain pine beetle is the lodgepole pine (*Pinus contorta*, Douglas ex Loudon), it inhabits most species of pines such as western white pine (*P.monticola* D.Don) and sugar pine (*P. lambertiana* Douglas) (Amman, 1978; Safranyik & Carroll, 2007).

Mountain pine beetle spends most of its lifetime within the phloem tissues of its host tree (Safranyik & Carroll, 2007). It has four life stages: egg, larva, pupa, and adult. Under optimal conditions, the beetles typically have a one-year life cycle, although at higher elevations in Alberta, a two-year life cycle has also been observed (Government of Alberta, 2023). During the summer, they lay eggs that hatch into larvae, which then spend the winter and spring under the bark until emerging as adults the following summer. Upon emergence, the adults leave the infested trees in search of new hosts. Once a suitable host is found, mountain pine beetles can carry up to three species of blue stain fungi, boring through the bark and mining the phloem, thereby cutting off the nutrient supply to the host tree. This action results in the killed tree turning red (known as a red-top tree) in the following year (Safranyik,

1989). MPB can cause massive mortality in large trees, with lodgepole pines greater than 10 cm in diameter at breast height (DBH) experiencing mortality rates of up to 100% as the DBH reaches 40 cm (Björklund & Lindgren, 2009).

Additionally, the behavior of MPB varies across different population stages, which include endemic, incipient epidemic, epidemic, and post-epidemic (Carroll et al., 2006). During the endemic stage, the beetles primarily attack weakened trees (Bleiker et al., 2014). In the incipient epidemic stage, they begin to successfully attack trees of average DBH, marking the onset of an epidemic. As the beetle population grows, they increasingly attack and reproduce in larger trees. Once most suitable pines are killed, beetles migrate and the local population enter the last stage post-epidemic (Safranyik & Carroll, 2007).

The range expansion by MPB increases the vulnerability of pine species not previously within the beetle's range (Safranyik & Wilson, 2006). In the late 1990s, the beetle population, previously at the endemic level, grew rapidly and caused a large outbreak in central British Columbia (Aukema et al., 2006). This outbreak was then exacerbated by forest fires and global warming (Kurz et al., 2008; Raffa et al., 2008), leading to the beetle's expansion into the northern Canadian Rocky Mountains (Nealis & Peter, 2008; Robertson et al., 2009a). Moreover, this invasion posed an immediate threat to jack pine (*Pinus banksiana* Lamb.), a species genetically closely related to lodgepole pine. Subsequent research has shown that MPB has successfully attacked jack pine, highlighting the potential risk of further spread across the boreal forest in Canada (Cullingham et al., 2011). Along with the invasion into Alberta, jack pine, a major component of Canada's boreal forests with habitats extending from northern Alberta to eastern Canada, is exposed to the risk of infestation (Cullingham et al., 2011; Lusebrink et al., 2013).

In addition to concerns about new host trees caused by the past range shift, climate change could continue range expansion (Sauchyn et al., 2003). According to Hayhoe and Stoner (2019), we can expect an increase in the number of days with summer

temperatures above 25°C and a decrease in cold days below −30°C. These summer changes will likely increase beetle activity, encouraging dispersal, while the milder winters could increase larval overwinter survival rates (Safranyik & Wilson, 2006). Moreover, climate change is expected to alter forest composition. Projections indicate that lodgepole pine habitats in Alberta will diminish by 2080, yet the quality of lodgepole pine is predicted to be higher (Monserud et al., 2008).

We explored the effects of jack pine and the impact of various climate factors under different climate change scenarios by constructing models that integrate environmental and ecological covariates. We chose our covariates based on the understanding of biology. Prior to emergence, beetles develop within the inner bark of trees, where their survival rates are notably affected by external temperatures and the thickness of the phloem (Safranyik & Wilson, 2006). Upon emergence, temperature largely impacts the dispersal patterns of the beetles. A warm and dry environment tends to benefit the beetle population, while a warm and humid environment increases the likelihood of flight (Safranyik, 1974; Safranyik et al., 1975; Shepherd, 1966). Furthermore, wind speed plays a crucial role in flight. Wind can assist beetles in sustaining flight for extended periods, although they typically do not fly if wind speeds exceed their maximum flight speeds (Seybert & Gara, 1970). Additionally, the resistance of pine trees influences infestation levels. The more resistant trees are often able to successfully defend against attacks.

We completed all analyses in the managed area classified by Alberta Environment. The annual aerial surveys conducted by Alberta Environment, which utilize heli-GPS technology to identify the numbers of the red-top trees, are based on the management zone. The whole province is divided into three zones: the leading-edge zone, the holding zone, and the salvage zone. The leading-edge zone is actively monitored and managed, while the holding zone is reserved for future management. The salvage zone, which rarely occurs, represents locations where immediate actions are required. Within the leading-edge zone, Alberta Environment conducts a costly exhaustive

search for red-top trees, ensuring that all infestations are monitored (Sustainable Resource Development, 2007). In this thesis, we made use of the comprehensive data in the leading-edge zone spanning the years from 2007 to 2020. We chose a 500m grid cell size, considering that the distance between two helicopter flight lines in the aerial surveys is approximately 500-1000 m.

1.2 Thesis Objectives

We addressed two biological questions related to infestations in Alberta. The first question concerns the potential range expansion of mountain pine beetle infestations into jack pine forests. The second explores the influence of climatic factors on the outbreak under climate change. The forest composition of lodgepole pine and jack pine presents an excellent opportunity to compare the behaviors of beetles in different pine environments and to discuss the likelihood of their continued eastward movement. Our goal was to provide insights into the interactions between pine species and beetle dynamics, and to assess how climatic factors influence these processes.

1.3 Thesis Outline

This thesis is presented across three content chapters. In the first content chapter, we tested the significance of the environmental and ecological covariates and conducted regression analyses to identify models with the best fit and performance. The subsequent chapter aimed to measure the impact of dominant pine species in Alberta, specifically lodgepole pine, lodgepole-jack hybrid pine, and jack pine, on the beetle's host selection. In the final chapter, we projected potential future outbreak risks considering varying climate change scenarios using the best-performing model from the first chapter. Following the three chapters, we concluded with a synthesis of our findings and reflected on the implications for MPB outbreaks in Alberta.

Chapter 2

A hierarchical model for mountain pine beetle dynamics

2.1 Introduction

The mountain pine beetle (MPB) is an invasive species, native to western North America, killing millions of lodgepole pines in British Columbia (Westfall & Ebata, 2007). While the primary host of mountain pine beetle is the lodgepole pine, it inhabits most species of pines within its range (Amman, 1978; Safranyik & Carroll, 2007). The range of mountain pine beetle is mainly governed by temperature and seasonal suitability (Jenkins et al., 2001; Logan & Bentz, 1999; Logan & Powell, 2001; Taylor & Safranyik, 2003). The outbreak in the late 1990s and 2000s has expanded its range, enabling the beetle to invade new territories east of the Rocky Mountains in northeastern British Columbia and northern Alberta (Nealis & Peter, 2008; Safranyik et al., 2010).

The study of MPB has been ongoing for a long time, with a primary focus on the development of mechanistic models like the age structure models highlighted in the works of Heavilin and Powell (2008) and Goodsman et al. (2016). These models provide valuable perspectives on the beetle's long-term dynamics. However, their precision diminishes in short-term analyses. This issue is particularly relevant for forest managers, who focus on short-term dynamics when planning. To address the need for accurate short-term predictions, advancements in computing power have

led researchers to adopt data-driven statistical models. Studies such as those by Kunegel-Lion and Lewis (2020), Preisler et al. (2012), and Wulder et al. (2006) have implemented logistic regression models to estimate the presence of infestations. Additionally, a zero-augmented beta regression model was developed to quantify infestation rates in Colorado (Kaufeld et al., 2014). Each of these models has demonstrated effectiveness in providing short-term predictive visualizations, aligning more closely with the planning needs of forest managers.

Contrasting with datasets in previous studies, our dataset from Alberta Environment stands out by quantifying the number of infested trees across an area that spans approximately half of Alberta. Using heli-GPS technology, Alberta Environment conducts annual aerial surveys within the managed areas, aimed at thoroughly identifying trees that were infested and killed in the previous year, commonly known as 'red-top trees'. Furthermore, these red-top trees, identified in September, will turn grey in the following year. Subsequent ground surveys and control programs are implemented in October and November. The ground survey detects the trees killed in the current year, known as 'green-dead trees', which look similar to healthy trees. These green-dead trees are identified by counting the number of tubes on the pine, with more than 40 tubes indicating infestation. The control program involves cutting and burning the green-dead trees (Government of Alberta, 2023; Sustainable Resource Development, 2007). These surveys are costly and thus maintain a high level of detection accuracy. In order to maximize the potential of this dataset, we have selected the number of infested trees in each cell as our response variable.

With an emphasis on biological interpretability, we selected Poisson-like regression models over machine learning models. The cold temperatures in Alberta, combined with active management operations, have led to a low level of infestations in the province's pine forests (Taylor & Safranyik, 2003; Wulder et al., 2009). The exhibited excess of cells without infestations further indicates the suitability of hierarchical models, which incorporate a Bernoulli distribution to classify zeros before applying

Poisson-like models.

The objective of this chapter was to establish links between environmental covariates and mountain pine beetle dynamics in the newly expanded range in Alberta and select a suitable model. We started with the covariates used in Kunegel-Lion and Lewis (2020). Our investigation was mainly guided by the following biological hypotheses:

1. Geographical covariates (elevation, northness, eastness, and slope) affect whether beetles are able to infest the stand and the number of trees that succumb to the beetle.

2. Climate related covariates (maximum temperature in summer, minimum temperature in summer, relative humidity, wind speed, soil moisture index, degree days, and overwinter survival rate) affect whether beetles are able to infest the stand and the number of trees that succumb to the beetle.

3. Ecological covariates (pine age, age^2, pine height, last year infestations within a cell, within the adjacent cells, the neighborhood of cells adjacent to those adjacent cells, and weighted influence within a 4 km radius) affect whether beetles are able to infest the stand and the number of trees that succumb to the beetle.

4. There is a quadratic relationship between pine age and resistance (Safranyik et al., 1975).

5. The height derived from an age-height model is significant (see Appendix A.2).

6. The influences of dispersal at local and non-local distances differ.

Additionally, past work demonstrated by Kunegel-Lion and Lewis (2020) revealed that covariates at epidemic and post-epidemic phases can present opposing influences on infestations. We divided our data into outbreak and peak growth declining stages

for analysis, using Ridge penalty and Smoothly Clipped Absolute Deviance penalty (SCAD).

To select the most appropriate model, we evaluated models based on two key criteria: their fit to the current distribution of infestations and their efficacy in predicting future infestations. To assess fitness, we made use of log-likelihood, Bayesian Information Criteria (BIC), and randomized quantile residuals (RQR). For comparing predictive performance, we used accuracy and root-mean-square error (RMSE). The presence of numerous cells without infestations and the allowed difference between the mean and variance in the Negative Binomial distribution lead us to expect that the more complex hierarchical Zero-Inflated Negative Binomial (ZINB) model may more accurately reflect the true distribution of infestations and present higher predicting accuracy.

2.2 Methods

2.2.1 Study area and data preparation

The MPB heli-GPS survey is considered the current best practice for conducting accurate and detailed aerial surveys. In Alberta, these surveys are conducted to exhaustively search for red-top trees within managed areas (Agriculture and Forestry, 2020; Sustainable Resource Development, 2007). The distance between flight lines is determined based on the helicopter's altitude. Along with the goal of minimizing travel distance between the survey blocks, a spacing of 300-500 m between two flight lines is generally suggested (Agriculture and Forestry, 2020). Considering that the distance between flight lines in the selected area shown in the Arcmap is approximately 500-1000 m, we have chosen a 500 m grid cell size for our analysis. The geographic scope of the selected area is illustrated in Figure 2.1, and an example of the track-log within this area is shown in Figure 2.2.

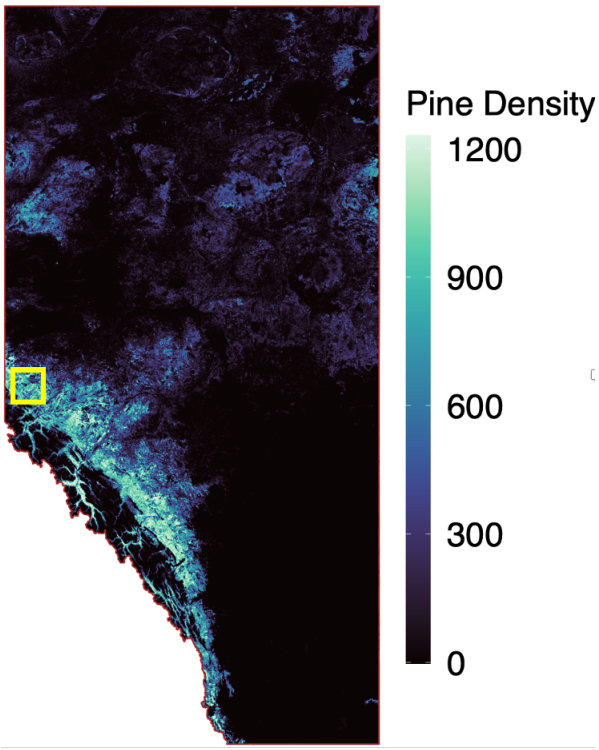


Figure 2.1: Position of the selected area in Alberta map (the yellow rectangle) along with pine density of lodgepole pine and jack pine (stems/hectare).

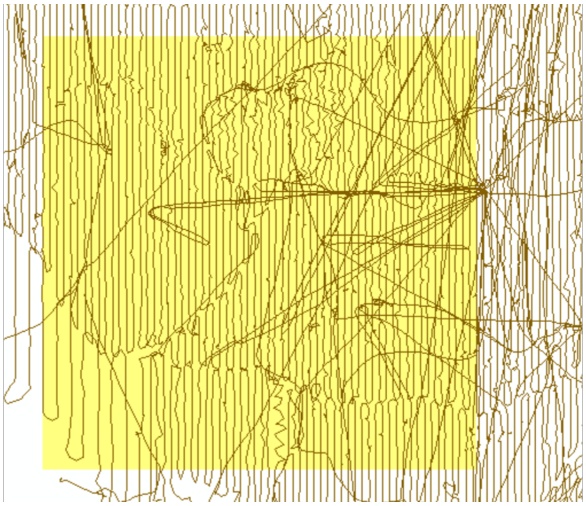


Figure 2.2: Example of the track-log of the selected area (yellow) described in Figure 2.1 (2019 track-log).

Our response variable (the number of infestations) is accounted by the detected red-top trees (aerial survey), green-dead trees (aerial survey and ground survey), and the controlled trees (control program). The number of infestations in year t

is calculated by adding the number of green-dead trees (current year dead trees; aerial survey and ground survey) in year t and the number of detected red-top trees (aerial survey) in year $t + 1$ to the number of controlled trees (current year cut and burned trees; control program) in year t. Our covariates, which include geographical, climatic, and ecological data, are sourced from Alberta Environment (Agriculture and Forestry, 2020), Alberta Vegetation Inventory (AVI) (Agriculture, Forestry and Rural Economic Development, 2022), and National Forest Inventory (NFI) (Beaudoin et al., 2014). For covariates not sourced from these three specified places, we used BioSIM to generate the necessary information. BioSIM calculates this by using a distance-based weighted average derived from the four nearest weather stations.

For the geographical covariates, we included aspect and slope. Elevation is excluded because it is a mere proxy for temperature and has shown a correlation with degree days with a value up to -0.79 in the selected area. Aspect is represented by two covariates: one indicating north-facing slopes, calculated using the cosine function, and the other for east-facing slopes, determined using the sine function. Two covariates reflect the orientation of slopes, which significantly influence sunlight availability, a critical factor in tree growth as highlighted in a previous study Stage (1976).

The climate-related covariates contain temperature, water availability and wind speed. The growth and development of MPB are highly sensitive to temperature, with summer and winter temperatures playing crucial roles. The beetle's development thrive in warm and dry summers, and for successful dispersal, the temperature range of 19°C to 41°C is required (Mccambridge, 1971). Summer temperatures also significantly influence beetle emergence, with the most emergences occurring when temperatures are above 20°C(Safranyik & Carroll, 2007). Winter temperatures affect the survival rates of egg, larvae, pupae, and adults during the overwintering period, as noted by Rosenberger et al. (2017b). Additionally, the accumulation of degree days is essential for the beetle's development, requiring at least 833 degrees above 5.6°C for a one-year life cycle completion (Safranyik & Wilson, 2006). To account for these

factors, our covariates specifically included the minimum and maximum temperatures in summer and degree days. We also introduced the overwinter survival rate as a covariate, which assesses the probability of larval survival over the winter (Régnière & Bentz, 2007). Furthermore, we considered relative humidity and soil moisture index. Soil moisture index reflects the moisture content of the land, which can influence tree health and, consequently, resistance to infestation (Hogg et al., 2013). Relative humidity accounts for the moisture content in the air, impacting both beetle and tree physiology.

The ecological covariates include pine attributes, and past infestations. The attributes of pine trees encompass age, height, and density of lodgepole pine, which is predominant in our research area. As discussed in Safranyik et al. (1975), researchers found that there is an upward quadratic relationship between age and beetle attacks during an outbreak. Thus, we hypothesized a parabolic pattern and included the square of age as a covariate. Since Alberta Vegetation Inventory provides only one year of height data, we used the age-height model depicted in Appendix A.2 to simulate height data. Moreover, we did not have explicit data of stems per cell. We had 2011 total volume of all trees measured in volume per hectare, lodgepole pine abundance percentage, and jack pine abundance percentage (Beaudoin et al., 2014). We estimated the pine stems per cell using the developed volume-numbers model from the appendix S1 of Goodsman et al. (2016). The procedure of estimating expected stems per cells is as follows :

1. volume per hectare of lodgepole pine and jack pine = total volume per hectare $\times$ (lodgepole pine abundance percentage + jack pine abundance percentage)

2. the expected number per hectare = 17.6 $\times$ volume per hectare of lodgepole pine and jack pine$\times$ exp(-0.00527 $\times$ volume per hectare)

3. the expected number per cell = the expected number per hectare $\times$ 25

Since we only had one year of volume per hectare values, we then adjusted the pine number by subtracting or adding the number of current year dead trees from the recorded year.

To evaluate the impact of beetles at various distances, we established four covariates. The observed annual expansion of red-top trees over kilometers leads us to hypothesize that the beetles disperse over intermediate distances (Carroll et al., 2017; Robertson et al., 2009b). However, the exact impact of past infestations at different spatial scales on the adjacent cell is not known. Based on Carroll et al. (2017), we proposed that the influence of one infested tree extends up to 4 km. The first three covariates consider the total number of last year dead trees within 1.25 km around the target cell. Meanwhile, the fourth covariate examines the influence of last year infested trees within a 4 km radius, excluding the impacts accounted for by the first three covariates. To calculate the fourth covariate, we summed the weighted influence of green-dead trees within 4 km of the target cell excluding the previous counted neighborhoods. Let $D_{t,j}$ represent the number of green-dead trees in cell j in year t where cell j is within the 4 km radius but is out of the grid of 1.25 km, λ represent the reproduction rate of females per stem, and f_{ij} represent the flying influence from cell j to cell i. The flying influence is from the empirical cumulative density function shown in the Chapter 3 of Carroll et al. (2017). The beetle pressure $B_{t,i}$ at time t in cell i is calculated :

$$B_{t,i} = \sum_{j} D_{t-1,j} * \lambda_j * f_{ij} \tag{2.1}$$

where we fix the reproduction rate λ to 166.7 (females/stem) (Cole, 1969). The predictors and response variable are summarized in Table 2.1.

Variables	Description
Northness	Face north (1) and south (-1) (cosine of aspect).
Eastness	Face east (1) and west (-1) (sine of aspect).
Slope	Slope of the cell (Burrough et al., 2015).
Pine density	Pine density (stems/cell; composing of lodgepole pine and jack pine) (Beaudoin et al., 2014).
Age	Average age of the leading pines in each cell (Beaudoin et al., 2014).
Age^2	Squared age (Beaudoin et al., 2014).
Pine Height	Average height of the leading pines in the cell (depicted in Appendix A.2).
Maximum temperature in summer	Highest daily temperature in summer in the current year.
Minimum temperature in summer	Minimum daily temperature in summer in the current year.
Relative Humidity	Average relative humidity from March to May in the current year.
Wind Speed	Average wind speed at 2 meters in July and August in the current year.
Soil Moisture Index	Average soil moisture index from May to July, estimating the water supply for tree growth; current year (%) (Hogg et al., 2013).
Degree Days	Cumulative degree days above 5.5°C from September of the previous year to August in the current year (Safranyik et al., 1975).
Overwinter Survival Rate	Probability of survival over the winter; the winter before the current year summer (%) (Régnière & Bentz, 2007).
Last year infestations within cell	Red-top trees (aerial survey) in the current year + green-dead trees (aerial survey and ground survey) in the previous year.
Last year infestations within 0.75 km	Red-top trees (aerial survey) in the current year in the adjacent cells + green-dead trees (aerial survey and ground survey) in the previous year in the adjacent cells excluding the center cell (as shown in Figure 2.3).
Last year infestations within 1.25 km	Red-top trees (aerial survey) in the current year adjacent to the adjacent cells + green-dead trees (aerial survey and ground survey) in the previous year adjacent to the adjacent cells excluding the center cell and adjacent cells (as shown in Figure 2.4).
Dispersal influence within 4 km	Expected number of beetles flying from all directions to the center cell excluding the center cell, adjacent cells of the center cells and cells adjacent to the adjacent cells (Carroll et al., 2017).
Infestations (Response variable)	Red-top trees (aerial survey) in the cell in the next year + green-dead trees (aerial survey and ground survey) +current year control trees (control program).

Table 2.1: Descriptions of the covariates and response variable used in this study.

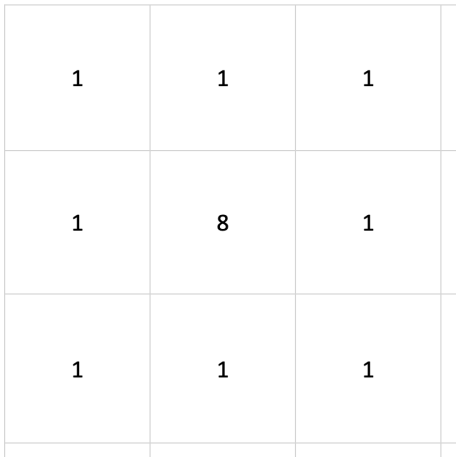


Figure 2.3: Description of the method we used to count the last year infestations within 0.75 km.

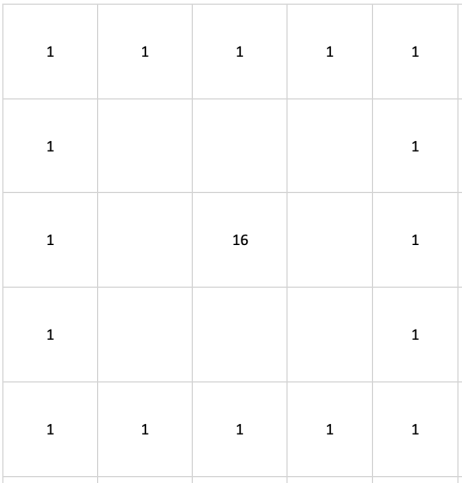


Figure 2.4: Description of the method we used to count the last year infestations within 1.25 km.

2.2.2 Hypothesis Testing

Our dataset featured a significant number of locations without infestations, suggesting the appropriateness of hierarchical models that initially employ a Bernoulli distribution to ascertain presence, followed by assessing the abundance through Poisson-like models. Accordingly, we examined four regression models: Poisson (P), Negative Binomial (NB), Zero-Inflated Poisson (ZIP), and Zero-Inflated Negative Binomial (ZINB). However, during our analysis, we faced convergence challenges with the pa-

rameters in the Negative Binomial model. This subsection focused on the tests of the three models that did not exhibit convergence issues. Let Y be the response variable. Let X be the covariate matrix for the abundance models, specifically Poisson and Negative Binomial models, and let Z represent the covariate matrix for the presence model, which follows the Bernoulli distribution. The probability density functions for Poisson, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial models were presented in equations 2.2, 2.3, and 2.4, respectively. In these models, μ denotes the expected value for the abundance model, π is the expected value for the presence model, β represents the vector of coefficients for the abundance model, γ indicates the vector of coefficients for the presence model, and ϕ is identified as the dispersion parameter in the Negative Binomial distribution.

$$\begin{aligned} \Pr(Y = y|X) &= \frac{\mu^y \cdot e^{-\mu}}{y!}, \quad y = 0, 1, 2, \dots \\ \log(\mu) &= X\beta \end{aligned} \tag{2.2}$$

$$\begin{aligned} \Pr(Y = y|X, Z) &= \begin{cases} \pi + (1-\pi) \cdot e^{-\mu}, & y = 0 \\ (1-\pi) \cdot \frac{e^{-\mu} \cdot \mu^y}{y!}, & y = 1, 2, 3, \dots \end{cases} \\ \pi &= f_{\text{Bernoulli}}(0; Z, \gamma) \\ \text{logit}(\pi) &= Z\gamma \\ \log(\mu) &= X\beta \end{aligned} \tag{2.3}$$

$$\begin{aligned} \Pr(Y = y|X, Z) &= \begin{cases} \pi + (1-\pi) \cdot \frac{\phi^\phi}{(\mu+\phi)^\phi}, & y = 0 \\ (1-\pi) \cdot \frac{\Gamma(y+\phi)}{y! \cdot \Gamma(\phi)} \cdot \frac{\mu^y \cdot \phi^\phi}{(\mu+\phi)^{y+\phi}}, & y = 1, 2, 3, \dots \end{cases} \\ \pi &= f_{\text{Bernoulli}}(0; Z, \gamma) \\ \text{logit}(\pi) &= Z\gamma \\ \log(\mu) &= X\beta \end{aligned} \tag{2.4}$$

In this subsection we focused on testing the biological hypothesis inherent in our covariates using three models. Our covariates addressed several hypotheses, including the significance of geographical, climatic, and ecological factors, the quadratic relationship between pine age and its resistance to beetle attacks (Safranyik et al.,

1975), the significance of pine height derived from an age-height model (see Appendix A.2), and the varying effects of dispersal at local and non-local distances. Specifically, we expected an upward quadratic relationship between pine age and resistance in the outbreak and a downward quadratic relationship in the growth peak declining (Safranyik et al., 1975; Safranyik & Wilson, 2006). We utilized data from 2007 to 2020 in the area depicted in Figure 2.1. We divided 14 years of data into two stages: outbreak and growth peak declining, as shown in Figure 2.5. We tested these hypotheses by employing regularization techniques and incorporating the covariates presented in Table 2.2 across three models. We described the regularization techniques in the subsequent paragraphs and presented the test results in the result section.

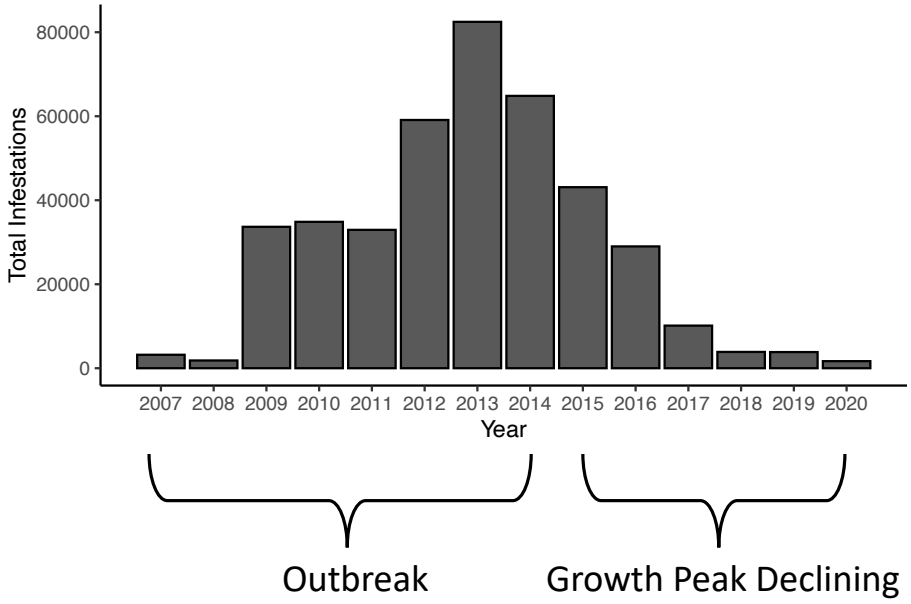


Figure 2.5: Histogram of the total number of infestations in the selected area (see Figure 2.1) that is separated into two stages: outbreak and growth peak declining.

In our analysis, we encountered difficulties due to the high correlation between covariates sourced directly from environmental data and those generated through models that use these environmental data as input. The presence of multicollinearity raised concerns regarding overdispersion and model instability. To address these issues, we incorporated regularization techniques to automatically handle multicollinear-

ity and perform variable selection.

Penalty techniques offer a fast and effective solution for handling multicollinearity and variable selection (Fan & Li, 2001; Hoerl & Kennard, 1970). Multicollinearity can significantly impact prediction accuracy, while variable selection helps reduce model complexity. In our dataset, we encountered both data-based multicollinearity, resulting from pure observations, and structural multicollinearity, where covariates use other covariates as input covariates. One of our objectives was to minimize these collinearities, ensuring that the covariates provide unique information in the regression model. To achieve this, we used the ridge regression (Hoerl & Kennard, 1970).

Let θ^T represent the parameter vector (β, γ, ξ), where β contains coefficients of the abundance model, γ contains coefficients of the presence model, and ξ encompasses all other parameters. Additionally, let $\ell(\theta)$ denote the log-likelihood function. The estimator for Ridge regression, as described in Hoerl and Kennard (1970), is defined as:

$$\hat{\theta} = \text{argmin}_\theta(-\ell(\theta) + \lambda_1 \sum_{i=1}^{n} \theta_i^2) \tag{2.5}$$

where λ_1 is the tuning parameter and n represents the number of non-intercept coefficients. The Ridge estimator effectively reduces multicollinearity by decreasing variance. For sufficiently small values of λ_1, the Ridge estimator yields a mean square error smaller than the maximum likelihood estimator (Hoerl & Kennard, 1970). The mean square error (MSE) is composed of the variance and the squared bias of our estimators, which can be expressed as:

$$\text{MSE} = \text{variance}(\theta) + \text{bias}^2(\theta) \tag{2.6}$$

In this context, by minimizing MSE, the Ridge estimator effectively reduces the variance of θ while introducing a controlled amount of bias. This trade-off is a key characteristic of Ridge Regression, where the increase in bias is compensated by a significant reduction in variance, particularly beneficial in the presence of multicollinearity among covariates.

While Ridge Regression effectively mitigates multicollinearity in the model, there remains a need to simplify the model complexity and concentrate specifically on the most significant covariates. However, the exhaustive best subset selection method employed in Kunegel-Lion and Lewis (2020) is computationally expensive, as the model with n covariates would result in 2^n subsets. To overcome this limitation, we employed a lasso-type penalty known as the Smoothly Clipped Absolute Deviation (SCAD) penalty for variable selection. A lasso-type penalty prunes small coefficients, effectively setting them to zero (Tibshirani, 1996). Initially, we considered three popular lasso-type penalties: Adaptive Lasso, Smoothly Clipped Absolute Deviation (SCAD), and Minimax Concave Penalty (MCP) (Fan & Li, 2001; Zhang, 2010; Zou, 2006). Among these options, we chose the SCAD penalty based on the findings of Z. Wang et al. (2016), who demonstrated that SCAD produces less biased estimators while maintaining continuity in model prediction. H. Wang et al. (2007) also noted that theoretically, BIC serves as a better information criterion when using the SCAD penalty. Compared to the lasso penalty, the SCAD penalty introduces smaller bias by employing boundary conditions. It assigns greater penalization to small coefficients while still considering the impact of larger coefficients. The SCAD penalty is defined as follows (Fan & Li, 2001):

$$p_{\lambda_2}(|\theta|) = \begin{cases} \lambda_2|\theta|, & \text{if } |\theta| \leq \lambda_2 \\ \frac{2\gamma_1\lambda_2|\theta| - \theta^2 - \lambda_2^2}{2(\gamma_1 - 1)}, & \text{if } \lambda_2 < |\theta| < \gamma_1\lambda_2 \\ \frac{\lambda_2^2(\gamma_1+1)}{2}, & \text{if } |\theta| \geq \gamma_1\lambda_2 \end{cases} \tag{2.7}$$

where γ_1 is a free parameter greater than 2 and λ_2 is the tuning parameter. In our analysis, we used the recommended value of 3.7 for λ_1 (Fan & Li, 2001). Furthermore, the smoothness of the estimators is ensured by the definition of the SCAD penalty, specifically through its first derivative:

$$p'_{\lambda_2}(|\theta|) = \lambda_2\{\mathbb{1}(|\theta| < \lambda_2) + \frac{(\gamma_1\lambda_2 - |\theta|)_+}{(\gamma_1 - 1)\lambda_2}\mathbb{1}(|\theta| \geq \lambda_2)\} \tag{2.8}$$

In order to incorporate the Ridge and SCAD penalties into our model, we as-

signed distinct weights to each penalty and these weights were set in such a way that their sum equals 1. The objective was to find the optimal balance between the two penalties, leveraging the variance-reducing property of Ridge Regression and the variable selection capability of the SCAD penalty within a single model framework. We estimated the parameters by minimizing the following equation (Z. Wang et al., 2016):

$$\hat{\theta} = \text{argmin}_\theta - \ell(\theta) + \frac{1}{2}(1-\alpha_1)\lambda_1 \sum_{i=1}^{b_1} \beta_i^2 + \frac{1}{2}(1-\alpha_2)\lambda_2 \sum_{j=1}^{b_2} \gamma_j^2 + \alpha_1\, p_{\lambda_1^*}(\beta) + \alpha_2\, p_{\lambda_2^*}(\gamma) \quad (2.9)$$

Here, α_1 represents the weight of the SCAD penalty in P or NB, while α_2 represents the weight of the SCAD penalty in the binomial model. λ_1 is the tuning parameter of the Ridge regression in P or NB, whereas λ_2 is the tuning parameter of the Ridge regression in the binomial model. λ_1^* is the tuning parameter of the SCAD penalty in P or NB and λ_2^* is the tuning parameter in binomial model. The variable b_1 represents the number of coefficients in the P or NB models excluding the intercept, and b_2 represents the number of coefficients in the binomial model excluding the intercept. We set the threshold parameters γ in the SCAD penalty to be the suggested number 3.7 (Fan & Li, 2001).

To implement the regularization, we utilized the `R package mpath` (Z. Wang et al., 2016). However, the `zipath()` function in `R` does not provide automatic support searching for different weights. Therefore, we conducted a grid search to determine the optimal weights. The optimal weights and the tuning parameters were selected based on the smallest Bayesian Information Criterion (BIC) value when allowing the function to conduct 100 values for each weight. The tuning parameters are assigned automatically by the `zipath()` function.

2.2.3 Model Selection

We employed various criteria to assess the quality of fit for two stages, including log-likelihood, Bayesian Information Criterion (BIC), and Randomized Quantile Residual (RQR). The log-likelihood estimates the probability that the response variable follows the assumed distribution given the input covariates. A higher log-likelihood suggests that the data is more likely to originate from the model being fitted. BIC, on the other hand, increases as the error variance and number of parameters increase. Considering our large dataset, we selected BIC rather than Akaike Information Criterion (AIC) as it tends to converge towards the best model with a probability of 1 when the sample size approaches infinity. The final criterion, Randomized Quantile Residual (RQR), is a residual method specifically developed for regression models with independent responses (Dunn & Smyth, 1996). Its purpose is to evaluate the consistency between the assumed distribution and the observed data, providing a visual indication of fitness through plots. Detailed explanations regarding RQR will follow in the subsequent paragraphs.

A study by Feng et al. (2020) demonstrated that RQR provides a more straightforward evaluation of model fit compared to Pearson and Deviance Residuals, particularly in the context of P, NB, ZIP, and ZINB models. Their study indicated that when P, NB, ZIP, and ZINB models are well-fitted, RQR generally adhere to a normal distribution. This conformity leads to a scenario where the values of RQR closely match those of the standard normal quantile, aligning neatly with the y=x line in the quantile-quantile (q-q) plot. By using their method, we calculated the RQR for the response variable, which is the number of infestations in cell i, denoted as y_i. We also computed the standard normal quantile represented as q_k. The standard normal quantile q_k was calculated by $\Phi^{-1}(\frac{k-3/8}{n+1/4})$ and k is the kth order statistic, $1 \leq k \leq$n and n is the sample size (McCullagh, 1985). We calculated RQR following specific

transformations (Dunn & Smyth, 1996; Klar & Meintanis, 2012):

$$\begin{aligned} \mathrm{F}^*(y_i;\hat{\mu}_i,\phi) &= \mathrm{F}(y_i-1;\hat{\mu}_i,\phi) + u_i \cdot \mathrm{p}(y_i;\hat{\mu}_i,\phi) \\ z_i &= \Phi^{-1}(F^*(y_i;\hat{\mu}_i,\phi)) \\ r_i^Q &= \frac{z_i - \mathrm{mean}(z)}{\mathrm{sd}(z)} \end{aligned} \tag{2.10}$$

In the above equations, $\mathrm{F}(y_i;\hat{\mu}_i,\phi)$ represents the cumulative distribution function (CDF) for a response variable y_i given the vector of covariates x_i, with estimated mean ($\hat{\mu}_i$) and dispersion parameter (ϕ), while $\mathrm{p}(y_i;\hat{\mu}_i,\phi)$ corresponds to the associated marginal probability function. Additionally, u_i denotes a random number generated from a uniform distribution with an interval of (0,1), and Φ follows a standard normal distribution. The vector z contains all z_i values, which are used in the calculations.

Furthermore, we implemented a procedure to visualize the 95% Monte Carlo simulated intervals for the RQRs. This procedure, described in Feng et al. (2020), involves the following steps:

1. generate 100 sets of random numbers that follow the target distribution,
2. re-fit the model using the generated random numbers as the response variable,
3. estimate the RQRs for the generated random numbers using the refitted model and arranging them in ascending order,
4. identify the values corresponding to the 2.5% and 97.5% percentiles of the RQRs for each distribution.

2.2.4 Model Validation

In this chapter, we validated model performance for a single period: the infestation growing period from 2007 to 2013. We did not validate the model using the data from the infestation declining period of 2014 to 2020, as most of the beetles died out during the cold winters of 2019 and 2020. Given the limited number of years available

in this period, we used a straight-forward validation method: we fit the model using data from 2007 to 2011 and validate the outcomes in 2012 and 2013.

Additionally, accurately predicting the exact number of infestations poses a challenge due to the response variable's extensive range, spanning from zero to hundreds in a cell. This broad spectrum adds considerable difficulty in precisely forecasting infestation counts. Therefore, our focus was on assessing the ability of our penalized models to accurately predict the severity classes of beetle infestations. These classes are outlined in Table 2.2, which is derived from the Gibson (2004). The procedures for model validation are depicted in Figure 2.6.

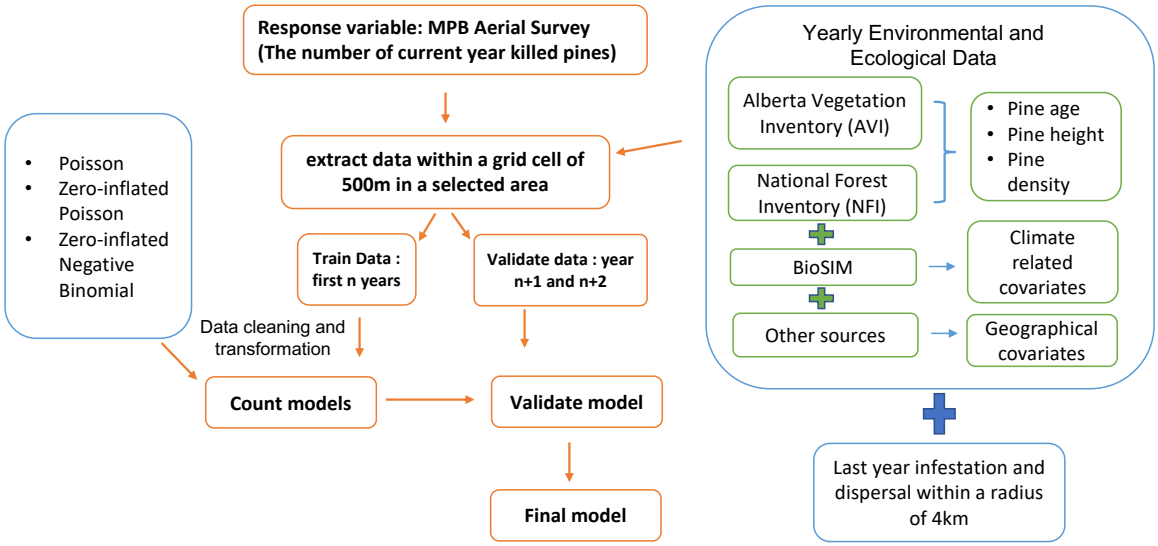


Figure 2.6: The modeling flow used to create the Poisson, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial models incorporated yearly geographical, climatic, and ecological covariates to select the model with the best predictive performance.

Severity	Number of infested trees in a 500m grid cell
no infestation (0)	0
light (1)	1-3
moderate (2)	4-25
very severe (3)	>25

Table 2.2: Severity Categories used to evaluate the severity of infestations in Alberta.

The model performance was evaluated using two indices: accuracy and root-mean-square error (RMSE). Let I_{ij} be the severity level at the cell i in the year j and $\hat{I}_{ij}$ be the predicted severity. The predicted severity class was identified as the one with the highest probability. Accuracy represents the correction rate and is computed by dividing the number of correct predictions by the total number of predictions. RMSE measures the difference between the predictions and the observations, which is defined as $\sqrt{\frac{\sum_{i=1}^{n}(I_{ij}-\hat{I}_{ij})^2}{n}}$ and n is the number of observations. We also presented the confusion matrix showing the number of correct predictions in each severity class.

2.3 Results

2.3.1 Hypothesis Testing

The optimal weights of penalties for three models over two stages were shown in Table 2.3. We presented covariates including age, age^2, height, last year infestations within cell, within 0.75 km, 1.25 km and dispersal influence within 4 km separately from other covariates. The non-zero coefficients of the standardized covariates corresponding to the hypotheses, along with their 95% confidence intervals were shown in Figures 2.7 through 2.12.

Model	Outbreak	Growth Peak Declining
P	1	0.99
ZIP	(Presence model: 1, Abundance model: 0.75)	(Presence model: 0.96, Abundance model: 0.72)
ZINB	Presence model: 1, Abundance model: 1)	(Presence model: 0.9, Abundance model: 1)

Table 2.3: Optimal weights of the penalties of P, ZIP, and ZINB for the outbreak and growth peak declining stages. Poisson model is an abundance model and does not contain a presence model.

In Figure 2.7, we observed that during the outbreak stage, all of the covariates are significant in the Poisson model. Additionally, the coefficients for last year infestations

within 1.25 km were small, with confidence intervals nearing zero. Given that spatial autocorrelation typically leads to underestimated confidence intervals, we can assert with greater certainty that last year infestations within 0.75 km and 1.25 km are of minor importance. The model in the peak growth declining stage identified several insignificant covariates, such as height, last year infestations within 1.25 km, and age. For the remaining covariates, as depicted in Figure 2.8, eastness and soil moisture index were of lesser importance during the outbreak stage, while northness, eastness, and soil moisture index were less significant during the growth peak declining stage.

The ZIP model also yielded many insignificant covariates. As depicted in Figure 2.9, during the outbreak stage, the presence model had only two non-zero coefficients. The abundance model's confidence intervals revealed that the square of age, height, last year infestations within 1.25 km, and age were not significant. During the other stage, the presence model identified one additional significant covariate: the influence of dispersal within 4 km. The abundance model identified only three significant covariates: last year infestations within cell, within 0.75 km and dispersal influence within 4 km. For the remaining covariates, as shown in Figure 2.10, ZIP model only identified four significant covariates in the presence model in the outbreak stage including wind speed, overwinter survival rate, minimum temperature in summer, and maximum temperature in summer. For the abundance model in the same stage, covariates including slope, pine density, northness, eastness, soil moisture index, and degree days were less significant. During the growth peak declining stage, both the presence and abundance models exhibited fewer significant covariates compared to the outbreak stage.

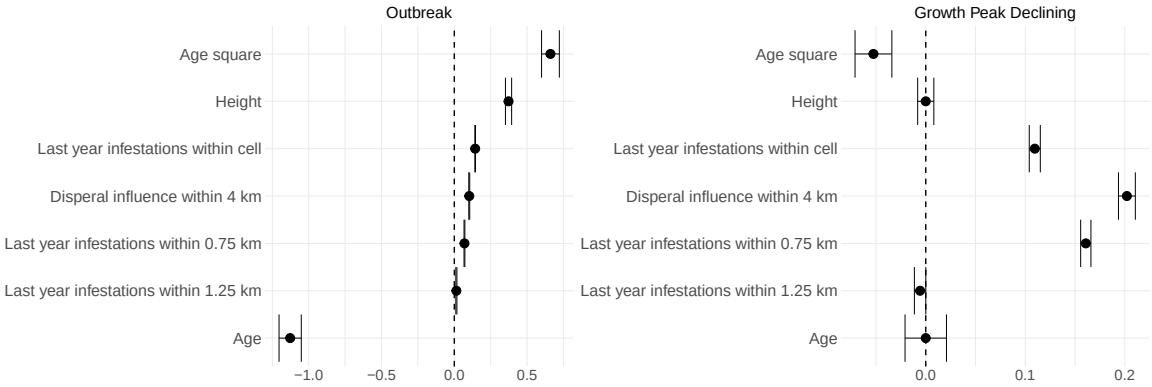
The results of the ZINB models were close to those of the ZIP models. As shown in Figure 2.11, during the outbreak stage, both height and age had coefficients of zero, while age square and dispersal influence within 4 km had confidence intervals closing to zero in the presence model. The abundance model showed non-zero coefficients for age square, height, dispersal influence within 4 km, and last year's infestations within

1.25 km, albeit with confidence intervals that included or were close to zero. In the subsequent stage, the ZINB model exhibited fewer significant covariates. Regarding the additional covariates in Figure 2.12, northness and eastness lacked significance, and the soil moisture index had a coefficient of zero in the presence model during the outbreak stage. In the same stage's abundance model, northness, slope, soil moisture index, degree days, eastness, and pine density were of lesser importance. In the growth peak declining stage, northness and pine density became less significant, while eastness, maximum temperature in summer, and slope showed coefficients of zero in the presence model. In this stage's abundance model, northness, soil moisture index, degree days, eastness, pine density, and relative humidity were also deemed less significant.

The covariates with non-zero coefficients in the Poisson model were different from those in the ZIP and ZINB models. In the Poisson model, the age square covariate was significant with contrasting signs across the outbreak and growth peak declining stages, while this covariate was found to be insignificant in both ZIP and ZINB models. As suggested by previous studies Safranyik et al. (1975), one signal of an outbreak is the mass-attacks to the older and stronger mature trees. However, the Poisson model suggested that in an outbreak, the beetle attacks young and old weakened trees, causing the least infestations in the strong mature trees. Thus, we decided to omit age square from subsequent analyses. Moreover, in the Poisson model, pine height was found to be significant during the outbreak phase, but this was not significant in the ZIP and ZINB models for two stages. Since two zero-inflated models better accounted for the excessive zero infestation in our data, we opted to exclude height from our analysis.

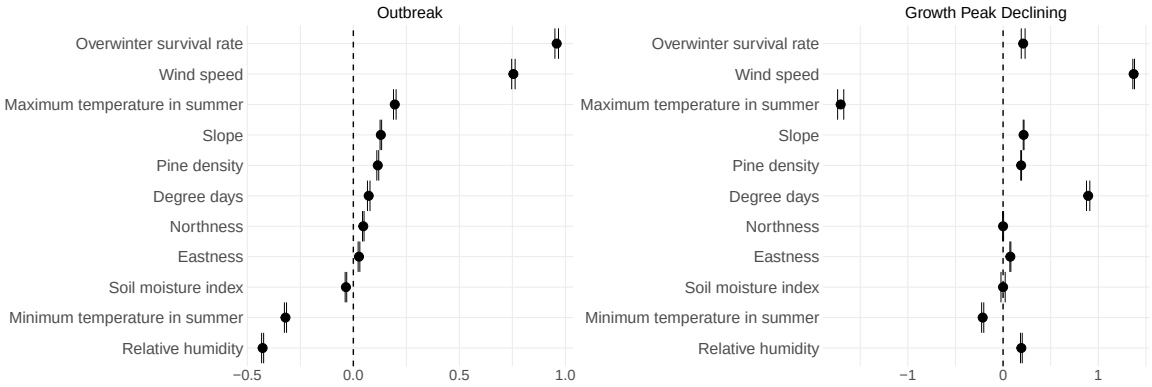
A similar pattern was observed with four covariates related to local and non-local dispersal. The last year infestations within cell, last year infestations within 0.75 km, and the dispersal influence within 4 km were significant in both stages for the Poisson model. However, in the ZIP and ZINB models, these covariates had less significance,

with last year's infestations within cell showing consistent importance across both stages. To align the three models, we decided to integrate covariates that include last year's infestations within 0.75 km and 1.25 km, as well as dispersal influence within 4 km, to be one covariate in our analysis.



(a) (b)

Figure 2.7: Coefficients of age, age^2, height, last year infestations within cell, within 0.75 km, 1.25 km and dispersal impacts within 4 km and their 95% confidence intervals in the (a) Poisson model in the outbreak, (b)Poisson model in the endemic. Covariates not shown in the plots indicate zero values.



(a) (b)

Figure 2.8: Coefficients of covariates not presented in Figure 2.7 and their 95% confidence intervals in the (a) Poisson model in the outbreak stage, (b)Poisson model in the growth peak declining stage. Covariates not shown in the plots indicate zero values.

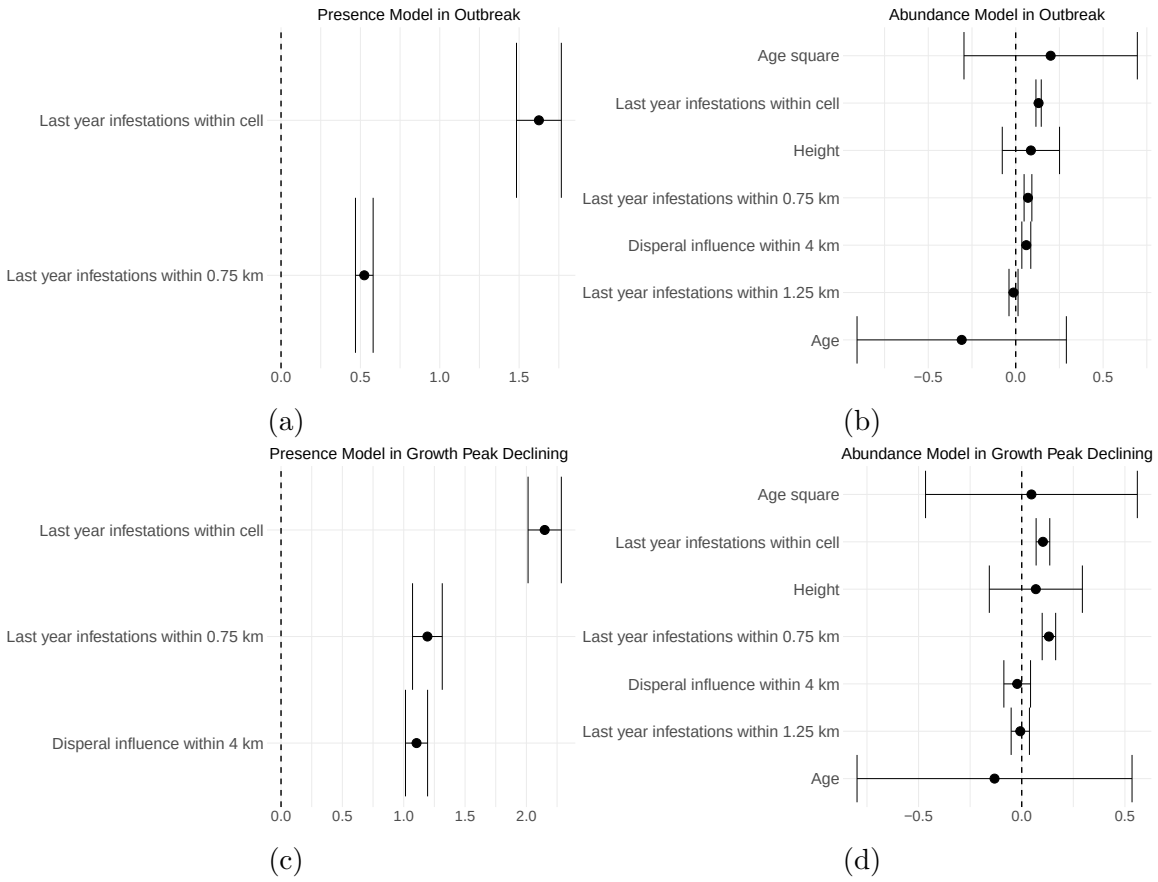


Figure 2.9: Coefficients of age, age^2, height, last year infestations within cell, within 0.75 km, 1.25 km and dispersal impacts within 4 km and their 95% confidence intervals in the (a) Presence model in the ZIP model in the outbreak stage (b) Abundance model in the ZIP model in the outbreak stage (c)Presence model in the ZIP model in the growth peak declining stage (d)Abundance model in the ZIP model in the growth peak declining stage. Covariates not shown in the plots indicate zero values.

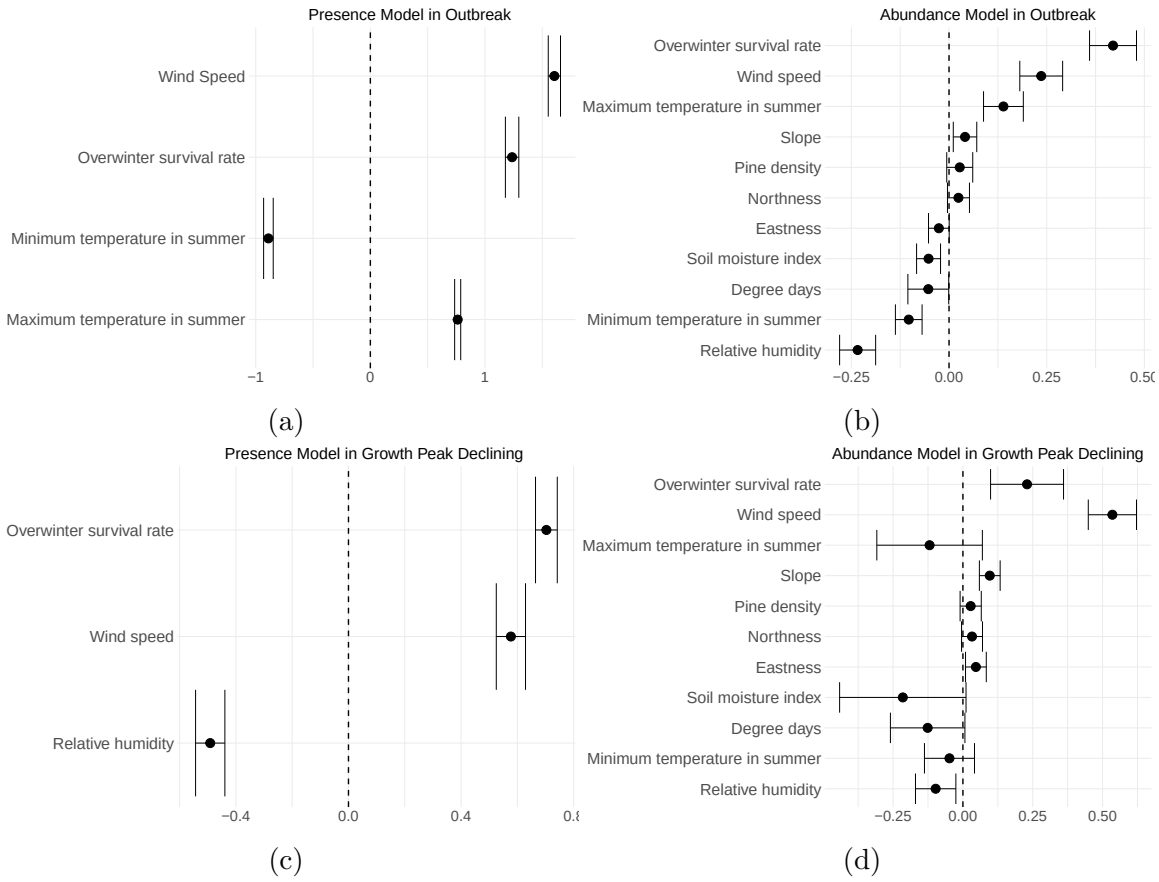


Figure 2.10: Coefficients of covariates not presented in Figure 2.9 and their 95% confidence intervals in the (a) Presence model in the ZIP model in the outbreak stage (b) Abundance model in the ZIP model in the outbreak stage (c)Presence model in the ZIP model in the growth peak declining stage (d)Abundance model in the ZIP model in the growth peak declining stage. Covariates not shown in the plots indicate zero values.

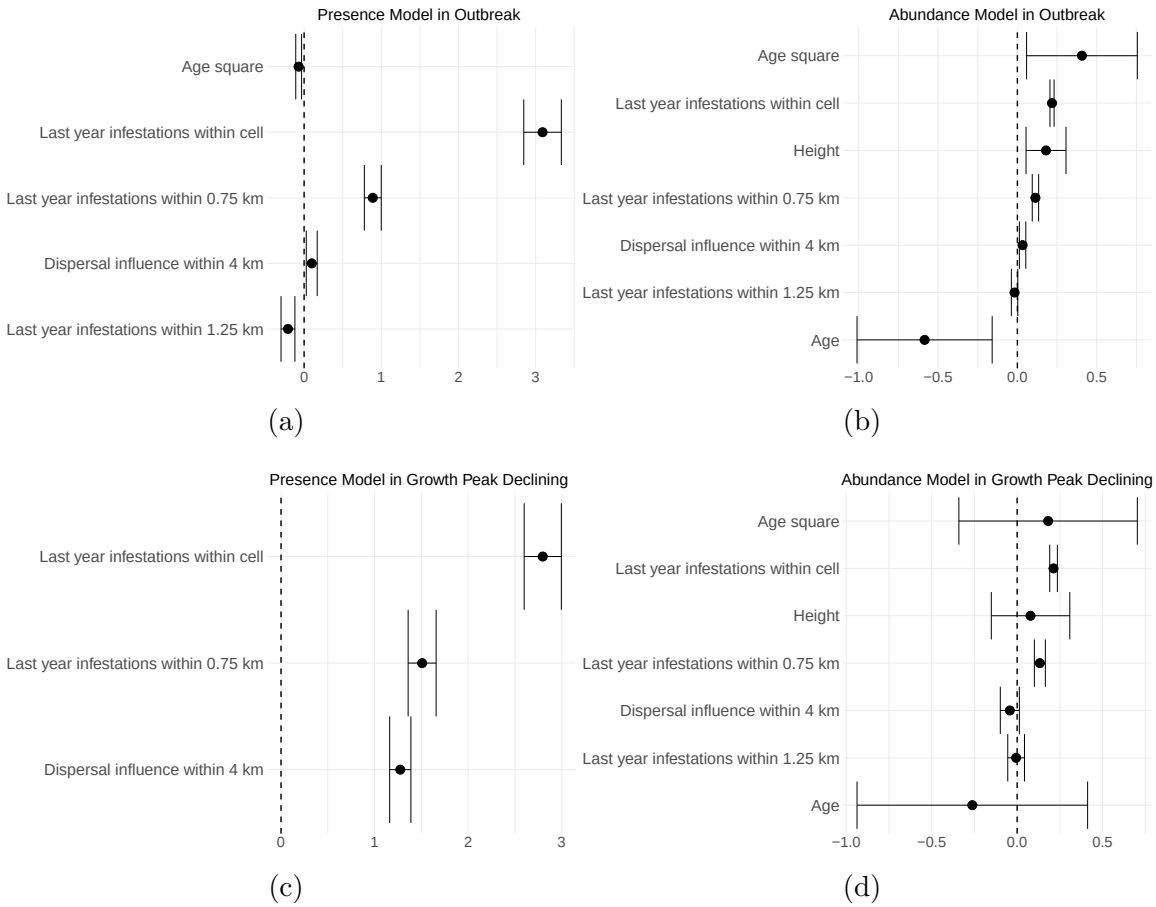


Figure 2.11: Coefficients of age, age^2, height, last year infestations within cell, within 0.75 km, 1.25 km and dispersal impacts within 4 km and their 95% confidence intervals in the (a) Presence model in the ZINB model in the outbreak stage (b) Abundance model in the ZINB model in the outbreak stage (c) Presence model in the ZINB model in the growth peak declining stage (d)Abundance model in the ZINB model in the growth peak declining stage. Covariates not shown in the plots indicate zero values.

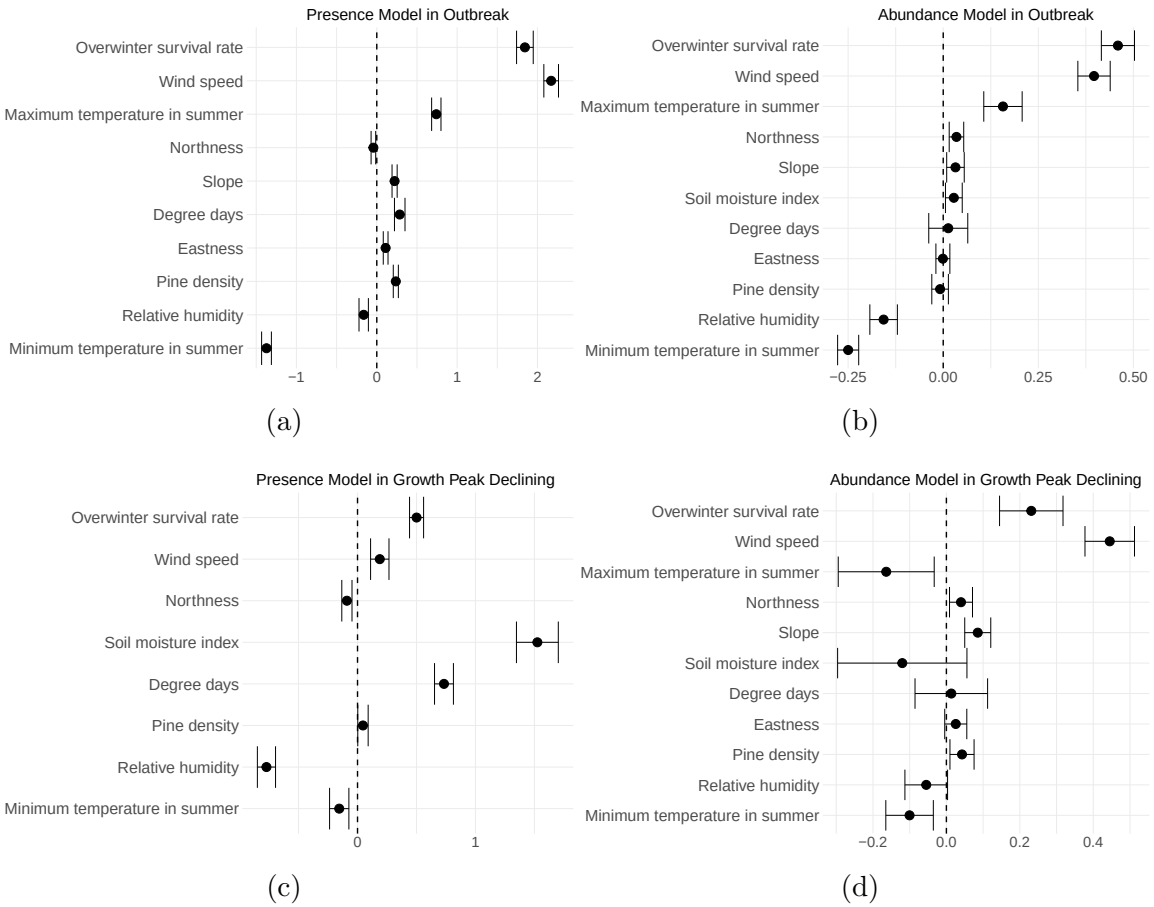


Figure 2.12: Coefficients of covariates not presented in Figure 2.11 and their 95% confidence intervals in the (a) Presence model in the ZINB model in the outbreak stage (b) Abundance model in the ZINB model in the outbreak stage (c) Presence model in the ZINB model in the growth peak declining stage (d)Abundance model in the ZINB model in the growth peak declining stage. Covariates not shown in the plots indicate zero values.

2.3.2 Model Selection

Based on the findings presented in Section 2.4.1, we have revised our covariates. This revision involved excluding the age square and height from our analysis. Additionally, we have consolidated the covariates representing last year's infestations within 0.75 km and 1.25 km, along with the dispersal influence within 4 km, into one covariate. This new covariate encapsulated the dispersal influence within 4 km but specifically excluded the area within the cell. The estimation method for this updated covariate remained similar to the previously used approach for calculating the dispersal influ-

ence within 4 km, albeit without considering last year's infestations within cell, 0.75 km, and 1.25 km.

The ZINB model exhibited the highest likelihood and the lowest BIC value for both stages, as shown in Table 2.4 and 2.5. Furthermore, the residual plots depicted in Figure 2.13 demonstrated that the ZINB model better fitted the data and had more consistent variance, although the dense patterns still suggested the presence of spatial autocorrelation. Similarly, as illustrated in Figure 2.14, the Q-Q plots for the ZINB model displayed a better fit within the 95% confidence interval and showed improved alignment with the $y = x$ line. Consequently, we have decided to select the ZINB model for further analysis.

Model	log-likelihood(Δ log-likelihood)	BIC(ΔBIC)
P	-350775.7(0)	701713.7(558661.1)
ZIP	-179733.8(171041.9)	359705.7(216653.1)
ZINB	-71364.05(279411.7)	143052.6(0)

Table 2.4: Table of log-likelihood and BIC of P, ZIP, and ZINB in the outbreak stage.

Model	log-likelihood(Δ log-likelihood)	BIC(ΔBIC)
P	-117683.5(0)	235507.6(176056.4)
ZIP	-57662.2(60021.3)	115562.4(56111.22)
ZINB	-29579.56(88103.94)	59451.18 (0)

Table 2.5: Table of log-likelihood and BIC of P, ZIP, and ZINB in the growth peak declining stage.

Model	Outbreak	Growth Peak Declining
P	0.99	0.99
ZIP	(Presence model: 1, Abundance model: 0.65)	(Presence model: 0.95, Abundance model: 0.43)
ZINB	(Presence model: 1, Abundance model: 0.79)	(Presence model: 1, Abundance model: 1)

Table 2.6: The optimal weights of the penalties in P, ZIP, and ZINB models (weights in the model with the smallest BIC).

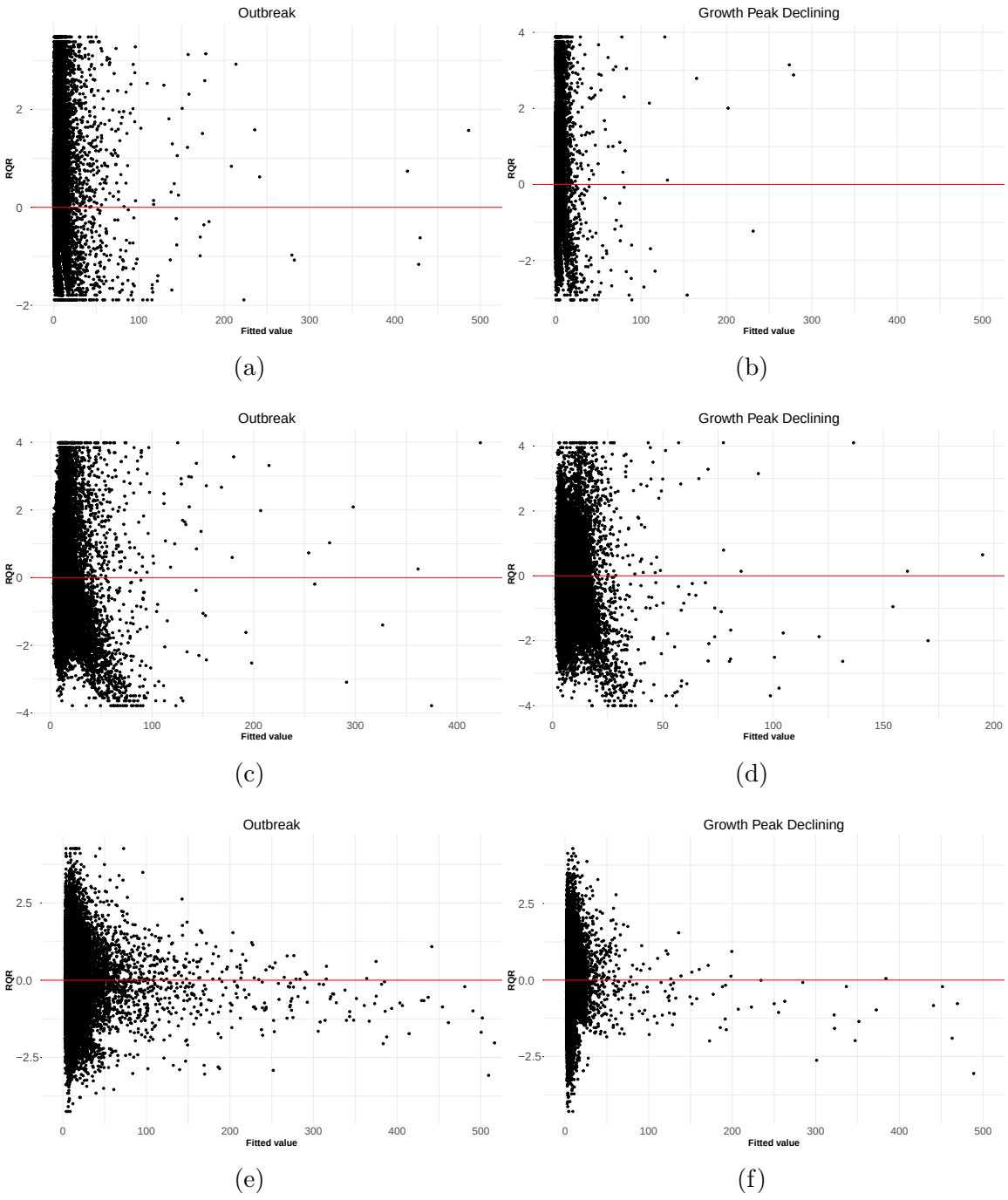
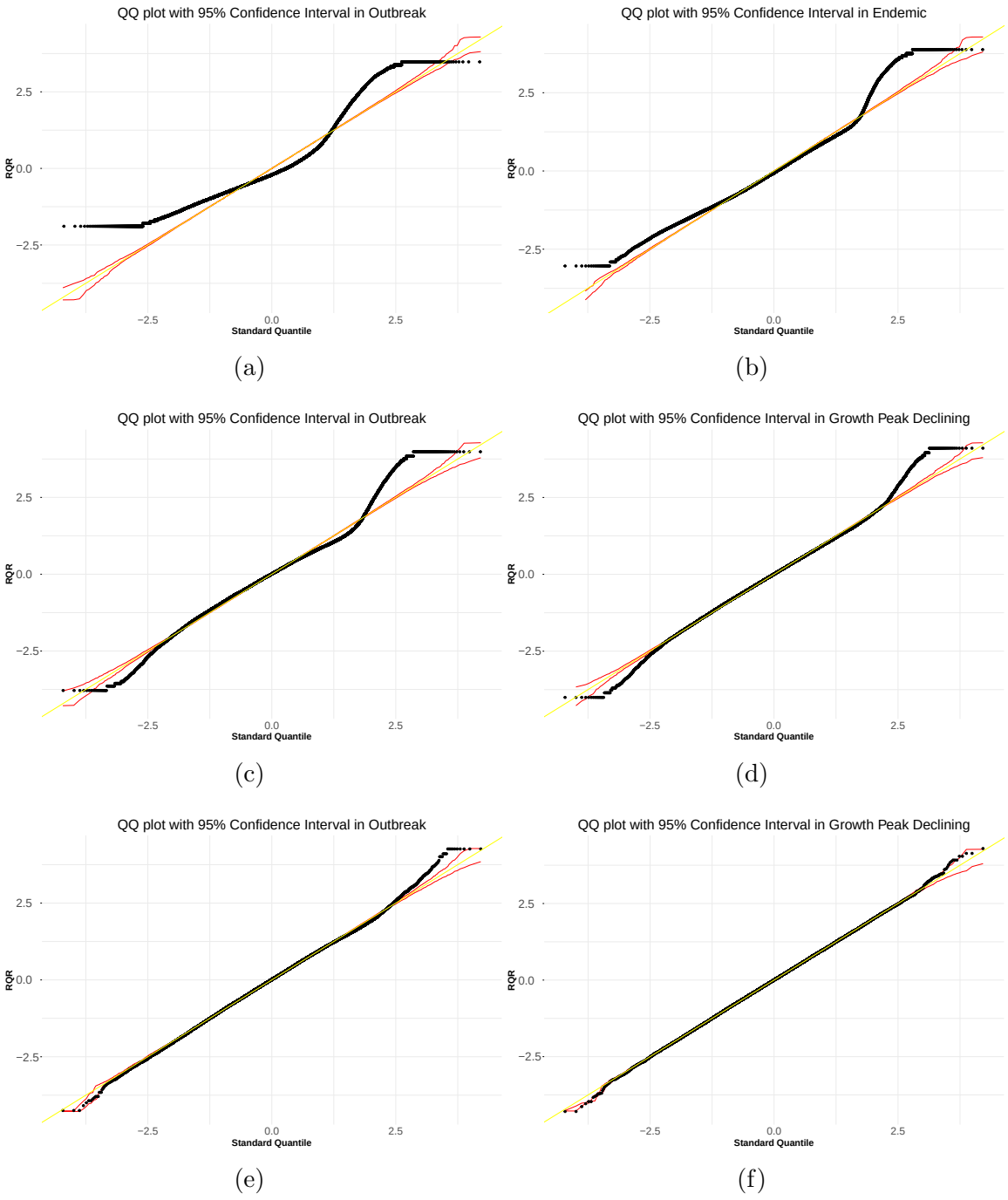


Figure 2.13: (a) Residual plot of Poisson model in the outbreak stage (b) Residual plot of Poisson model in the growth peak declining stage (c) Residual plot of ZIP in the outbreak stage (d) Residual plot of ZIP in the growth peak declining stage (e) Residual plot of ZINB in the outbreak stage (f) Residual plot of ZINB in the growth peak declining stage.



(a) (b) (c) (d) (e) (f)

Figure 2.14: (a) QQ-plot of Poisson model in the outbreak stage (b) QQ-plot of Poisson model in the growth peak declining stage (c) QQ-plot of ZIP in the outbreak stage (d) QQ-plot of ZIP in the growth peak declining stage (e) QQ-plot of ZINB in the outbreak stage (f) QQ-plot of ZINB in the growth peak declining stage (the yellow line is y=x).

2.3.3 Model Validation

The optimal weights for the penalties were detailed in Table 2.7. We have reported the accuracy and RMSE in Table 2.8, where a good model is characterized by high accuracy and low RMSE. These two statistical indices showed marked improvements when using the ZINB model. The confusion matrices for predictions in 2012 and 2013, as seen in Tables 2.9, 2.10, and 2.11, further demonstrated that the ZINB model achieves higher accuracy, particularly in predicting zero infestation cases.

The prediction plots for 2012 and 2013, when compared to the actual severity, were illustrated in Figures 2.15 and 2.16. It is important to note that due to the fluctuation of the coverage of aerial surveys across different years, the exact number of cells represented in the plots varies slightly. These plots revealed that the ZINB model outperformed the other two models in predicting infestations for 2012. For 2013, its predictions closely aligned with those of the ZIP model.

(a)

α
0.99

(b)

$\alpha_{abundance}$	$\alpha_{presence}$
1	0.99

(c)

$\alpha_{abundance}$	$\alpha_{presence}$
0.68	1

Table 2.7: The optimal weights for the two penalties have been identified for (a) P, (b) ZIP, and (c) ZINB.

Statistical index	Poisson	ZIP	ZINB
Accuracy	0.21	0.64	0.73
RMSE	1.54	1.18	1.07

Table 2.8: The estimated statistical indexes of the three models .

Table 2.9: Confusion matrix of the Poisson model (containing 2012 and 2013)

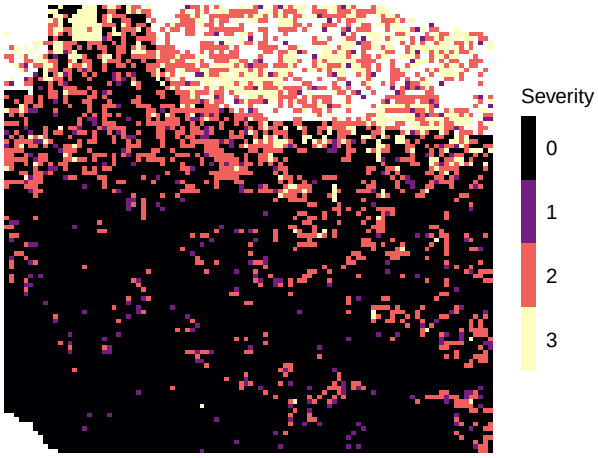
	predicted 0	predicted 1	predicted 2	predicted 3	Row Sum
observed 0	466	3271	8471	303	12511
observed 1	0	60	444	72	576
observed 2	0	122	2510	523	3155
observed 3	0	5	707	731	1443

Table 2.10: Confusion matrix of the ZIP model (containing 2012 and 2013)

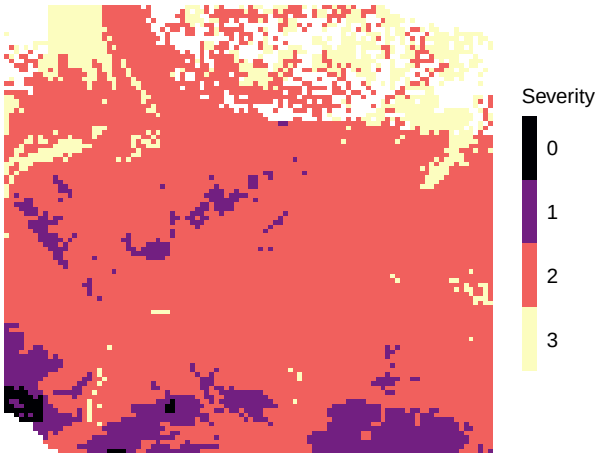
	predicted 0	predicted 1	predicted 2	predicted 3	row sum
observed 0	9428	0	2918	165	12511
observed 1	336	0	213	27	576
observed 2	1719	0	1117	319	3155
observed 3	372	0	350	721	1443

Table 2.11: Confusion matrix of the ZINB model (containing 2012 and 2013)

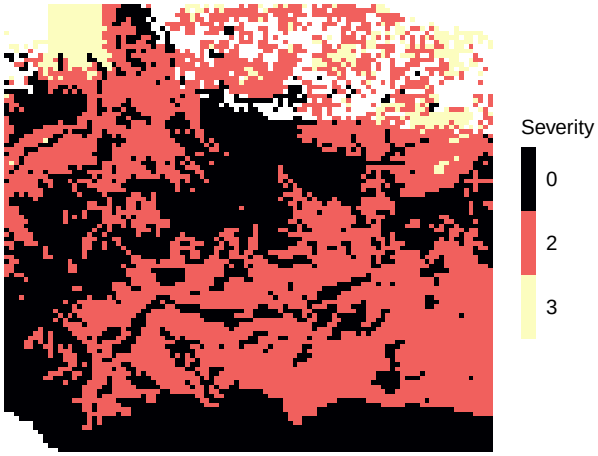
	predicted 0	predicted 1	predicted 2	predicted 3	row sum
observed 0	11970	0	480	61	12511
observed 1	496	0	72	8	576
observed 2	2608	0	445	102	3155
observed 3	693	0	283	467	1443



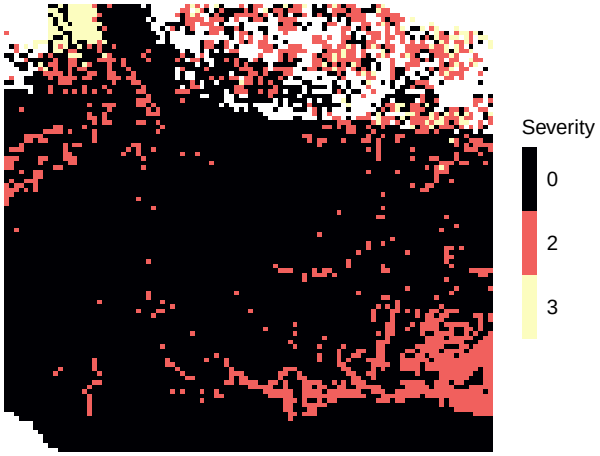
(a) Real severity in 2012



(b) Predicted severity in 2012 using P

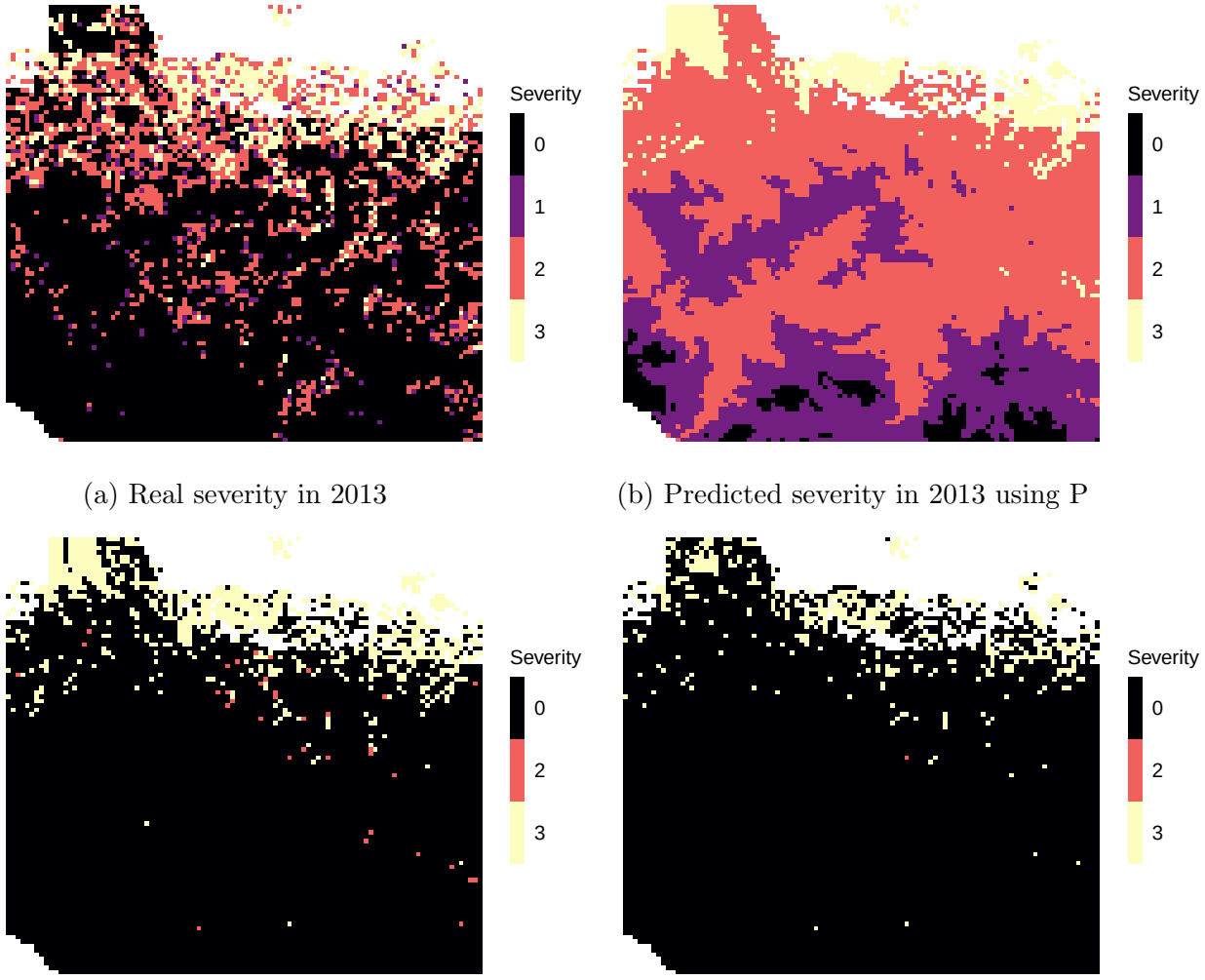


(c) Predicted severity in 2012 using ZIP



(d) Predicted severity in 2012 using ZINB

Figure 2.15: Comparisons of Real and Predicted Severity for the P, ZIP and ZINB models in 2012



(a) Real severity in 2013

(b) Predicted severity in 2013 using P

(c) Predicted severity in 2013 using ZIP

(d) Predicted severity in 2013 using ZINB

Figure 2.16: Comparisons of Real and Predicted Severity for the P, ZIP and ZINB models in 2013

2.4 Discussion

In this chapter, we have established the biological factors and models that are further explored in the subsequent two chapters. Our results demonstrated that the hierarchical Zero-Inflated Negative Binomial (ZINB) model better fits the infestation distribution, outperforming both Poisson (P) and Zero-Inflated Poisson (ZIP) models. However, our hypothesis testing revealed that the Poisson model identified more significant covariates compared to ZIP and ZINB models, likely due to its inadequate

fit for our data.

Focusing on the two hierarchical models, the hypothesis testing suggested that variables such as the square of the pine's age and its height were less critical. The expected quadratic relationship between pine age and resistance to infestations was not evident. We attempted to compute height information from a sigmoid age-height model using limited pine height data from the AVI dataset. However, we found that the simulated height did not show significance in the hierarchical models. For beetle pressure, only last year infestations within cells was consistently significant in both ZIP and ZINB models. We faced challenges in distinguishing the significance levels of the other three covariates related to dispersal at various distances. To address this, we combined three variables – last year infestations within 0.75 km and 1.25 km, and the dispersal influence within 4 km excluding neighboring counts within 1.25 km – into a single covariate. This was achieved using an estimation approach similar to the one previously employed for assessing dispersal influence within 4 km.

During model selection, ZINB demonstrated the best fit, as indicated by the highest log-likelihood, the lowest BIC, and the RQR plots that confirmed ZINB's better fit to the data. In the model validation section, the ZINB model demonstrated the highest accuracy, with the lowest RMSE. Its confusion matrix further illustrated its predictive ability compared to the other two models. Prediction maps also provided a clear comparison of the performances of all three models. Overall, the modified covariates and the ZINB model are employed in the next two chapters, building on these findings.

The model employed in this study differs from those used in previous studies, such as those by Kunegel-Lion and Lewis (2020), Preisler et al. (2012), Srivastava and Carroll (2023), and Wulder et al. (2006), as it estimates both the presence and abundance of infestations. Due to the wide range of infestation numbers, from zero to hundreds, the accuracy of our models was considerably limited. For example, the Zero-Inflated Negative Binomial (ZINB) model showed a maximum accuracy of

73% when predicting 2012 and 2013 across four severity classes. However, despite this trade-off in accuracy, our model provided a more comprehensive overview of the impacts of covariates on both the presence and abundance of infestations, thereby offering enhanced interpretability.

Additionally, our approach to handling multicollinearity and variable selection differs from previous methods. Unlike the use of the variance inflation factor (VIF) and best subset selection method in Kunegel-Lion and Lewis (2020), and the consideration of fewer covariates to avoid multicollinearity issues in Srivastava and Carroll (2023), our study employed penalties. This method allows for a faster process while still retaining comprehensive information.

Besides, as detailed in Appendix A.1, hypothesis testing revealed a significant level of spatial autocorrelation within the study area. The residual plots in Figure 2.9 showed patterns grouping together, indicative of spatial autocorrelation present across all three models. This finding underscored the need for cautious interpretation of our results, particularly regarding the significance of covariates, as spatial autocorrelation can lead to an underestimation of their confidence intervals (Legendre, 1993). Furthermore, standard regression models are limited in their ability to capture population dynamics, posing challenges in predicting future trends and applicability to different areas. The development of new spatial regression models could yield more realistic covariates and higher prediction accuracy (Chi & Zhu, 2008).

Overall, we have focused on testing our covariates and selecting the most suitable model for the subsequent chapters. The hierarchical model, incorporating a Bernoulli distribution to classify the presence of zero infestations, has demonstrated higher suitability compared to the other two models. In the following two chapters, we employ the ZINB model to examine the impacts of different covariates on mountain pine beetle population dynamics.

Chapter 3

The influence of pine host on beetle's attack

3.1 Introduction

The Mountain Pine Beetle (MPB), considered one of North America's most destructive native insects, has caused the death of millions of lodgepole pines in British Columbia (Government of Canada, 2021; Westfall & Ebata, 2007). While the primary host of mountain pine beetle is the lodgepole pine, it inhabits most species of pines within its range (Amman, 1978; Safranyik & Carroll, 2007). The range of mountain pine beetle is mainly governed by temperature and seasonal suitability (Jenkins et al., 2001; Logan & Bentz, 1999; Logan & Powell, 2001; Taylor & Safranyik, 2003). Further exacerbated by global warming, suitable habitats for MPB have expanded eastward and northward, enabling these beetles to infest pine species in regions where they had not previously been present (Safranyik & Wilson, 2006; Taylor & Safranyik, 2003). The outbreak in the late 1990s and 2000s enabled them to breach the Rocky Mountains into Alberta, thereby posing a threat to the boreal forest in Saskatchewan and beyond (Nealis & Peter, 2008; Robertson et al., 2009a; Safranyik et al., 2010).

In central and northwestern Alberta, lodgepole pine hybridizes with jack pine, a major component of Canada's boreal forests, with habitats extending from northern Alberta to eastern Canada (Rudolph & Laidly, 1990). Jack pine is commonly regarded as closely related to lodgepole pine (Moss, 1949; Rice et al., 2007b). However, in

regions where their ranges overlap, there exists a distinct elevation separation between jack pine and both lodgepole and hybrid pine (Rweyongeza et al., 2007). Hybrid pine is distinguished from jack pine as it is more similar to lodgepole pine than jack pine (Rweyongeza et al., 2007; Yang et al., 1999).

Although the range of lodgepole pine overlaps with that of jack pine, different pines prefer different habitats (Critchfield, 1985). Jack pine thrives in excessively drained, nutrient-poor lands and is most commonly found on soils of the podzolic order, ranging from coarse to fine sands and gravels (Kenkel et al., 1997). In comparison to lodgepole pine, jack pine typically has a shorter lifespan, about 80-100 years for a medium-sized tree. It also grows shorter, reaching heights of 15-25 m, and is thinner, with a diameter at breast height of approximately 20-30 cm (Kenkel et al., 1997). Previous studies have indicated that jack pine could be a suitable host with limited defensive capabilities (Clark et al., 2014; Cullingham et al., 2011). However, cold winter temperatures can hinder brood development, significantly limiting the range of beetles (Taylor & Safranyik, 2003). The colder climate in jack pine forests, combined with the trees' thinner phloem, can lead to high winter mortality rates among beetles (Amman, 1976; Chang, 1954; Rosenberger et al., 2017a). The similarities and differences between lodgepole pine forests and jack pine forests raise questions about how population dynamics of MPB in jack pine forests may differ from those in lodgepole pine forests.

In this chapter, our goal was to measure the comparative performance of the mountain pine beetle in lodgepole and jack pines. For this goal, we made use of a hierarchical model with selection process detailed in Chapter 2. We estimated the likelihood of jack pine infestation and the reproduction rate in infested jack pines, in comparison to lodgepole pine. This estimation took into account the physical environment, as well as the biology of both the beetle and the trees. Based on the limitations inherent in the natural environment of jack pine forests and the findings reported in Bleiker et al. (2023), we anticipated a lower likelihood and reproduction rate of beetles in

jack pine compared to lodgepole pine.

We assessed the risk of infestation using relative risk, which is derived from absolute risk. Absolute risk is defined as the probability of an event occurring (Nelson et al., 2008; Suter II, 2016), whereas relative risk is the ratio of the absolute risk of one event over another (Nelson et al., 2008). In our study, the 'event' is the occurrence of infestations: the first event being infestation in a cell with lodgepole pine, and the second in a cell with jack pine. We focused on relative risk rather than absolute risk, due to the varying precision levels of environmental and ecological covariates as well as our less precise model. Let y represent the response variable being the number of infested trees and θ the vector of parameters in the model. We calculated absolute risk as $\Pr(y > 0|\theta)$. The relative risk is then $\Pr(y > 0|\theta\&\text{lodgepole pine})/\Pr(y > 0|\theta\&\text{jack pine})$. Furthermore, the reproduction rate was analyzed as the average productivity of jack pine relative to lodgepole pine, assessed by quantifying the impact of past infested jack pine on the new host tree.

Our analyses were done by a hierarchical model, using the number of pines killed in the current year as the response variable, and incorporating both geographical, climatic and ecological covariates. In addition to the set of covariates used in the model selection and validation in the previous chapter, we included a covariate 'pine identity' that differentiates between lodgepole and jack pine (Burns et al., 2019). We selected an area in central Alberta, where we can observe the genetic transition from lodgepole pine to jack pine with similar numbers of cells dominated by lodgepole pine, lodgepole× jack hybrid pine and jack pine as shown in Figure 3.1. We further segmented population in this area into different stages, aligning with the dynamics observed, and created distinct models for each stage (Kunegel-Lion & Lewis, 2020). We trained our models on a dataset that covers a 14-year period from 2007 to 2020. The outbreak stage was modeled using data from 2007 to 2014, while the endemic stage was based on data from 2015 onwards, as depicted in Figure 3.2. We only focused on the outbreak stage because we did not get significant results in the endemic stage.

Similar to the approach in the second chapter, we utilized the Ridge regression and the Smoothly Clipped Absolute Deviation (SCAD) penalty to mitigate multicollinearity and facilitate variable selection (Fan & Li, 2001; Hoerl & Kennard, 1970). The analysis was conducted in the `R` programming language using the `package-mpath` and `package-pscl`. In line with the penalties applied, our null hypothesis posits that all coefficients except intercept are negligible. The covariates conform to the null hypothesis if their coefficients prune to 0 or if their 95% confidence intervals encompass 0.

Overall, this study conducted a quantitative analysis of the impacts of different pine species on infestations, focusing particularly on the potential differences between lodgepole pine and jack pine forests. We explored the possibility of range expansion of mountain pine beetles into jack pine forests by assessing relative risk and the reproduction rate. To facilitate this, we incorporated a range of geographical, climatic and ecological covariates into our model.

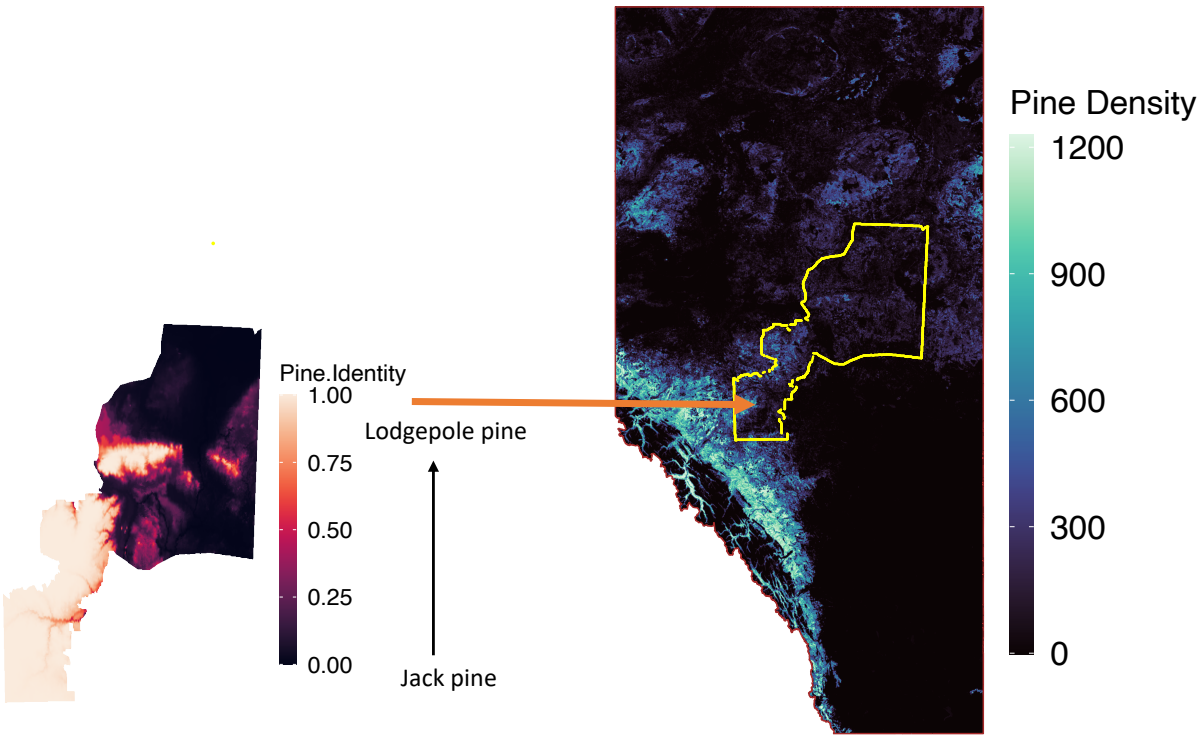


Figure 3.1: The selected area (marked by yellow color).

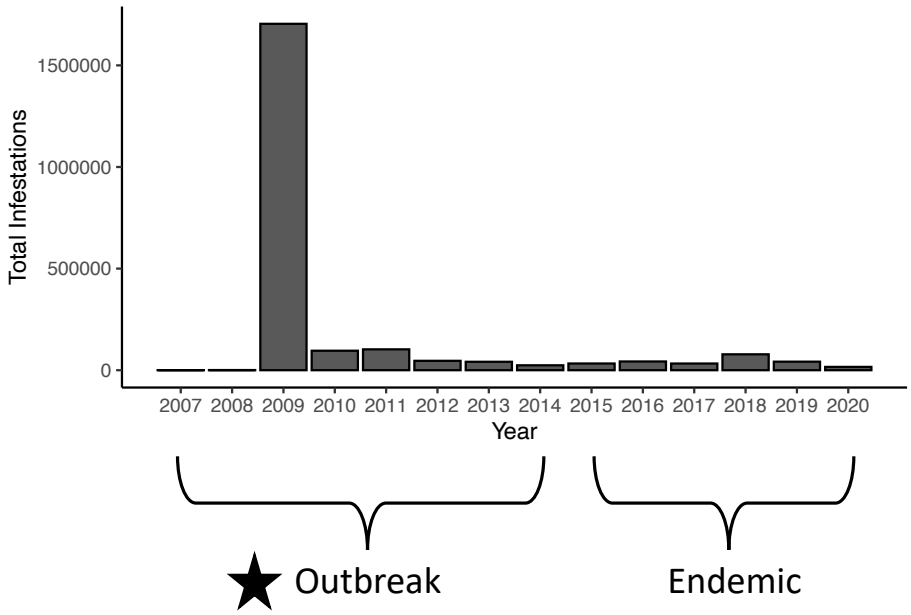


Figure 3.2: The infestations observed from 2007 to 2020 in a selected area in central Alberta.

3.2 Method

3.2.1 Study area and data preparation

To assess the risk associated with infestations in lodgepole pine versus jack pine, our focus area differed from that of Chapter 2. We have selected a region within the central forests where a clear transition in tree composition from lodgepole pine to jack pine is observable when traveling from south to north. Moreover, the northern extent of this area was determined by the locations of infestations identified in aerial surveys. The southern extent was established to ensure that the number of cells for the three pine species, as identified by pine identity, is closely balanced across the cells. Pine identity, an index ranging from 0 to 1, was developed through a logistic model incorporating environmental covariates (Burns et al., 2019). If its value ranges from 0 to 0.1, the cell is jack pines. When its value is greater than 0.9, the cell is lodgepole pine. All other covariates remained consistent with those used in the model selection and validation sections of Chapter 2. Similar to Chapter 2, our covariates, which include geographical, climatic, and ecological data, were sourced from Alberta Environment (Agriculture and Forestry, 2020), Alberta Vegetation Inventory (AVI) (Agriculture, Forestry and Rural Economic Development, 2022), and National Forest Inventory (NFI) (Beaudoin et al., 2014), as summarized in Table 3.1. Our research spanned the period from 2007 to 2020, employing a grid consisting of 196587 cells measuring 500 m $\times$ 500 m, as depicted in Figure 3.1.

Variables	Description
Northness	Face north (1) and south (-1) (cosine of aspect).
Eastness	Face east (1) and west (-1) (sine of aspect).
Slope	Slope of the cell (Burrough et al., 2015).
Pine Identity (Q)	Genetic score used to identify pine type. It identifies trees as jack pine if $Q <$ 0.1, as lodgepole pine if >0.9, and between is hybrid pine (Burns et al., 2019).
Pine density	Pine density (stems/cell; including jack pine, hybrid pine and lodgepole pine) (Beaudoin et al., 2014). The method is described in Chapter 2.
Age	Average age of the leading pines in each cell (Beaudoin et al., 2014).
Maximum temperature in summer	Highest daily temperature in summer in the current year.
Minimum temperature in summer	Minimum daily temperature in summer in the current year.
Relative Humidity	Average relative humidity from March to May in the current year.
Wind Speed	Average wind speed at 2 meters in July and August in the current year.
Soil Moisture Index	Average soil moisture index from May to July, estimating the water supply for tree growth; current year (Hogg et al., 2013).
Degree Days	Cumulative degree days above 5.5°C from September of the previous year to August in the current year (Safranyik et al., 1975).
Overwinter Survival Rate	Probability of survival over the winter; the winter before the current year summer (%) (Régnière & Bentz, 2007).
Last year infestations	Number of pines killed last year within the cell.
Dispersal impacts within 4 km	Expected number of beetles flying from all directions to the center cell excluding the counted within cell infestations (Carroll et al., 2017).
Infestations (response variable)	Number of pines killed this year.

Table 3.1: The covariates and response variable considered in Chapter 3.

3.2.2 Fitting and Analysis

In Chapter 2, we identified that the Zero-inflated Negative Binomial (ZINB) model provides the best fit for the distribution of infested trees. We defined β and γ as the coefficient vectors for the abundance (following Negative Binomial distribution) and presence (following Bernoulli distribution) models, respectively. We considered the same covariates in two submodels and used the pine identity value (Q) to determine the pine species in the cell. We further defined μ and π as the mean for the abundance model and the probability of zero infestation in the presence model. We used Ridge regression to handle multicollinearity (Hoerl & Kennard, 1970) and SCAD penalty to simplify the model complexity and concentrate specifically on the most significant

covariates (Fan & Li, 2001). Two penalties were simultaneously applied, with their respective influences weighted such that their combined weights sum to 1. We did a grid search for the optimal weights that minimize Bayesian Information Criteria (BIC) (H. Wang et al., 2007). We assigned weights to one penalty ranging from 0.01 to 1 in increments of 0.01, while the weight of the other penalty was set to be 1 minus the first penalty's weight. The ZINB model comprises two sub-models: a logistic model and a Negative Binomial model, with each model having two penalties applied to it. When incorporating covariates into the model, we standardized their values. In the outbreak stage, we had 890504 observations and the endemic stage contained 1123473 observations. The fewer cells in the outbreak stage was because, within the chosen area, the extent of the aerial survey coverage fluctuates annually. The aerial survey data showed that the number of cells detected within the survey range was lower from 2007 to 2011 compared to the higher detection rates observed from 2012 onwards.

In order to evaluate the likelihood of infesting a jack pine as compared to a lodgepole pine, we examined both the hypothesized absolute and relative risks. Let y_i be the number of infested trees in the cell i. Let x_i be the vector containing all x_{ij} values, each representing a value of covariate j in cell i. We calculated the hypothesized absolute risk through varying the value of the covariate pine identity, while keeping all other covariates at their original values. We used Q to represent pine identity in the following equations. The absolute risk in cell i, assuming the pines to be jack pine, hybrid pine, and lodgepole pine, is represented by the following equations respectively:

$$\mathrm{P}(y_i > 0 | 0 \leq x_{iQ} \leq 0.1) \tag{3.1}$$

$$\mathrm{P}(y_i > 0 | 0.1 < x_{iQ} < 0.9) \tag{3.2}$$

$$\mathrm{P}(y_i > 0 | 0.9 \leq x_{iQ} \leq 1) \tag{3.3}$$

We further estimated the hypothesized relative risk, which is derived from the absolute

risks. The relative risk of a cell containing lodgepole pine relative to jack pine is represented as:

$$\frac{\mathrm{P}(y_i > 0 | 0.9 \leq x_{iQ} \leq 1)}{\mathrm{P}(y_i > 0 | 0 \leq x_{iQ} \leq 0.1)} \tag{3.4}$$

We also calculated the reproduction rate of jack pine relative to lodgepole pine. We added an additional parameter s into our hierarchical model, measuring the influence of the past jack pine infestations to the nearby new host tree while using the influence of past infestations in non-jack pine as the benchmark. We assumed that the influence of past infested non-jack pines is 1 and estimated the influence of past infested jack pine by multiplying the parameter s with the observed jack pine trees and dispersal influence. We assumed that the influence of one infestation, regardless of the pine species, extends only up to 4 kilometers. The new introduced parameter s ranging from 0 to ∞ specifically affected two covariates: last year infestations and dispersal impacts within 4 km.

Let the first m+1 covariates, numbered from 0 to m, be those that are not affected by the new parameter s. The subscript 0 indicates the intercept. We defined last year infestations and dispersal impacts within 4 km as the m+1 and m+2 covariates, respectively. We divided these two covariates, into two categories using superscripts: 'not jack' for infested non-jack pine and 'jack' for infested jack pine. We also denoted $\mathbb{1}$ as an indicator function with a bracket indicating the condition for having a value 1. For the cell i, we estimated μ_i and π_i by

$$\begin{aligned}
\log(\mu_i) &= \beta_0 + \sum_{j=1}^{m} x_{ij}\beta_j + \mathbb{1}(Q \leq 0.1) \cdot s \cdot x_{im+1}\beta_{m+1} \\
&\quad + \mathbb{1}(Q > 0.1) x_{im+1}\beta_{m+1} + (x_{im+2}^{\text{not jack}} + x_{im+2}^{\text{jack}} \cdot s)\beta_{m+2} \\
\operatorname{logit}(\pi_i) &= \gamma_0 + \sum_{j=1}^{m} x_{ij}\gamma_j + \mathbb{1}(Q \leq 0.1) \cdot s \cdot x_{im+1}\gamma_{m+1} \\
&\quad + \mathbb{1}(Q > 0.1) x_{im+1}\gamma_{m+1} + (x_{im+2}^{\text{not jack}} + x_{im+2}^{\text{jack}} \cdot s)\gamma_{m+2}
\end{aligned} \tag{3.5}$$

For simplicity, when estimating this new parameter s, we did not employ two penalties. The estimation of the parameter s was carried out by minimizing the negative

log-likelihood function, utilizing the 'BFGS' gradient function specified within the `optim()` function in R.

3.3 Results

The coefficients in the presence and abundance models correspond to the probability of infestations and the number of infested trees if present, respectively. Positive coefficients indicate that an increase in the value of a covariate is associated with an increased likelihood of infestations and a higher number of infested trees. In this section, we focused solely on the results from the outbreak stage. The reason for this is that the results in the endemic stage were insignificant for both the presence model (modeled with a Bernoulli distribution) and the abundance model (modeled with a Negative Binomial distribution), as detailed in Appendix B. The optimal weights for the penalties applied in these models in the outbreak stage were detailed in Table 3.2.

The values of the standardized coefficients were presented in Table 3.3, and we also displayed these coefficients along with their 95% confidence intervals in Figures 3.3 and 3.4. In both models, the coefficient values and confidence intervals for pine identity underscored its significance during an outbreak. Its values suggested a higher likelihood and number of infestations in lodgepole pine forest.

The presence model revealed more insignificant covariates: Pine density, northness, and easternness appear to have minimal importance. In the abundance model, the coefficients and their 95% confidence intervals suggested that tree age is not a significant factor, and northness has a lesser impact. This interpretation considered the potential underestimation of confidence intervals due to spatial autocorrelation (Legendre, 1993). The coefficient of wind speed in the abundance model indicated the importance of flight for new infestations in the central forest. In both models, the minimum temperature in summer, maximum temperature in summer and overwinter survival rate provided results that are less straight-forward to interpret.

The absolute and relative risks presented in Figures 3.5, 3.6, and 3.7 offered addi-

tional insights. As depicted in Figures 3.5 and 3.6, the probability of having at least one infested tree in both lodgepole and jack pine species was low, with the majority of cells having values less than 0.025. Compared to the absolute risk of jack pine shown in Figure 3.5, the risk for lodgepole pine was higher. Moreover, the relative risk illustrated in Figure 3.7 suggested that more than half of the cells are more likely to be infested if they contain lodgepole pine. The relative risk of lodgepole pine compared to jack pine had an average of 1.89 indicating that, on average, each cell is 1.89 times more likely to be infested if it contains lodgepole pine. It also had a median of 1.67 meaning that for 50% cells, lodgepole pine is 1.67 times more likely to be infested. Additionally, the parameter s had a value of 0.558 during the outbreak stage, demonstrating that infested jack pine has a lesser influence on the spread to nearby new host trees. It meant that the reproduction rate in jack pine is only 0.558 times that of lodgepole pine. Together, our findings suggested a lower likelihood of infestations in jack pine compared to lodgepole pine.

	weight in presence model	weight in abundance model
outbreak stage	0.93	1

Table 3.2: The optimal weights of SCAD penalty for the outbreak stage.

Covariates	Presence Model	Abundance Model
Northness	-0.07	0.091
Eastness	-0.102	-0.192
Slope	-0.332	0.183
Pine identity	3.089	0.691
Pine density	0.142	0.317
Age	1.096	0.042
Minimum temperature in summer	1.899	-1.35
Maximum temperature in summer	-2.192	0.394
Relative Humidity	-5.424	0.303
Wind Speed	0.696	0.863
Soil Moisture Index	-2.458	0.608
Degree Days	-2.331	0.51
Overwinter Survival Rate	-2.474	-0.609
Last year infestations	-28.138	1.05
Dispersal impacts within 4 km	1449.039	0.257

Table 3.3: The coefficients of the standardized covariates in the outbreak data model.

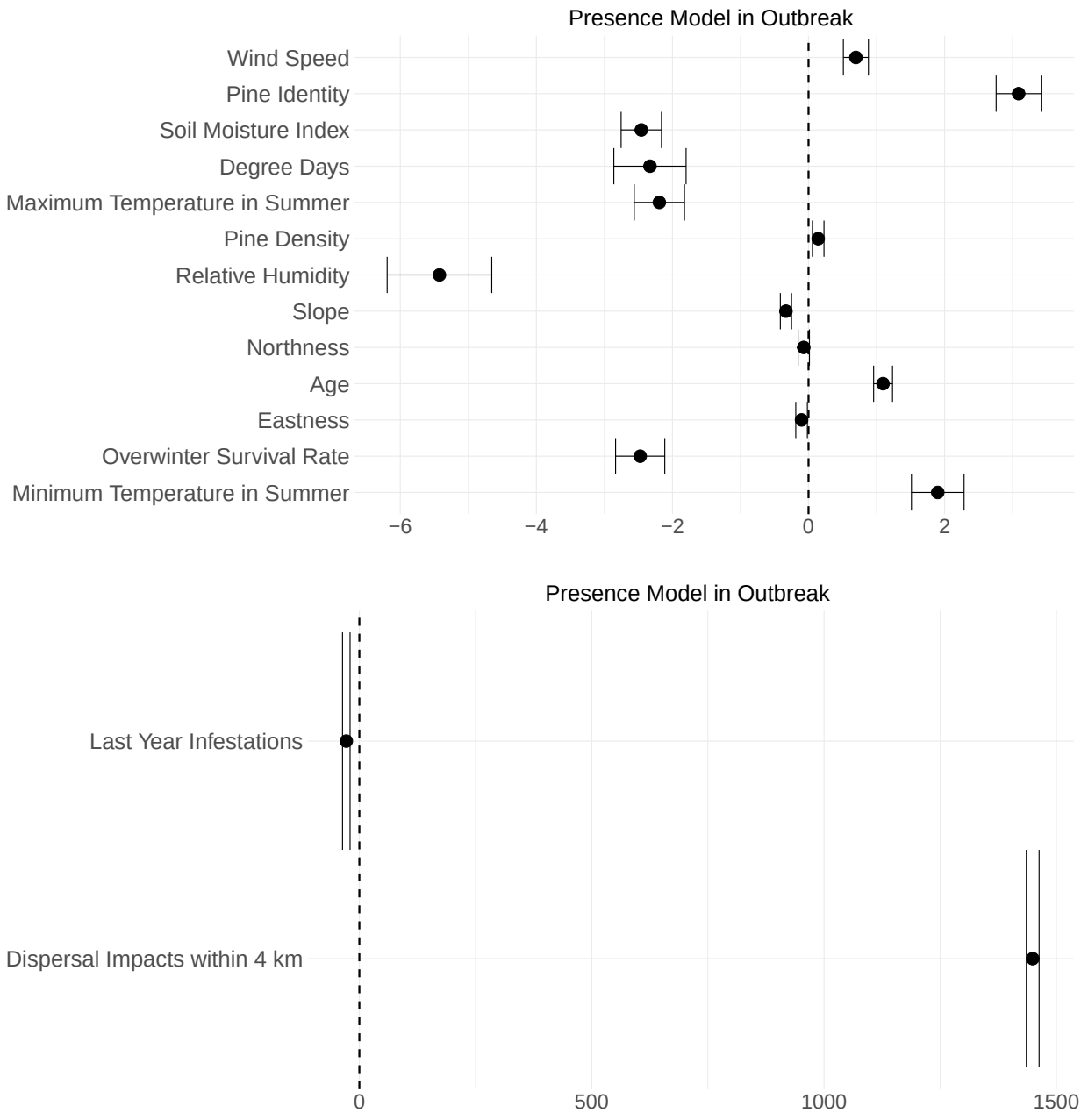


Figure 3.3: The 95% confidence intervals of the standardized non-zero coefficients of our covariates in the presence model in the outbreak stage.

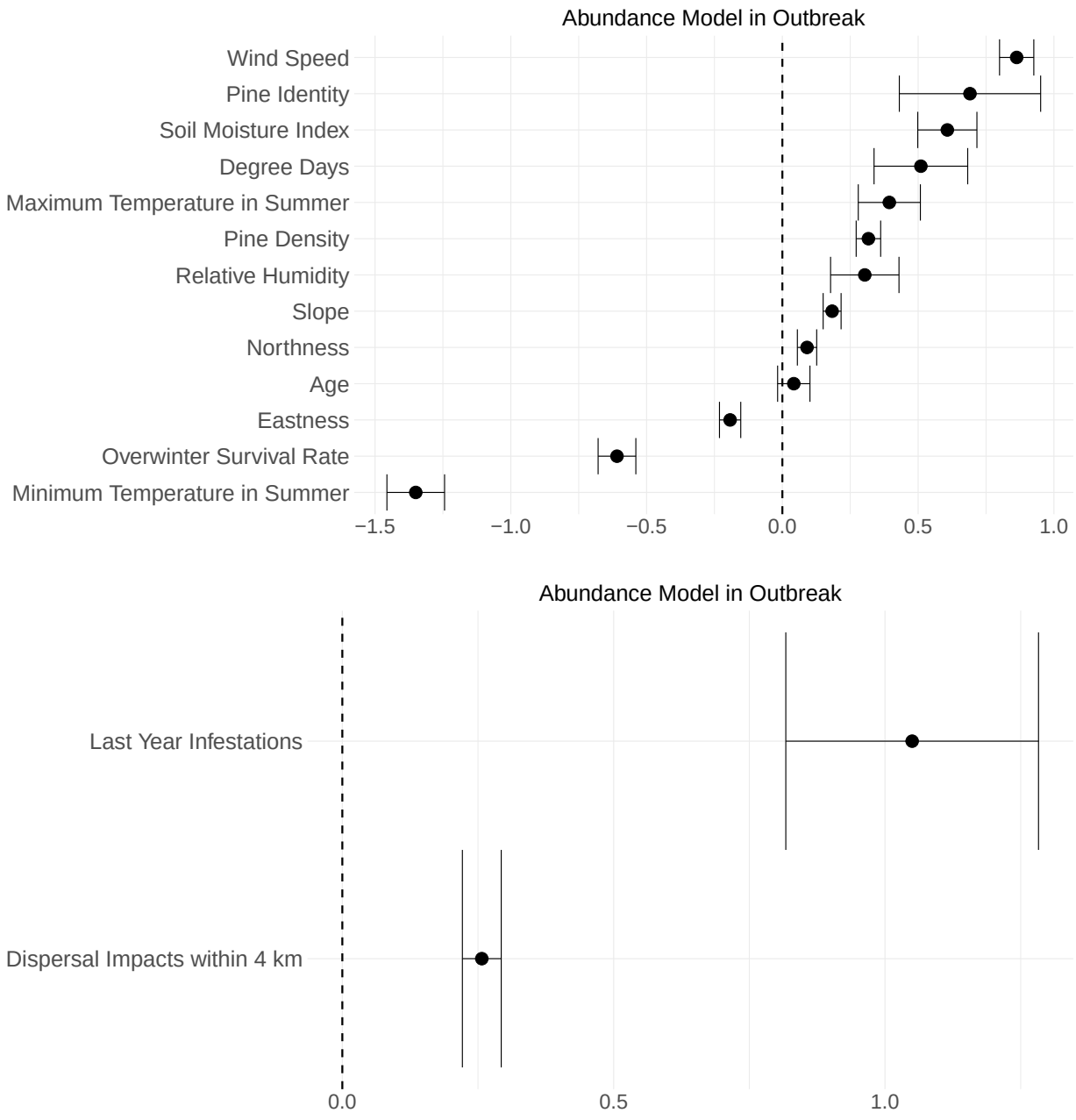


Figure 3.4: The 95% confidence intervals of the standardized non-zero coefficients of our covariates in the abundance model in the outbreak stage.

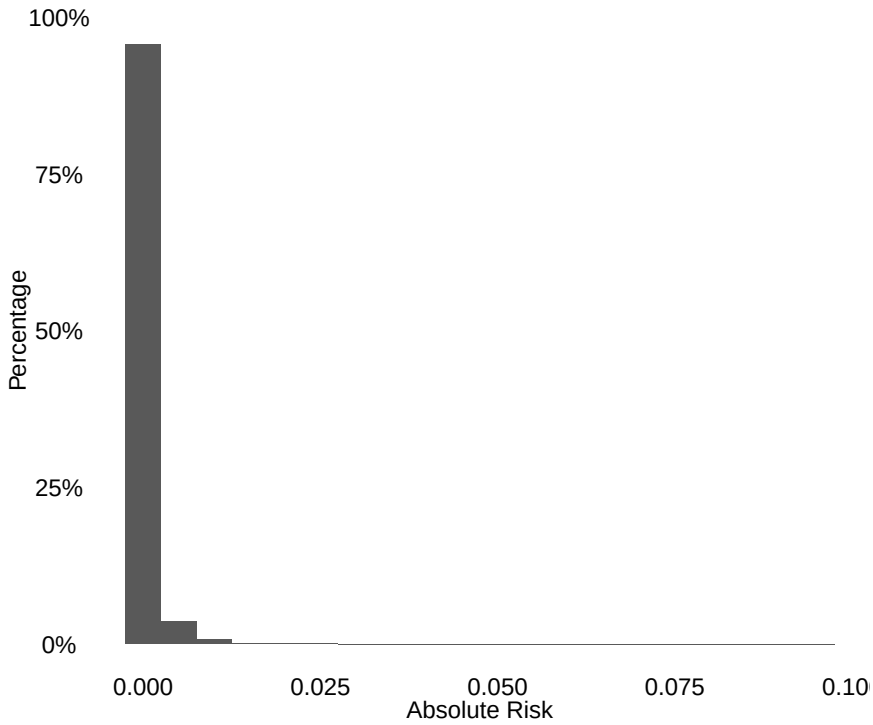


Figure 3.5: The histogram of the absolute risk showing the likelihood of jack pine infestation corresponding to equation 3.1 during the outbreak stage.

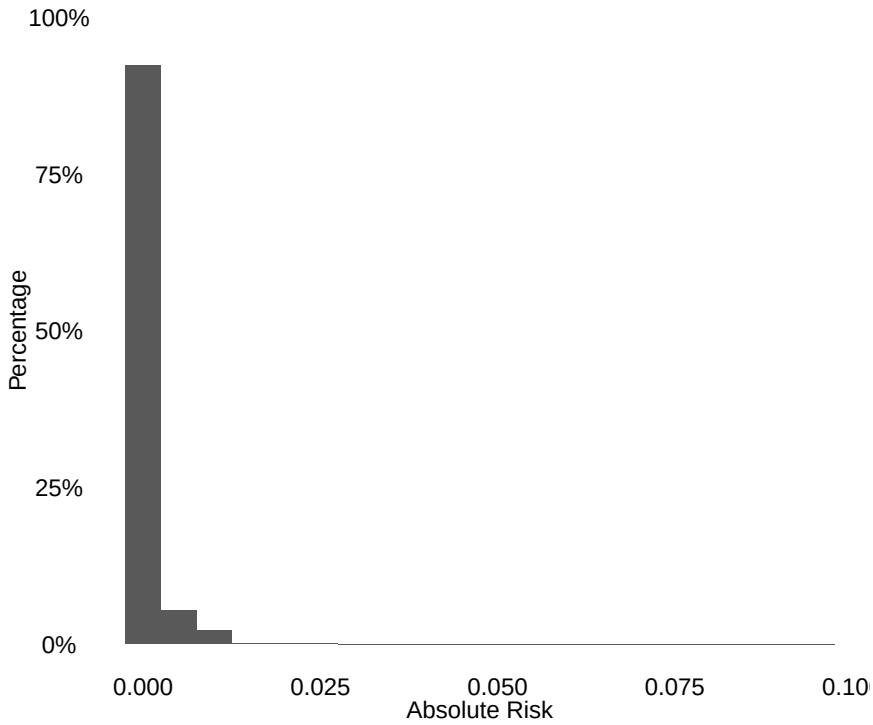


Figure 3.6: The histogram of the absolute risk showing the likelihood of lodgepole pine infestation corresponding to equation 3.3 during the outbreak stage.

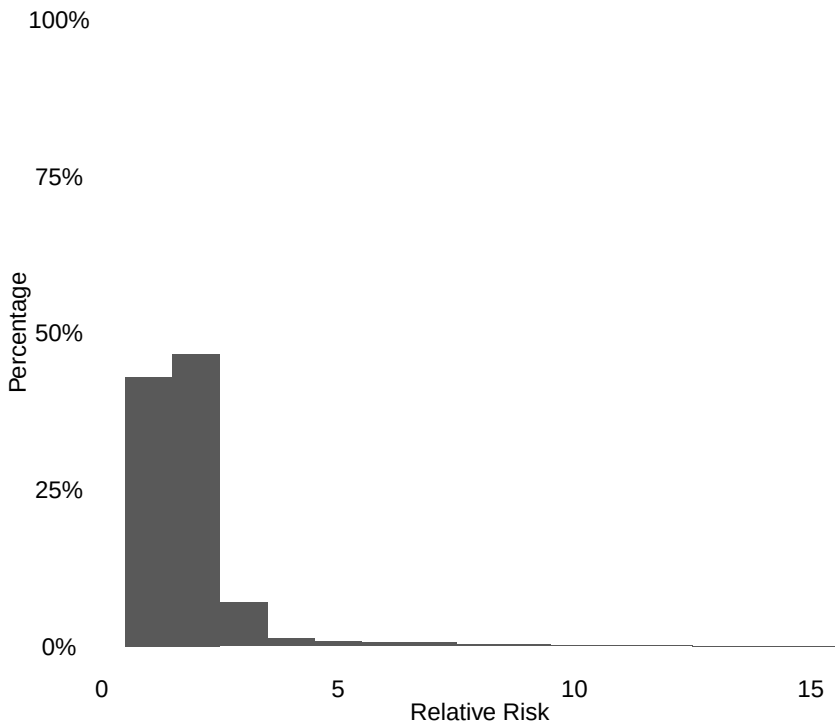


Figure 3.7: Histogram of the relative risk showing the ratio of the likelihood of lodgepole pine infestation to jack pine infestation corresponding to equation 3.4 during the outbreak stage.

3.4 Discussion

This study's findings suggested that the new host, jack pine, is a less suitable host compared to lodgepole pine. Our results confirmed the initial hypothesis that jack pine has a lower probability of infestation and rate of reproduction. This supported the findings reported in Bleiker et al. (2023), which state that the likelihood of infesting a jack pine is lower than lodgepole pine.

The newly introduced covariate, pine identity, played a significant role in our hierarchical model, exhibiting positive signs in both the presence and abundance models. It had the second largest coefficient value in the presence model and the third largest in the abundance model. These values highlighted the importance of pine species in the host selection process, indicating a strong preference for lodgepole pine. For other covariates, such as the minimum temperature in summer, maximum tempera-

ture in summer, and overwinter survival rate, the results were less interpretable. This discrepancy is likely a result of the inconsistent impacts before and after the peak infestation, coupled with the scarcity of data available prior to the peak.

Previous studies in tree biology have shown that jack pine is susceptible and can be poorly defended (Cullingham et al., 2011; Moss, 1949; Rice et al., 2007a, 2007b). Incorporating environmental and ecological covariates, our study further examined and quantified the relative susceptibility of jack pine compared to lodgepole pine. The findings indicated that jack pine has a lower probability of infestation and reduced productivity. Specifically, during an outbreak, lodgepole pine is on average 1.89 times more likely to be infested. Additionally, the reproduction rate of jack pine is only 0.558 times that of lodgepole pine. These rates implied a potentially reduced severity of infestation in jack pine forests compared to lodgepole pine forests, which mitigated concerns about their spread into the boreal forests in Canada, as discussed in Nealis and Peter (2008), Robertson et al. (2009a), Safranyik et al. (2010), and Taylor and Safranyik (2003). In conclusion, our results indicated the possibility of further invasion into the boreal forest, with jack pine serving as a bridge. However, the severity of infestations was expected to be lower compared to those in lodgepole pine.

While our results aligned with our expectations, their accuracy was constrained by the limited reports of jack pine infestations in aerial surveys. The current focus of these surveys is primarily on lodgepole pine forests (Agriculture and Forestry, 2020; Sustainable Resource Development, 2007). Aerial surveys conducted in 2009, 2010, and 2011 indicated a potential threat to jack pine forests in northwest Alberta. However, detailed information about the actual conditions is scant, as aerial surveys did not include the northern areas after 2011. Additionally, ongoing climate change could influence our findings. The anticipated rise in winter temperatures in Alberta (Hayhoe & Stoner, 2019) might decrease beetle mortality during winter (Safranyik & Wilson, 2006), and changes in forest compositions due to climate change could

increase uncertainty in beetle migration patterns (Monserud et al., 2008).

Furthermore, the accuracy of our results was limited by the precision of our covariates. These covariates were detected and estimated at varying levels of precision. For instance, elevation, aspect, and slope maps had a resolution of 100 m, while pine density was estimated by using maps with a resolution of 250 m (Beaudoin et al., 2014; Goodsman et al., 2016), and the pine identity value was estimated from a logistic model presented in a map with a resolution of 1000 m (Burns et al., 2019). Given these limitations in both the aerial surveys and the covariates, the reliability of the absolute risk of infestations was reduced. This suggested a need to shift our focus towards the relative risk, which compares the absolute risk between lodgepole pine and jack pine.

Our results was also constrained by the inherent limitations of the hierarchical model we selected. Our regression model did not account for spatial autocorrelation, which could lead to an underestimation of the significance of the covariates (Legendre, 1993). Consequently, it is essential to collect more comprehensive data on jack pine and acquire covariate maps with finer resolution. Additionally, adopting a more advanced modeling approach will be crucial for enhancing the accuracy of future evaluations concerning the beetle's impact on jack pine forests.

Chapter 4

Impact of climate factors on infestations under climate change

4.1 Introduction

Climate change has been seen as the central factor in many species' invasions (Iverson & Prasad, 1998). It affects temperature and precipitation allowing more locations to become suitable habitat for invasive species. The past three decades (1980-2010) have been recorded as the warmest in history, featuring carbon dioxide (CO2) levels 40% higher than those in 1750 (Hartmann et al., 2013). As reported in Hayhoe and Stoner (2019), Alberta has experienced a rise in summer temperatures ranging from +0.1 to +0.3°C per decade. This change is accompanied by a significant increase in the number of days with temperatures above 25°C and a decrease in the frequency of cold days. The report also highlights a projected change in precipitation patterns, anticipating a 5-10% increase in precipitation from September to April in the future. More specifically, the study of Jiang et al. (2017) shows that the seasonal precipitation could change from -25 to 36%.

The Mountain Pine Beetle (MPB), a native and destructive insect of North America, has a developmental cycle that is highly dependent on temperature (B. J. Bentz et al., 1991). Typically, it completes its life cycle in one year, progressing through four life stages: egg, larva, pupa, and adult (Safranyik & Wilson, 2006). In the summer, MPBs lay eggs that hatch into larvae. These larvae spend winter and spring beneath

the bark, emerging as adults in the following summer. The adults then leave their hosts to find new trees to infest. For a one-year life cycle, beetles typically require a cumulative temperature of at least 833 degrees above 5.6°C (Safranyik & Wilson, 2006). Most beetles emerge when temperatures exceed 20°C (Safranyik & Carroll, 2007), and a range of $19°C$ to 41°C is necessary for successful dispersal in summer (Mccambridge, 1971). Additionally, overwintering success is crucial for maintaining MPB populations (**aukema2008movement**; Safranyik, 1978). The larval stage of the mountain pine beetle, typically occurring in winter, is recognized as the stage with the most cold tolerance (B. Bentz & Mullins, 1999; Safranyik & Wilson, 2006). The commonly accepted lower lethal temperature (LLT) for this stage is around −40°C (B. Bentz & Mullins, 1999; Cooke, 2009). Other life stages have lower cold tolerance. The egg and pupal stages reach 50% mortality at around −20°C (Bleiker & Smith, 2019; Bleiker et al., 2017), and adult beetles rarely survive the winter.

Climate change exerts a multifaceted impact on both beetles and lodgepole pines, affecting various aspects of their biology. Rising temperatures across all seasons lessen the challenge of accumulating the minimum degree-days required for the beetle's life stage development, leading to more beetles in the following summer (Jiang et al., 2017). An increase in the number of days with temperatures above 25°C during summer leads to heightened beetle activity, facilitating greater dispersal (Safranyik & Wilson, 2006). The decrease in the frequency of cold days during winter significantly increases the overwinter survival rate for all beetle life stages, resulting in more beetles emerging in the subsequent summer (Régnière & Bentz, 2007).

Rising temperatures also have the potential to directly increase forest productivity by providing extra heat and lengthening the growing season. However, this benefit could be negated if the increase in warmth is not accompanied by a sufficient rise in precipitation, potentially exacerbating drought conditions and adversely affecting tree growth (Chhin et al., 2008; Monserud et al., 2008; Sauchyn et al., 2003). In Alberta, the dryness index is expected to rise by 20% to 30%, suggesting a more

pronounced drying trend (Barrow & Yu, n.d.). This increased dryness could diminish pine trees' resistance to beetle attacks. Additional research by Monserud et al. (2008) predicts an increase in the site index value, indicating that dominant tree heights in areas with adequate precipitation could grow by up to 5 meters. The same study also underscores the potential decrease in lodgepole pine habitats, forecasting a loss of 59% by the 2080s. Such changes in forest composition could further complicate the prediction of beetle migration patterns. Given the potential changes in temperature and precipitation and the conclusion of Sambaraju et al. (2012), which posits a continuing increase in habitat suitability in Alberta, we anticipated an increase in both the frequency and number of infestations in lodgepole pine forests.

To accurately project beetle dynamics under climate change, we employed the hierarchical Zero-inflated Negative Binomial (ZINB) model discussed in the second and third chapters. Unlike the models in **aukema2008movement**; Kunegel-Lion and Lewis (2020) and Srivastava and Carroll (2023), which estimate only the presence/absence, our regression model assessed both presence/absence and abundance, utilizing comprehensive biological information to describe the habitat and beetle activity. Our model was more sensitive to variations in infestations and offered a more comprehensive understanding of the MPB population dynamics.

By utilizing the number of infestations as our response variable, we incorporated a range of environmental and ecological covariates. To analyze the impact of climate change, we projected two risks: the likelihood and the numbers of infestations. These two risks were estimated using two components within the Zero-inflated Negative Binomial (ZINB) model: the presence model, which is based on a Bernoulli distribution, and the abundance model, which employs a Negative Binomial distribution.

In this chapter, we concentrated on areas with pure lodgepole pine, as most infestations in Alberta have occurred in such forests. Pine identity is an index developed to distinguish between lodgepole and jack pine, with values ranging from 0 to 1 (Burns et al., 2019). Values less than 0.1 typically represent stands containing jack pine,

while those greater than 0.9 indicate lodgepole pine. A value of 1 is indicative of a pure lodgepole pine stand. For our analysis, the selected area is characterized as pure lodgepole pine, denoted by a pine identity value of 1.

Our objective was to create risk maps that visually represent both the likelihood of occurrence and the projected expected number of infestations. To achieve this, we utilized similar data sources as in the previous two chapters. These included aerial survey data from Alberta Environment (Agriculture and Forestry, 2020), pine attribute information from the National Forestry Inventory (NFI) (Beaudoin et al., 2014), and climate data sourced from BioSIM. Our study encompassed a 14-year dataset spanning from 2007 to 2020. Following a similar approach to Kunegel-Lion and Lewis (2020), we have divided the data for the selected area to separate the outbreak infestation period from the subsequent declining stage. As illustrated in Figure 4.2, the outbreak period included the years from 2007 to 2014, while the growth peak declining stage comprised the years from 2015 onwards. The selected area was depicted in Figures 4.1, while the observed dynamics were presented in Figures 4.2.

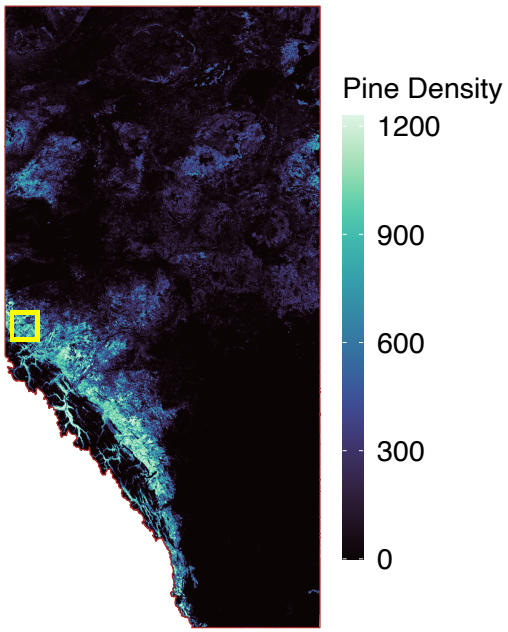


Figure 4.1: The area we select to study impacts of climate change (yellow)

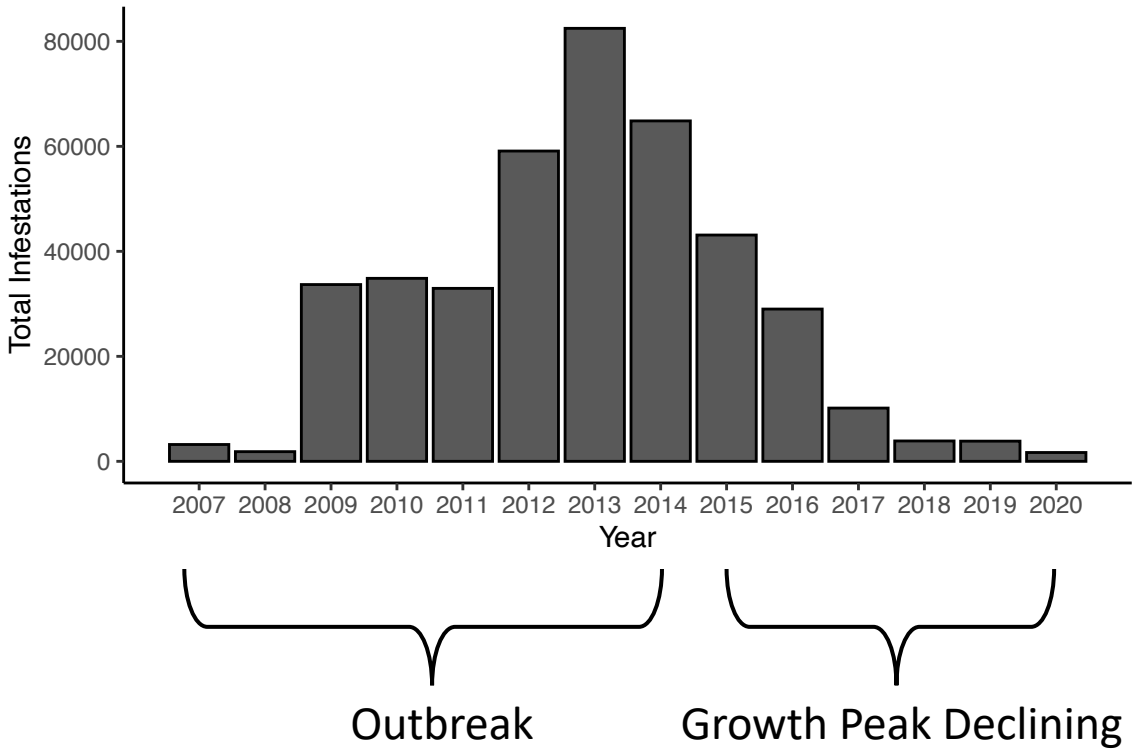


Figure 4.2: Histogram of the observed infestations in the selected area as shown in Figure 4.1.

4.2 Method

4.2.1 Study area and data preparation

As illustrated in Figures 4.1, this chapter focused on the study of lodgepole pine stands located in the western part of Alberta. The selected area, shown in Figure 4.1, corresponded to the area selected in Chapter 2 containing pure lodgepole pines. We have chosen this area due to the clear outbreak observed as shown in Figure 4.2.

In this chapter, we focused on the dynamics of the MPB in relation to climate-related variables such as relative humidity, soil moisture index, degree days, and overwinter survival rate. Our emphasis on broad-scale patterns arising from climate meant that we did not include the more detailed topographic variables listed in Table 3.1. To avoid complexities arising from mixed pine genetics (as discussed in Chapter 3), our analysis was concentrated on regions where the predominant pine species is

consistently lodgepole pine. The maximum temperature in summer was excluded from the analysis due to its strong correlation with degree days, and the minimum temperature in summer was excluded due to interpretability issues.

We have established three climate change scenarios: Representative Concentration Pathway (RCP) 2.6, 4.5, and 8.5. A higher RCP value indicates a scenario of more severe global warming, reflecting greater greenhouse gas emissions. For these scenarios, we substituted the asterisk-marked covariates listed in Table 4.1 with corresponding climate change values. These values were generated using BioSIM, based on projections from the HadGEM2 model with a 10 km resolution (Martin et al., 2006; Ringer et al., 2006). In BioSIM, we produced 10 replications for each covariate under each RCP scenario and then used the mean of these 10 replicates to replace the original covariates. Following the approach in Chapters 2 and 3, we utilized grid cells with a resolution of 500 m for our analysis. The selected area, as depicted in Figure 4.1, comprised 9810 cells.

Variables	Description
Pine density	Pine density (stems/cell; including jack pine, hybrid pine and lodgepole pine) (Beaudoin et al., 2014). The computing method is mentioned in Chapter 2.
Age	Average age of the leading pines in each cell (Beaudoin et al., 2014).
*Relative Humidity	Average relative humidity from March to May in the current year.
Wind Speed	Average wind speed at 2 meters in July and August in the current year.
*Soil Moisture Index	Average soil moisture index from May to July, estimating the water supply for tree growth; current year (Hogg et al., 2013).
*Degree Days	Cumulative degree days above 5.5°C from the September of the previous year to the August in the current year (Safranyik et al., 1975).
*Overwinter survival rate	Probability of survival over the winter; the winter before the current year summer (%) (Régnière & Bentz, 2007).
Last year infestations within cell	Number of pines killed last year within the cell.
Dispersal influence within 4 km	Expected number of beetles flying from all directions to the center cell excluding the counted within cell infestations (Carroll et al., 2017).
Infestation (Response variable)	Number of trees killed this year.

Table 4.1: The description of covariates and response variable considered in this chapter.

4.2.2 Comparative analysis

Since we only selected covariates that have been shown to be significant in the previous two chapters, we fitted the hierarchical ZINB model without penalties. Our aim was to analyze the influence of key climatic covariates, including relative humidity, soil moisture index, degree days, and overwinter survival rate, on infestations under climate change. We defined risks differently from those in Chapter 3. We assessed the risk of infestations and the number of infestations through the presence and abundance models, following Bernoulli and Negative Binomial distributions, respectively. We employed three RCP levels: 2.6, 4.5, and 8.5. Within each RCP scenario, we estimated the probability of infestations from the presence model and the expected number of infestations using the abundance model. We set the value of last year's infestations within a cell and the dispersal influence within 4 km to be the 10% quantile of the non-negative values, which are 3 and 0.05, respectively. The asterisk covariates

in Table 4.1 were replaced by the mean values of 10 replications generated in BioSIM using the inherent HadGEM2 model (Martin et al., 2006; Ringer et al., 2006). We replaced the non-asterisk covariates with the median values for each cell during the outbreak and peak growth decline periods to mitigate the influence of extreme values. Five maps were created under each RCP level for both presence and abundance models. Four maps were generated by altering one asterisk covariate at a time, and in the fifth map, all asterisk covariates were changed. The general procedure was presented in Figure 4.3.

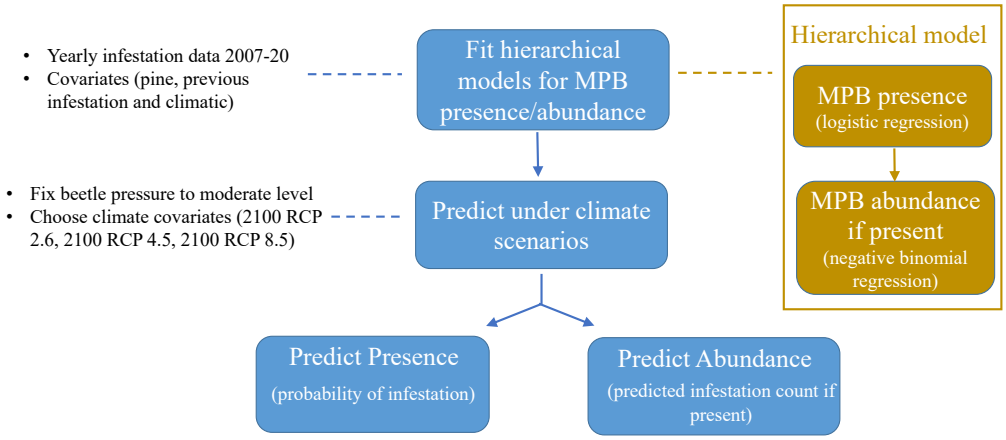


Figure 4.3: The modeling flow used to create risk maps incorporated yearly climatic, and ecological covariates.

4.3 Results

We fitted two hierarchical models for two different stages in the selected area shown in Figure 4.1. We have included the plots from the growth peak declining stage in Appendix C. As shown in Figures 4.4 and 4.5, our covariates indicated that during an outbreak, the soil moisture index was less important in both models and pine density was not significant in the abundance model. In the presence model, overwinter survival rate and degree days positively impacted the probability of occurrence, while a higher relative humidity was associated with a reduced likelihood. The abundance model showed that degree days and overwinter survival rate positively affect beetles,

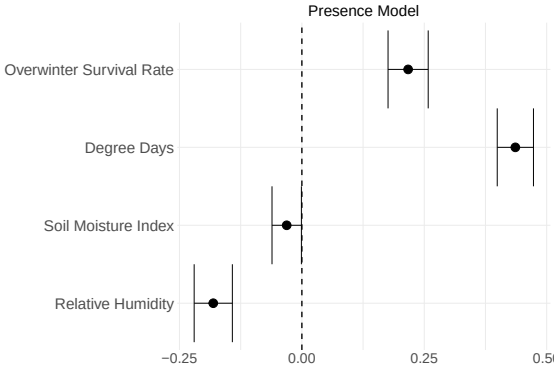
whereas relative humidity has a negative impact.

We also showed the predicted value of four climatic covariates from BioSIM under climate change scenarios. As shown in Figure 4.6, as the temperature goes up, relative humidity rose with the south area showing more significant difference of moisture. Figure 4.7 illustrated that soil moisture decreases as global temperatures rise, potentially resulting in diminished pine resistance to beetle attacks. Figures 4.8 indicated that degree days increase with rising global temperatures and were much higher than the mean of the historical values. Additionally, Figure 4.9 showed that compared to the mean of historical values in the outbreak, the overwinter survival rate increases, while the rate under RCP 2.6 is higher than under RCP 4.5 and the rate under RCP 8.5 is higher than two other RCPs'.

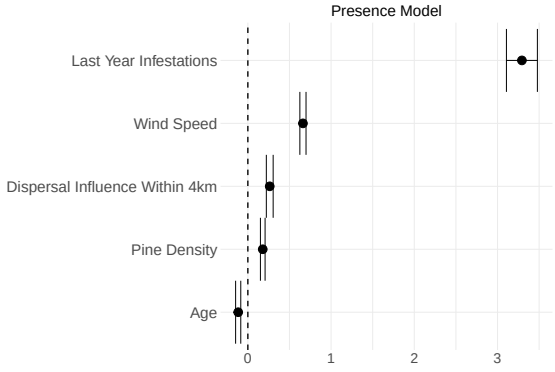
By substituting four climate covariates simultaneously, we observed that the probability of infestation and the number of infestations increase as the global mean temperature rises. The probability of occurrence and the log scale of the expected infestations were estimated using the median value of each cell over years for all covariates, except the last year infestations and dispersal influence within 4 km, which are fixed at 3 and 0.05.

Our results showed that changes in relative humidity lead to a decrease in both the probability of occurrence and the number of infestations, as illustrated in Figures 4.10 and 4.11. The influence of soil moisture index under climate change on beetles was minimal. Figures 4.12 and 4.13 did not present significantly different results under various climate scenarios. Figures 4.14 and 4.15 demonstrated that an increase in degree days significantly raises both the likelihood of beetle presence and the number of infestations. Conversely, the impact of the overwinter survival rate under climate change was relatively minor. Figures 4.16 (b, c) and 4.17 (b, c) showed a slight decrease in both likelihood and number of infestations as the overwinter survival rate decreases when transitioning from RCP 2.6 to RCP 4.5. However, moving from RCP 4.5 to 8.5, we saw an increase in both values, due to the increased cold tolerance, as

shown in Figures 4.16 (c, d) and 4.17 (c, d). As depicted in Figures 4.18 and 4.19, both the probability of occurrence and the number of infestations were higher than those estimated using the median of historical values during outbreaks, suggesting an overall positive impact of climate change on the MPB.

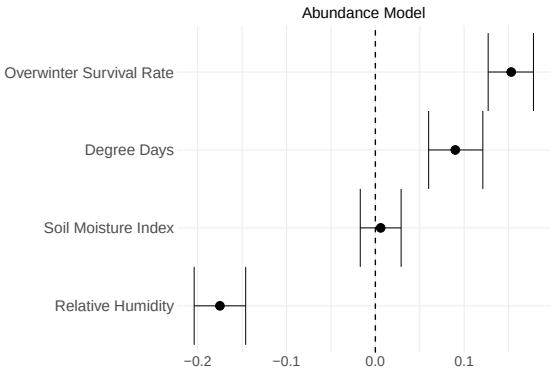


(a) The coefficients of the four standardized climate covariates in the presence model in the outbreak stage.

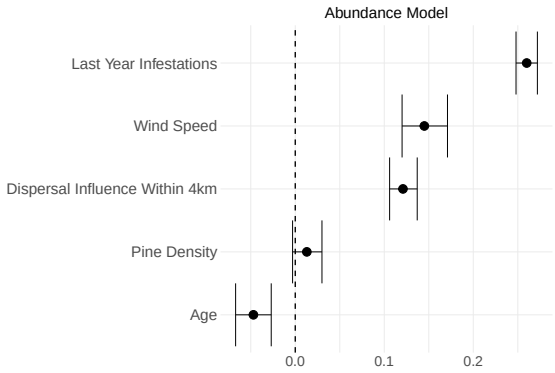


(b) The coefficients of the five standardized non-climate covariates in the presence model in the outbreak stage.

Figure 4.4: Presence model coefficients for the selected area (Figure 4.1) in the outbreak stage.

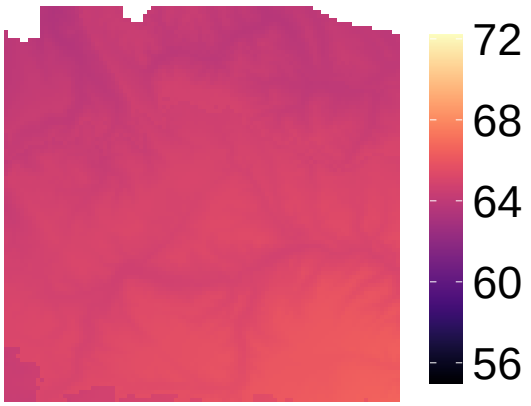


(a) The coefficients of the four standardized climate covariates in the abundance model in the outbreak stage.

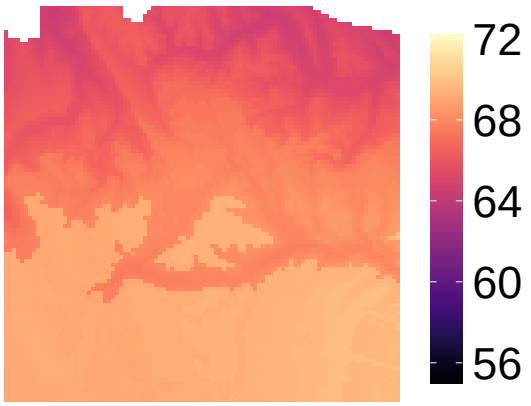


(b) The coefficients of the five standardized non-climate covariates in the abundance model in the outbreak stage.

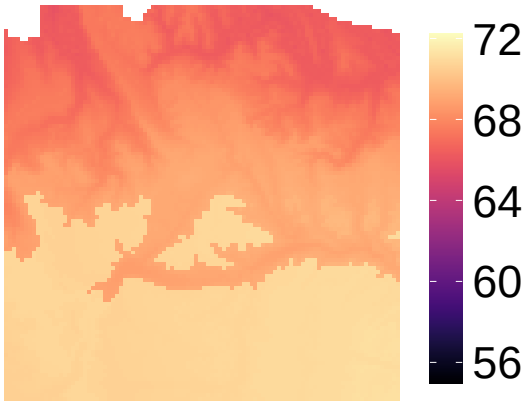
Figure 4.5: Abundance model coefficients for the selected area (shown in Figure 4.1) in the outbreak stage.



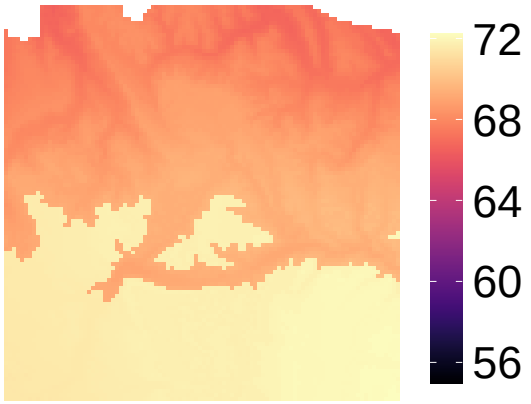
(a) Mean of the relative humidity from 2007 to 2014 from BioSIM in the selected area (shown in Figure 4.1).



(b) Projected relative humidity from BioSIM in the selected area (shown in Figure 4.1) under RCP 2.6 in 2100.

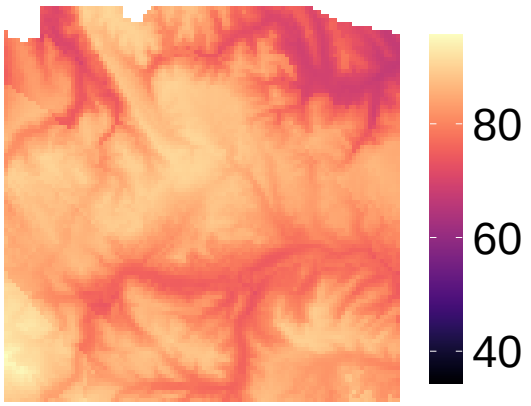


(c) Projected relative humidity from BioSIM in the selected area (shown in Figure 4.1) under RCP 4.5 in 2100.

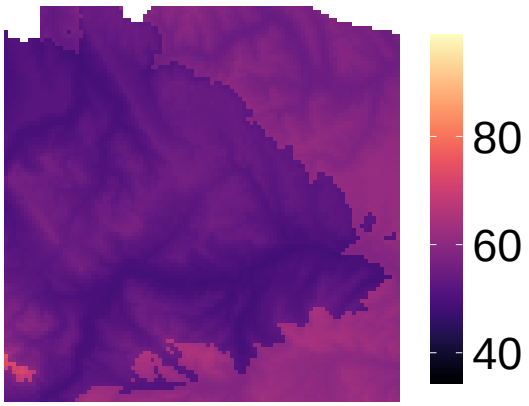


(d) Projected relative humidity from BioSIM in the selected area (shown in Figure 4.1) under RCP 8.5 in 2100.

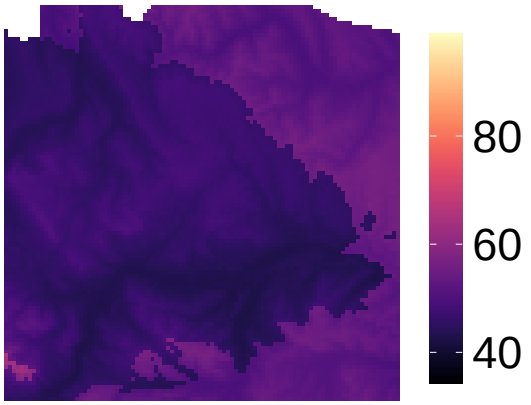
Figure 4.6: Relative humidity in the selected area (shown in Figure 4.1): Current and projected values under different RCP scenarios.



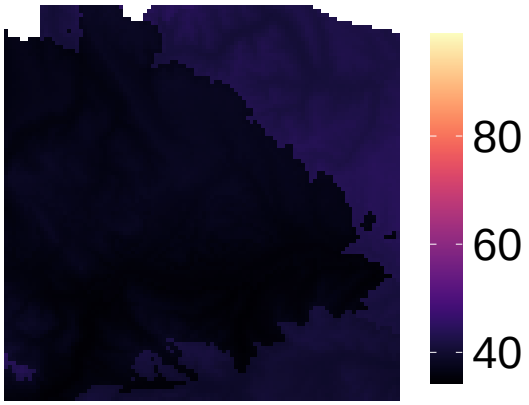
(a) Mean of the soil moisture index from 2007 to 2014 from BioSIM in the selected area (shown in Figure 4.1).



(b) Projected soil moisture index from BioSIM in the selected area (shown in Figure 4.1) under RCP 2.6 in 2100.

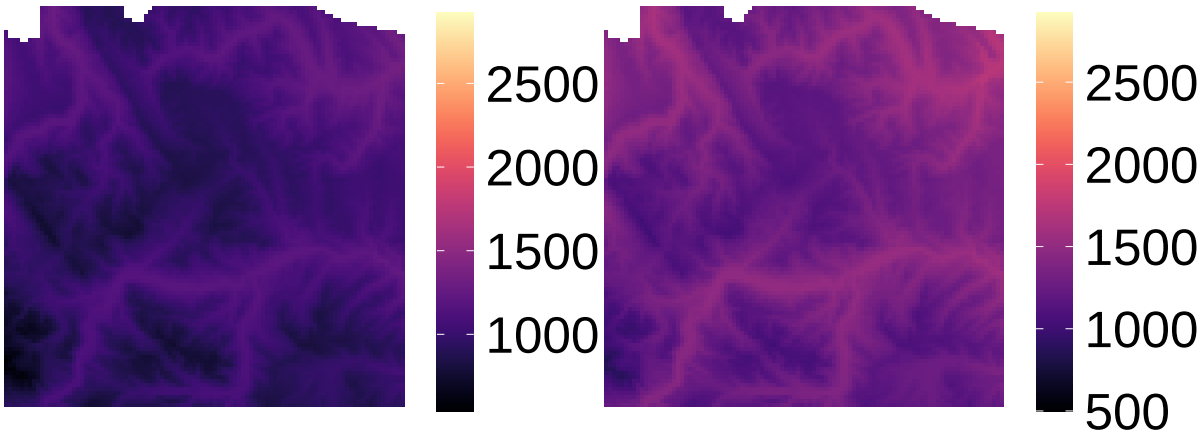


(c) Projected soil moisture index from BioSIM in the selected area (shown in Figure 4.1) under RCP 4.5 in 2100.



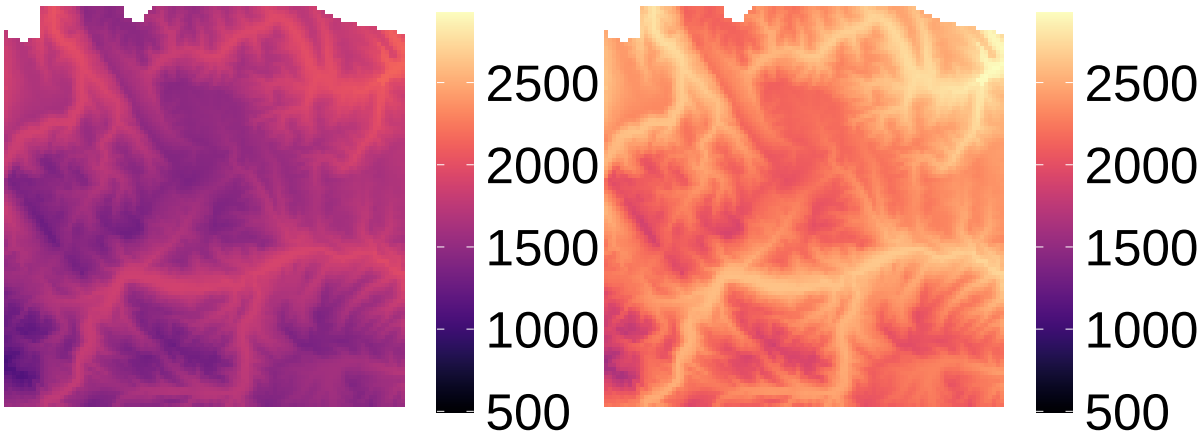
(d) Projected soil moisture index from BioSIM in the selected area (shown in Figure 4.1) under RCP 8.5 in 2100.

Figure 4.7: Soil moisture index in the selected area (shown in Figure 4.1): Current mean and projected values under different RCP scenarios.



(a) Mean of the degree days from 2007 to 2014 in the selected area (shown in Figure 4.1).

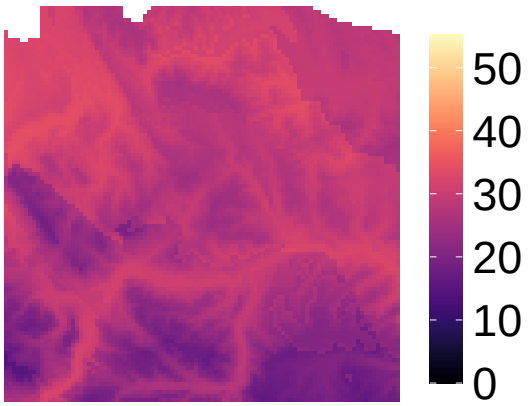
(b) Projected degree days from BioSIM in the selected area (shown in Figure 4.1) under RCP 2.6 in 2100.



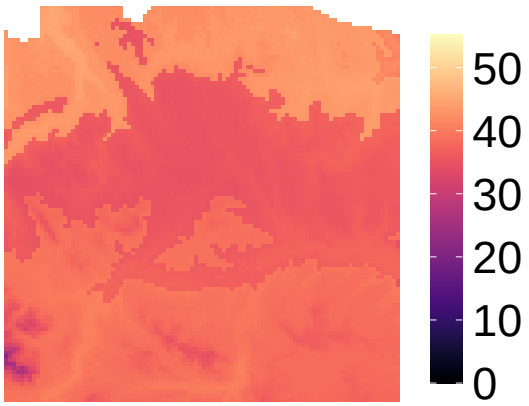
(c) Projected degree days from BioSIM in the selected area (shown in Figure 4.1) under RCP 4.5 in 2100.

(d) Projected degree days from BioSIM in the selected area (shown in Figure 4.1) under RCP 8.5 in 2100.

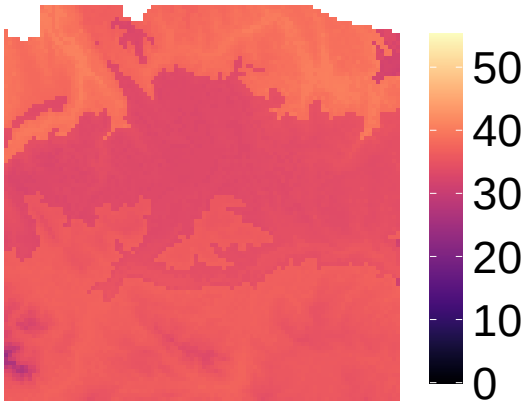
Figure 4.8: Degree days in the selected area (shown in Figure 4.1): Current mean and projected values under different RCP scenarios.



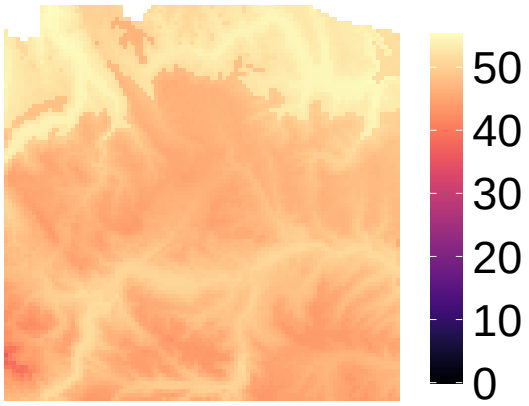
(a) Mean of the overwinter survival rate from 2007 to 2014 in the selected area (shown in Figure 4.1).



(b) Projected overwinter survival rate from BioSIM in the selected area (shown in Figure 4.1) under RCP 2.6 in 2100.

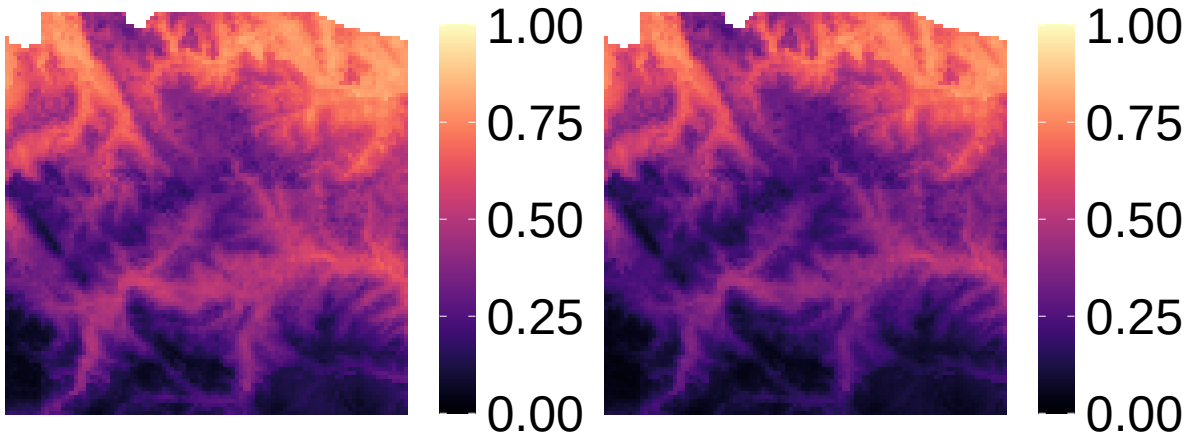


(c) Projected overwinter survival rate from BioSIM in the selected area (shown in Figure 4.1) under RCP 4.5 in 2100.



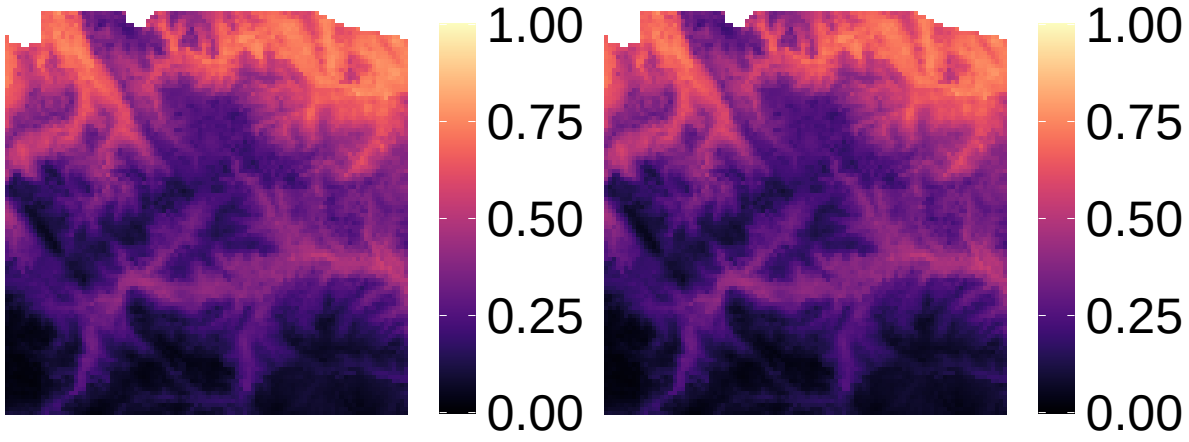
(d) Projected overwinter survival rate from BioSIM in the selected area (shown in Figure 4.1) under RCP 8.5 in 2100.

Figure 4.9: Overwinter survival rate in the selected area (shown in Figure 4.1): Current and projected values under different RCP scenarios.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.

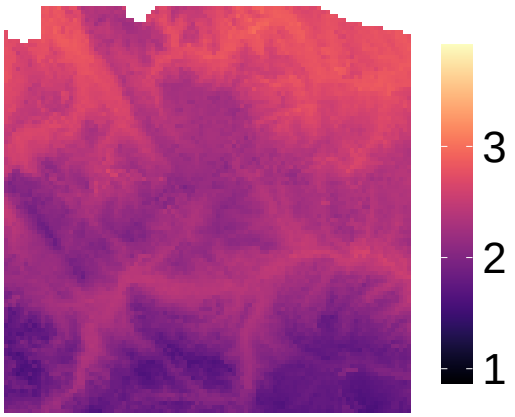
(b) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



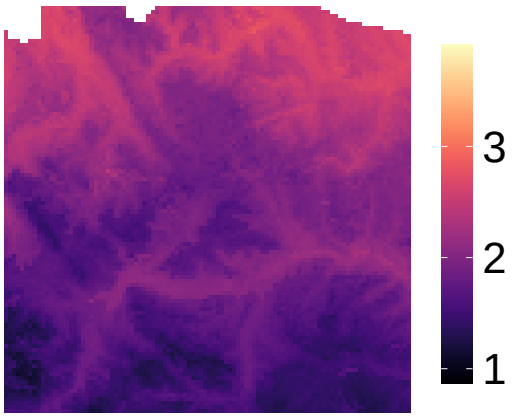
(c) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

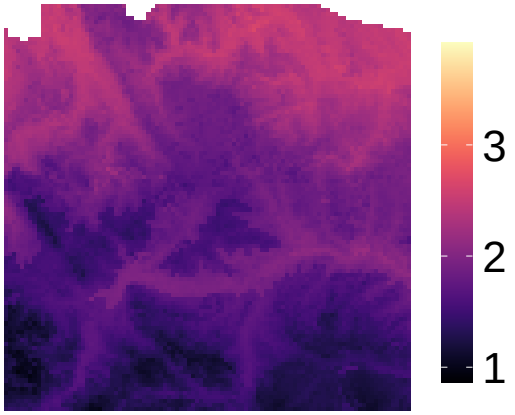
Figure 4.10: Comparative analysis of current and future presence under three RCP scenarios, illustrating a decreasing of probability of occurrence as temperatures increase when varying relative humidity. Only relative humidity is varied under four scenarios.



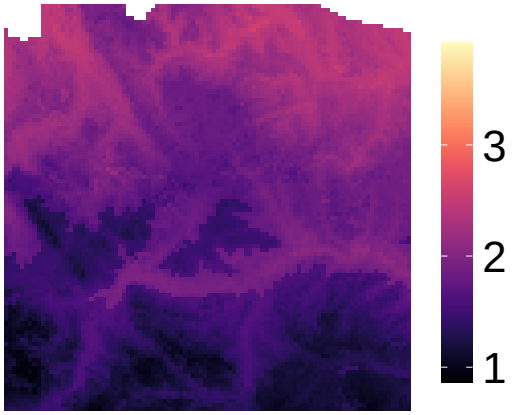
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.

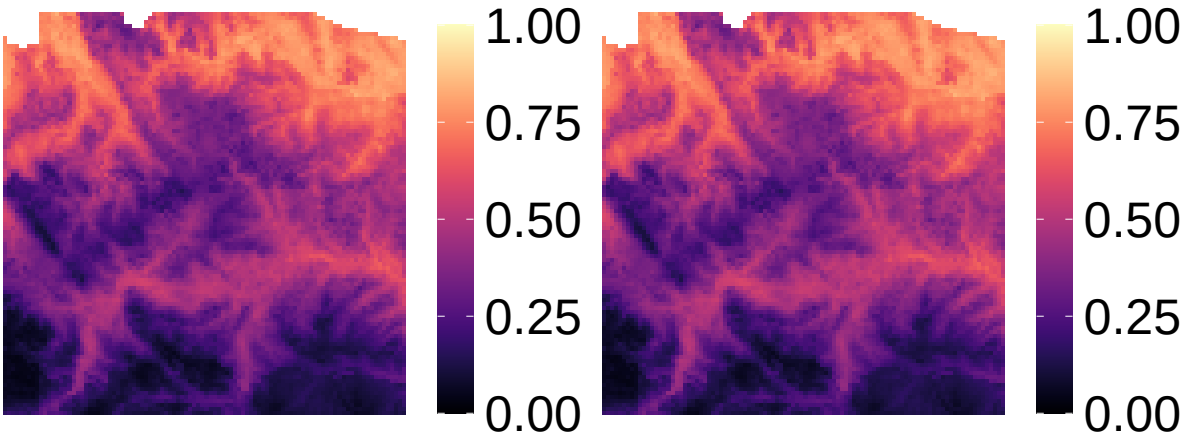


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.



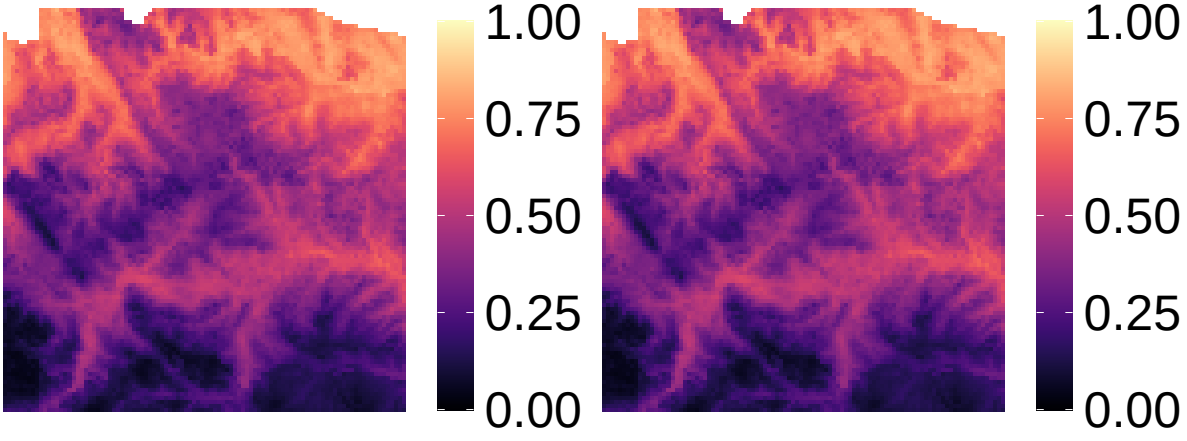
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

Figure 4.11: Comparative analysis of current and future abundance under RCP scenarios, illustrating a decreasing of number of infestations from current to RCP 2.6, an increasing from RCP 2.6 to 4.5 and a decreasing from RCP 4.5 to 8.5. Only relative humidity is varied under four scenarios.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.

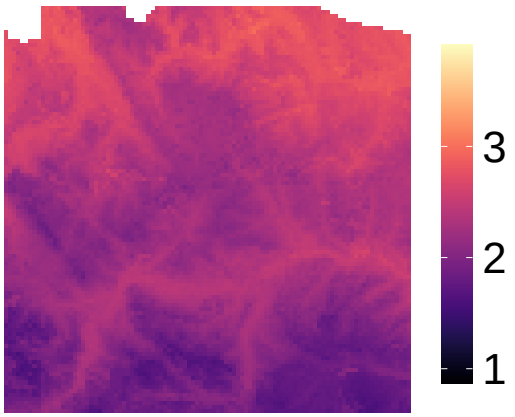
(b) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



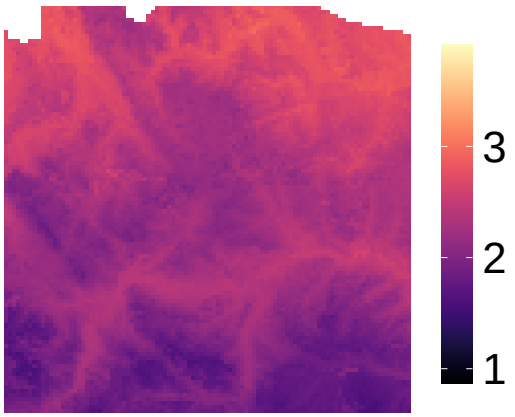
(c) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

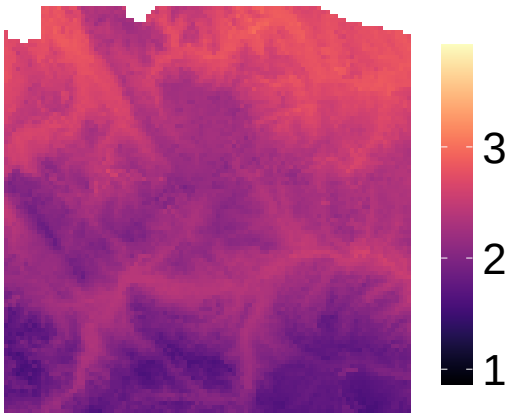
Figure 4.12: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperatures increase when varying soil moisture index. Only soil moisture index is varied under four scenarios.



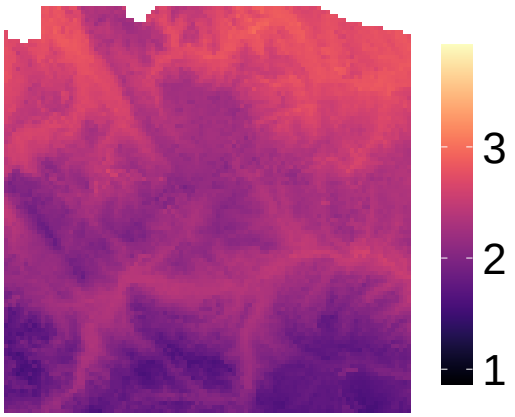
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.

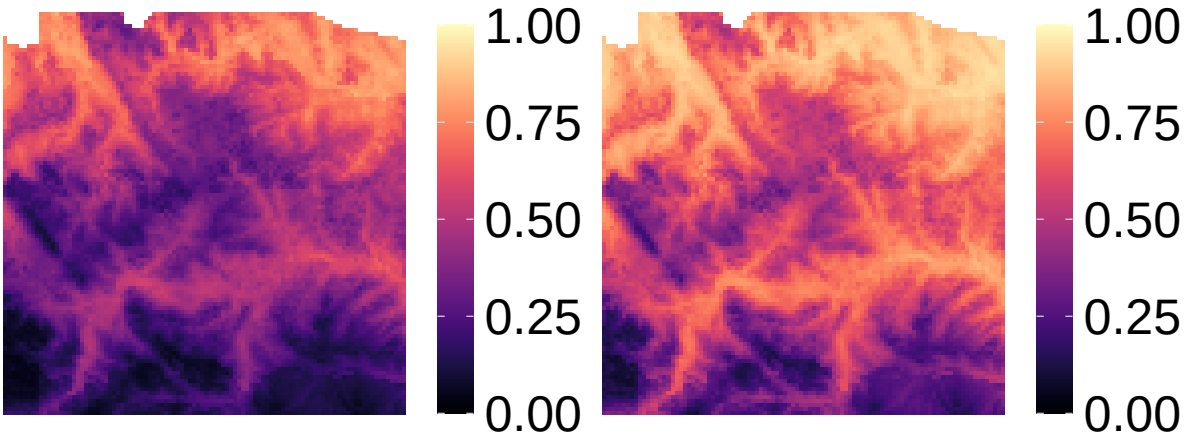


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.



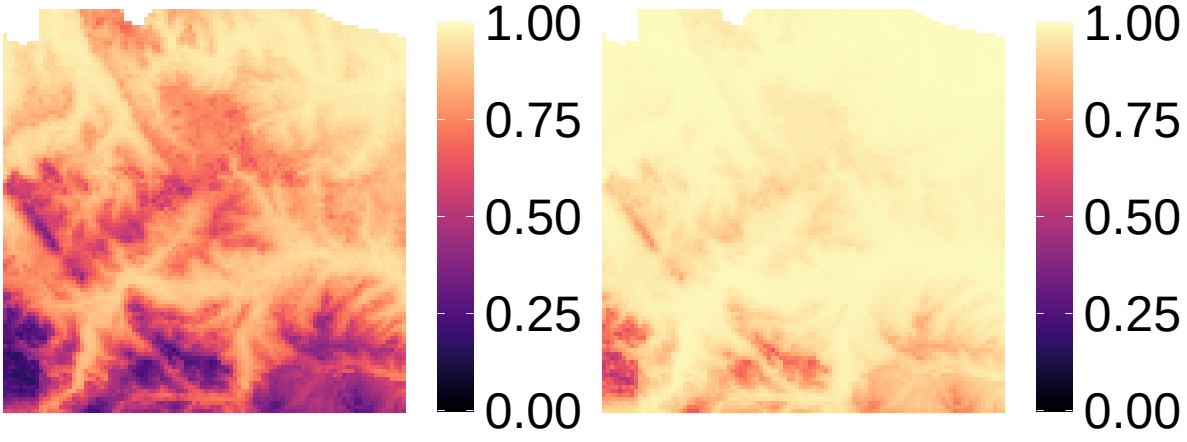
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

Figure 4.13: Comparative analysis of current and future abundance under three RCP scenarios, illustrating no significant difference as temperatures increase when varying soil moisture index. Only soil moisture index is varied under four scenarios.



(a) Probability of occurrence estimated in the presence model in selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.

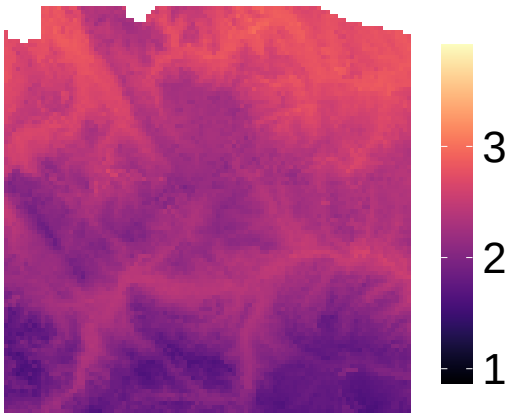
(b) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



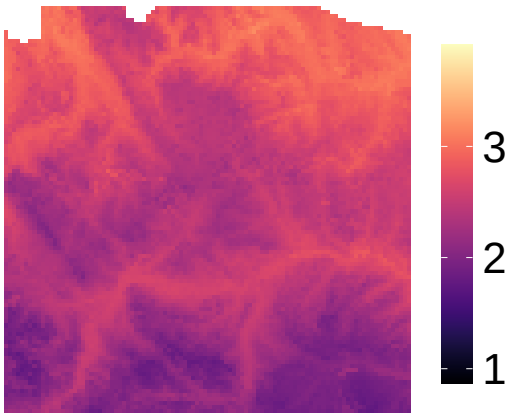
(c) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

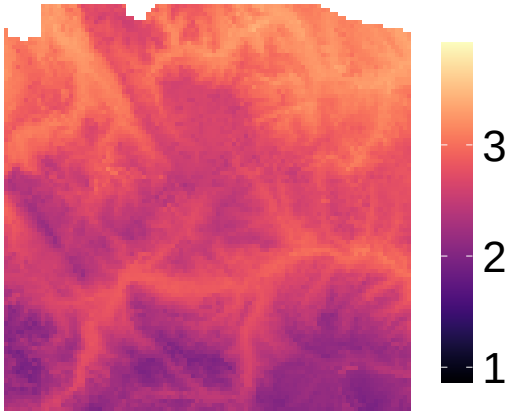
Figure 4.14: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperature increase when varying degree days. Only degree days is varied under four scenarios.



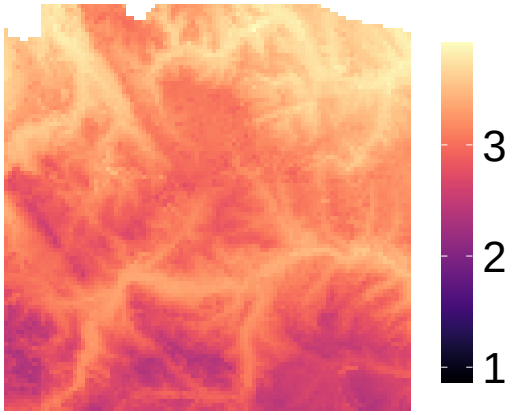
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.

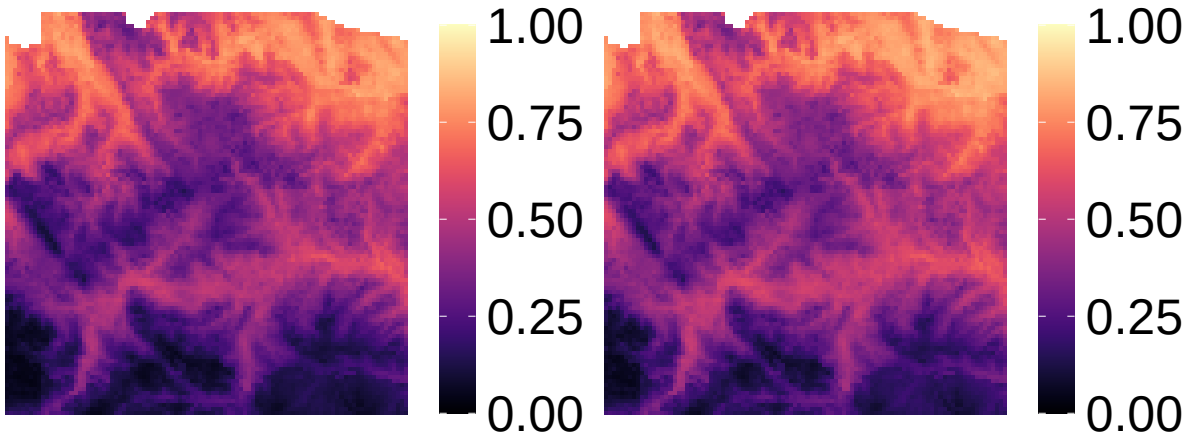


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.



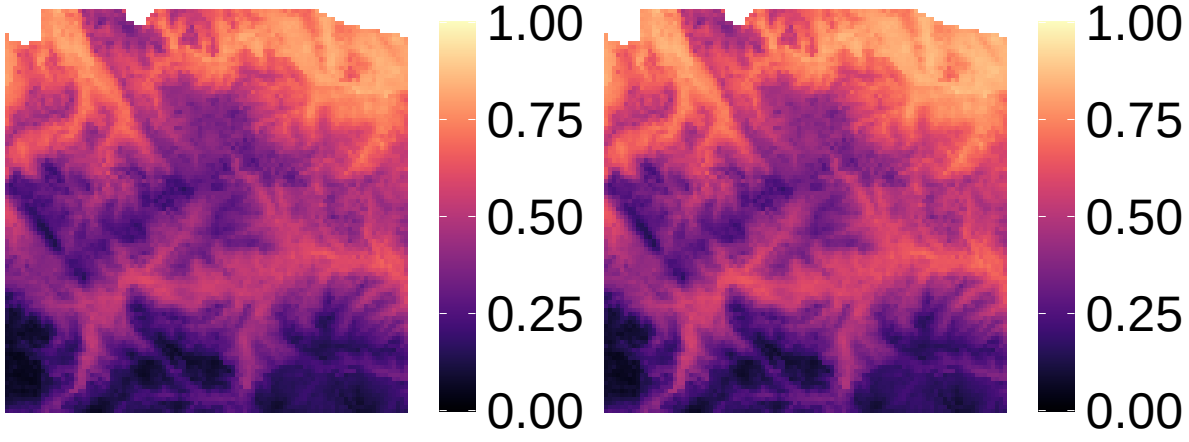
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

Figure 4.15: Comparative analysis of current and future abundance under three RCP scenarios, illustrating an increasing of numbers of infestations as temperatures increase when varying degree days. Only degree days is varied under four scenarios.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.

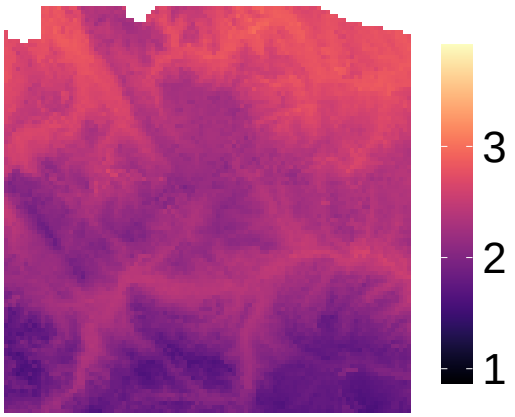
(b) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



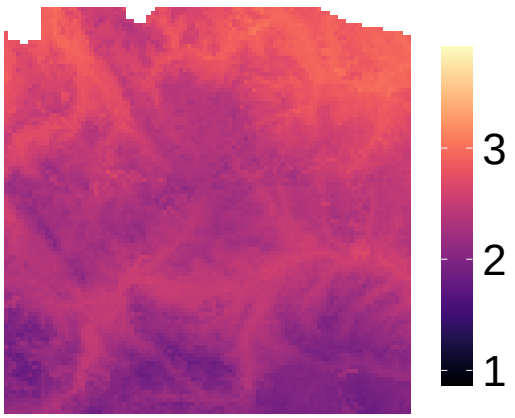
(c) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

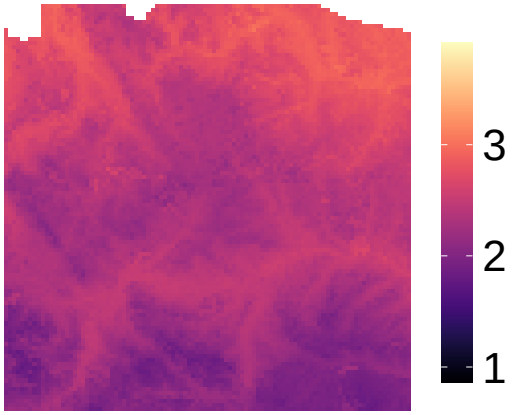
Figure 4.16: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperature increase when varying overwinter survival rate. Only overwinter survival rate is varied under four scenarios.



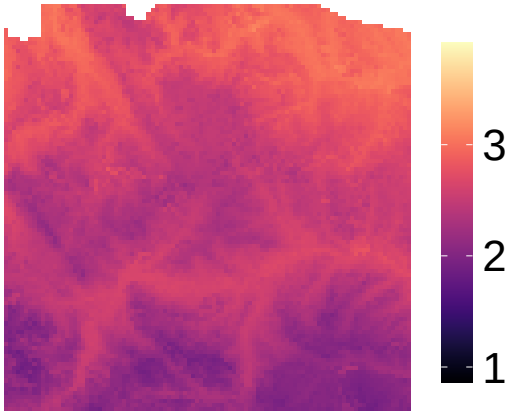
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate value in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.

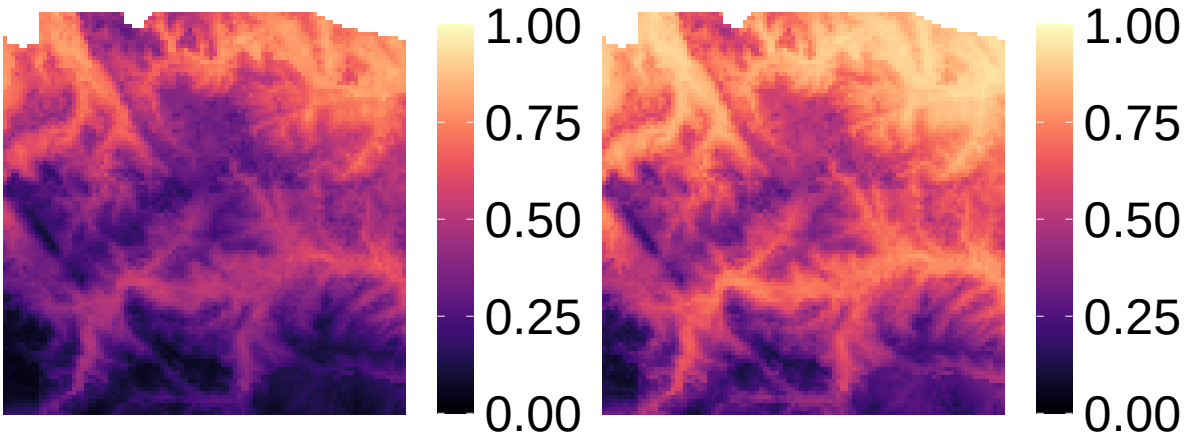


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.



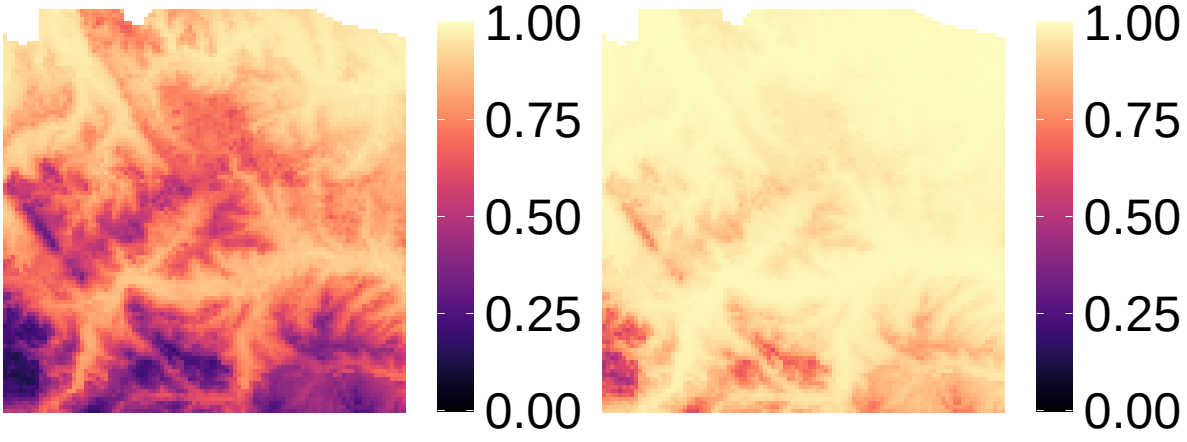
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

Figure 4.17: Comparative analysis of current and future abundance under three RCP scenarios, illustrating an increasing of the number of infestations as temperatures increase when varying overwinter survival rate. Only overwinter survival rate is varied under four scenarios.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.

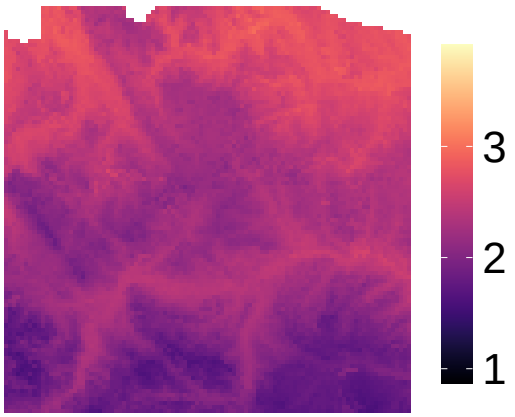
(b) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



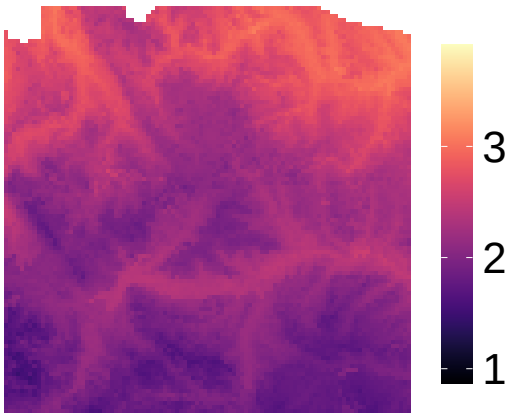
(c) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5 in 2100.

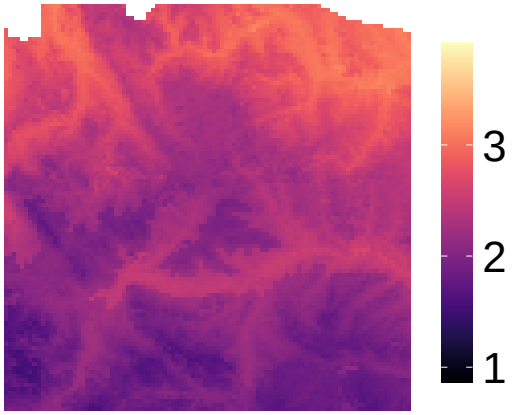
Figure 4.18: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperature increase when varying four covariates. Only four climate covariates are varied under four scenarios.



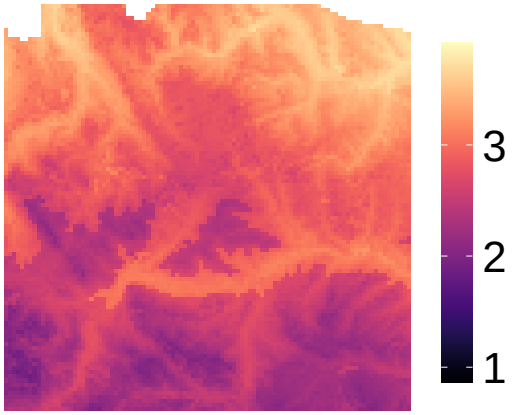
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using outbreak model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 2.6 in 2100.



(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 4.5 in 2100.



(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using outbreak model under RCP 8.5.

Figure 4.19: Comparative analysis of current and future abundance under three RCP scenarios, illustrating an increasing of the number of infestations as temperatures increase when varying four covariates simultaneously. Only four climate covariates are varied under four scenarios.

4.4 Discussion

Our results suggested that climate change could benefit beetles, leading to a higher probability and number of infestations in lodgepole pine forests. In our model, degree days were a crucial factor, guiding our projections of infestation probability and numbers under climate change scenarios. With rising temperatures, degree days could increase the egg hatch successful rate, encourage more dispersal and increase adult emergence (Bleiker & Van Hezewijk, 2016; Safranyik & Wilson, 2006; Sambaraju et al., 2012). Winter temperature was also a crucial factor showing less impacts on infestations compared to degree days. As global mean temperatures rise, the resulting milder winters lead to increased beetle survival rates, subsequently resulting in larger beetle populations in the following summer (Régnière & Bentz, 2007; Taylor & Safranyik, 2003). According to the projections from BioSIM, relative humidity was expected to show an overall increasing trend, albeit with local variations. This increase in relative humidity, when combined with its negative impact on beetles, led to a general reduction in both the probability and number of infestations. The soil moisture index's impact remained ambiguous in our model, as its confidence intervals in both presence and abundance models included zero, indicating insignificance. This result could be attributed to the minimal variation in soil moisture within the selected study area.

Our findings were consistent with previous studies suggesting an increased likelihood of outbreak under climate change scenarios (Pacific Forestry Centre & Carroll, Allan L, 2006; Sambaraju et al., 2012; Srivastava & Carroll, 2023; Taylor & Safranyik, 2003). These studies primarily focused on how climate change could reshape MPB habitats and extend their range. Our research specifically examined the impact of key climate-related covariates on MPB dynamics in lodgepole pine forests. Our results showed that future infestations are largely driven by degree days and overwinter survival rates. Seasonal temperatures were expected to increase by approximately

0.048°C annually (Jiang et al., 2017), with the annual mean minimum temperature projected to rise from around −5°C to 0°C by 2100 (Eum et al., 2023). These trends suggested that the degree days and overwinter survival rates of the beetle will intensify, leading to more suitable habitats at higher latitudes. While Chapter 3 indicated that jack pine is less susceptible to infestation and supports fewer reproductions, the evolving climate - characterized by changing temperatures and precipitation patterns – still poses a risk of future infestations.

Additionally, our results should be primarily interpreted as indicative of general trends under climate change, owing to inherent uncertainties in our covariates and model. The reliability of estimates from BioSIM heavily depends on the proximity to the nearest weather stations. The absence of weather stations north of the study area introduced significant uncertainties, particularly affecting the patterns of relative humidity, soil moisture index, and overwinter survival rate. Moreover, the applicability of our regression model was limited to the scope of the available data and as demonstrated in Chapter 2, prediction accuracy was sacrificed for greater interpretability of MPB population dynamics. Therefore, our findings lacked precision and should be regarded as indicative rather than definitive. Additionally, the HadGEM2 model, known for its conservative projections, suggested that the actual future impact of climate change could be more severe than our projections indicated (Collins et al., 2011).

Chapter 5

Conclusion

The potential range expansion of the mountain pine beetle has raised concerns among people, particularly following the outbreak in the late 1990s and 2000s. This outbreak enabled the beetles to breach the barrier of the Rocky Mountains, reaching Northwest Alberta (Nealis & Peter, 2008; Robertson et al., 2009a; Safranyik et al., 2010; Sambaraju et al., 2012; Taylor & Safranyik, 2003). As a consequence, this expansion directly exposed jack pine, a predominant species in Canada's boreal forests extending from Northern Alberta to Nova Scotia, to the risk of infestation (Rudolph & Laidly, 1990). Furthermore, the study by Cullingham et al. (2011) demonstrated the successful attack of beetles on jack pine, heightening concerns about potential outbreaks in jack pine forests.

Additionally, climate change has been seen as a crucial factor in past expansions (Taylor & Safranyik, 2003). Warmer summer temperatures could enhance the beetles' dispersal (Safranyik & Wilson, 2006), while milder winters could significantly increase their survival rates (Régnière & Bentz, 2007; Taylor & Safranyik, 2003). Although beetle activity will be suppressed if the temperature is excessively high, particularly above 41°C, the colder climate of Alberta provides sufficient opportunity for beetles to reproduce before reaching the temperature threshold (Safranyik & Carroll, 2007; Safranyik & Wilson, 2006).

In this thesis, we discussed the infestation in jack pine and the overall trend of in-

festations in lodgepole pine forests under climate change, utilizing data from Alberta. We will briefly summarize the related results and initiate a discussion for future work in the following paragraphs.

5.1 Model for mountain pine beetle dynamics

Before examining two biological questions, our initial step involved variable and model selection to identify covariates and a model that best match the population dynamics of mountain pine beetle in Alberta. Through statistical hypothesis testing, we concluded that our study should include geographical covariates (such as northness, eastness, and slope), climate-related covariates (maximum and minimum temperatures in summer, overwinter survival rate, degree days, relative humidity, soil moisture index, wind speed), and ecological covariates (pine age, pine density, last year infestations, and dispersal influence within 4 km).

Employing this revised set of covariates, we discovered that the Zero-Inflated Negative Binomial (ZINB) model offered a better fit for the dynamics in Alberta and demonstrated higher accuracy in predicting infestations during 2012 and 2013. Although the Zero-Inflated Negative Binomial (ZINB) model outperformed the Poisson and Zero-inflated Poisson models, its limitations were evident in the residual plots. A perfectly fitting model would produce a residual plot without any dense patterns or clear increasing or decreasing trends in the residuals, as noted by Feng et al. (2020). Such a plot would also exhibit homogeneity. Unfortunately, our data displayed patterns of residuals grouping together and showed heterogeneity, which was likely due to the model's failure to account for spatial autocorrelations.

The standard regression model always has lower accuracy when predicting the future compared with machine learning models. Our hierarchical model exhibited low accuracy, despite predicting only four severity classes rather than the exact number of infestations in the near future. However, our hierarchical model offered stronger interpretability, aiding in a better understanding of the dynamics of the mountain

pine beetle. It allowed us to analyze separately the impacts of covariates on the presence and abundance of infestations.

The application of penalties in our model also merited mention. They allowed for the inclusion of comprehensive information when mitigating concerns about multicollinearity (Hoerl & Kennard, 1970). The integration of variable selection within the model fitting process offered a time-efficient approach, particularly beneficial considering 18 covariates, which would otherwise result in 2^{18} subsets using the best subset selection method (Fan & Li, 2001; Hocking & Leslie, 1967). A potential solution to the current model's drawbacks could be the utilization of a spatial hierarchical model with penalties (Chi & Zhu, 2008).

5.2 Mountain pine beetle in jack pine forest

A key conclusion of this thesis, which supported the findings of Bleiker et al. (2023), was the lower likelihood of an outbreak in jack pine compared with lodgepole pine. Utilizing the best model selected in Chapter 2, our analysis revealed that lodgepole pine is, on average, 1.89 times more likely to be infested compared to jack pine. Jack pine exhibited a lower reproduction rate, with a value only 0.558 times that of lodgepole pine. By considering the complex physical environment, our outcome contrasted with previous suggestion based on the examination of tree biology, which showed that jack pine can be poorly defended (Clark et al., 2014). Our results, from the perspective of spatial ecology, provide an overview of the beetle's behavior, indicating that, given past climate and precipitation patterns, it might be challenging for beetles to establish themselves in jack pine forests.

However, these dynamics could shift under the influence of climate change. The range of the mountain pine beetle is largely determined by weather patterns and seasonal suitability (Jenkins et al., 2001; Logan & Bentz, 1999; Logan & Powell, 2001). Among the climate covariates, cold winters play a crucial role in restricting the mountain pine beetle population (**aukema2008movement**; Taylor & Safranyik,

2003). As illustrated in Chapter 4, climate change could lead to a significant increase in degree days and overwinter survival rates, potentially expanding suitable habitats for the mountain pine beetle. Tree availability is also a key factor influencing beetle behavior (Safranyik et al., 2010; Safranyik & Wilson, 2006). Projections suggest that by the 2080s, the lodgepole pine forests in Alberta could face a reduction of 59% (Monserud et al., 2008). This decrease in lodgepole pine forests may reduce food sources for the mountain pine beetle, potentially driving them to infest jack pine forests. Although our findings indicated a low likelihood of infestations in jack pine forests, ongoing climate change still poses a risk of future infestations.

5.3 Impacts of climate factors under climate change

In Chapter 4, we analyzed the impact of four climate-related covariates - degree days, overwinter survival rate, soil moisture index, and relative humidity - on mountain pine beetle population dynamics in lodgepole pine forests under the influence of climate change. Our findings indicated that degree days plays the most significant role, exhibiting a strong positive effect on the mountain pine beetle population. The overall trend suggested an increase in both the probability and number of infestations, pointing to more frequent outbreaks approaching 2100.

It is important to note that our results should be considered indicative of potential future scenarios rather than definitive predictions. As outlined in Chapter 2, the accuracy of our hierarchical model was limited to 73% when predicting peak years infestations in the near future. Moreover, the covariates used in our analysis contained inherent uncertainties. The four climate-related covariates, generated using BioSIM, were based on simulations from the four nearest weather stations, and their precision was closely linked to the distance to these stations. This was evident in Figures 4.6, 4.7, 4.8, and 4.9, where edges can be observed due to the uneven distances to the weather stations. Additionally, the relatively small size of the selected study area may obscure certain insights related to climate change. The area's limited extent

resulted in more uniformity in temperature, soil moisture, and relative humidity. This uniformity could potentially explain the ambiguous role of the soil moisture index in our model and the less pronounced impacts of the overwinter survival rate on infestations. Furthermore, the HadGEM2 model, known for its conservative climate change projections (Collins et al., 2011), suggested that future trends could be more severe than our current projections. Future studies with more representative area could provide more insights on climate change.

5.4 Conclusion

This thesis examined two key biological questions: the dynamics of the mountain pine beetle population in jack pine forests and the impact of climate change on lodgepole pine forests. By incorporating environmental and ecological covariates, our study shed light on the interactions among the physical environment, forests, and beetles. These covariates enhanced the realism of our estimates regarding the relative risk of infestations in jack pine compared to lodgepole pine, as well as the reproduction rate. The indicative trends in lodgepole pine forests under climate change suggested that temperature is a primary driver of beetle dynamics. Furthermore, the local variations projected for relative humidity indicated an increase in the spatial variability of future infestations. Although our results based on past data showed a low probability of infestations in jack pine forests, the potential consequences of climate change implied that the risk of infestations in these forests should remain a concern.

Bibliography

Agriculture and Forestry. (2020). Alberta agriculture and forestry forest health aerial survey guide 2020 [Updated: May 1, 2020]. https://open.alberta.ca/dataset/dfa5303c-7101-45bc-a1a8-bb24780159f9/resource/fd67b33e-4db2-4e06-90b5-a933d6094ecf/download/af-forest-health-aerial-survey-guide-2020.pdf

Agriculture, Forestry and Rural Economic Development. (2022). Alberta vegetation inventory standards. version 2.1.5 [Issued: 2022-01-21]. *Agriculture, ForestryRural Economic Development*. https://open.alberta.ca/dataset/a35dcb86-eaa2-492f-9541-93d35174ee06/resource/5124f3e7-05d7-4d88-b489-252d749e81a8/download/afred-alberta-vegetation-inventory-standards-version-2-1-5.pdf

Amman, G. D. (1976). *Optimum egg gallery densities for the mountain pine beetle in relation to lodgepole pine phloem thickness* (Vol. 209). Intermountain Forest & Range Experiment Station.

Amman, G. D. (1978). Biology, ecology, and causes of outbreaks of the mountain pine beetle in lodgepole pine forests. *Theory and practice of mountain pine beetle management in lodgepole pine forests: proceedings of the symposium*, 25–27.

Aukema, B. H., Carroll, A. L., Zhu, J., Raffa, K. F., Sickley, T. A., & Taylor, S. W. (2006). Landscape level analysis of mountain pine beetle in british columbia, canada: Spatiotemporal development and spatial synchrony within the present outbreak. *Ecography*, *29*(3), 427–441.

Barrow, E., & Yu, G. (n.d.). Climate scenarios for alberta, prairie adaptation.

Beaudoin, A., Bernier, P., Guindon, L., Villemaire, P., Guo, X., Stinson, G., Bergeron, T., Magnussen, S., & Hall, R. (2014). Mapping attributes of canada's forests at moderate resolution through k nn and modis imagery. *Canadian Journal of Forest Research*, *44*(5), 521–532.

Bentz, B. J., Logan, J. A., & Amman, G. D. (1991). Temperature-dependent development of the mountain pine beetle (coleoptera: Scolytidae) and simulation of its phenology. *The Canadian Entomologist*, *123*(5), 1083–1094.

Bentz, B., & Mullins, D. (1999). Ecology of mountain pine beetle (coleoptera: Scolytidae) cold hardening in the intermountain west. *Environmental Entomology*, *28*(4), 577–587.

Björklund, N., & Lindgren, B. S. (2009). Diameter of lodgepole pine and mortality caused by the mountain pine beetle: Factors that influence their relationship and applicability for susceptibility rating. *Canadian Journal of Forest Research*, *39*(5), 908–916. https://login.ezproxy.library.ualberta.ca/login?

url=https://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=43017368&site=eds-live&scope=site

Bleiker, K., Ethier, C. A., & Van Hezewijk, B. H. (2023). Suitability of a historical, novel, and occasional host for mountain pine beetle (coleoptera: Curculionidae). *Forests*, *14*(5), 989.

Bleiker, K., O'Brien, M. R., Smith, G. D., & Carroll, A. L. (2014). Characterisation of attacks made by the mountain pine beetle (coleoptera: Curculionidae) during its endemic population phase. *The Canadian Entomologist*, *146*(3), 271–284.

Bleiker, K., & Smith, G. (2019). Cold tolerance of mountain pine beetle (coleoptera: Curculionidae) pupae. *Environmental entomology*, *48*(6), 1412–1417.

Bleiker, K., Smith, G., & Humble, L. (2017). Cold tolerance of mountain pine beetle (coleoptera: Curculionidae) eggs from the historic and expanded ranges. *Environmental Entomology*, *46*(5), 1165–1170.

Bleiker, K., & Van Hezewijk, B. (2016). Flight period of mountain pine beetle (coleoptera: Curculionidae) in its recently expanded range. *Environmental Entomology*, *45*(6), 1561–1567.

Burns, I., James, P. M., Coltman, D. W., & Cullingham, C. I. (2019). Spatial and genetic structure of the lodgepole× jack pine hybrid zone. *Canadian Journal of Forest Research*, *49*(7), 844–853.

Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of geographical information systems*. Oxford University Press, USA.

Carroll, A. L., Aukema, B., Raffa, K., Linton, D., Smith, G. D., & Lindgren, B. (2006). Mountain pine beetle outbreak development: The endemic—incipient epidemic transition. *Canadian Forest Service, Mountain Pine Beetle Initiative Project*, *1*, 22.

Carroll, A. L., Seely, B., Welham, C., & Nelson, H. (2017). Assessing the effectiveness of alberta's forest management program against the mountain pine beetle. *Final report for fRI Project*, *246*.

Chang, Y.-P. (1954). *Bark structure of north american conifers*. US Department of Agriculture.

Chhin, S., Hogg, E. T., Lieffers, V. J., & Huang, S. (2008). Potential effects of climate change on the growth of lodgepole pine across diameter size classes and ecological regions. *Forest ecology and management*, *256*(10), 1692–1703.

Chi, G., & Zhu, J. (2008). Spatial regression models for demographic analysis. *Population Research and Policy Review*, *27*, 17–42.

Clark, E. L., Pitt, C., Carroll, A. L., Lindgren, B. S., & Huber, D. P. (2014). Comparison of lodgepole and jack pine resin chemistry: Implications for range expansion by the mountain pine beetle, dendroctonus ponderosae (coleoptera: Curculionidae). *PeerJ*, *2*, e240.

Cole, W. E. (1969). *Mountain pine beetle infestations in relation to lodgepole pine diameters* (Vol. 95). US Department of Agriculture, Forest Service, Intermountain Forest & Range ...

Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., et al. (2011). Development

and evaluation of an earth-system model–hadgem2. *Geoscientific Model Development*, *4*(4), 1051–1075.

Cooke, B. (2009). Forecasting mountain pine beetle-overwintering mortality in a variable environment. *Mountain Pine Beetle Working Paper-Pacific Forestry Centre, Canadian Forest Service*, (2009-03).

Critchfield, W. B. (1985). The late quaternary history of lodgepole and jack pines. *Canadian Journal of Forest Research*, *15*(5), 749–772.

Cullingham, C. I., Cooke, J. E., Dang, S., Davis, C. S., Cooke, B. J., & Coltman, D. W. (2011). Mountain pine beetle host-range expansion threatens the boreal forest. *Molecular ecology*, *20*(10), 2157–2171.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and graphical statistics*, *5*(3), 236–244.

Eum, H.-I., Fajard, B., Tang, T., & Gupta, A. (2023). Potential changes in climate indices in alberta under projected global warming of 1.5–5° c. *Journal of Hydrology: Regional Studies*, *47*, 101390.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Feng, C., Li, L., & Sadeghpour, A. (2020). A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology*, *20*(1), 1–21.

Gibson, E. K. (2004). Management guide for mountain pine beetle [Accessed: July 2010]. https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5187520.pdf

Goodsman, D. W., Koch, D., Whitehouse, C., Evenden, M. L., Cooke, B. J., & Lewis, M. A. (2016). Aggregation and a strong a llee effect in a cooperative outbreak insect. *Ecological Applications*, *26*(8), 2623–2636.

Government of Alberta. (2023). Mountain pine beetle in alberta [Accessed: 2023]. https://www.alberta.ca/mountain-pine-beetle-in-alberta#:~:text=Contact-,Current%20status,term%2C%20pine%2Dfocused%20strategy.

Government of Canada. (2021). Mountain pine beetle (factsheet) [Accessed: 2021-02-01]. https://natural-resources.canada.ca/forests/fire-insects-disturbances/top-insects/13397#

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., & Kaplan, A. (2013). Observations: Atmosphere and surface. In *Climate change 2013 the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change* (pp. 159–254). Cambridge University Press.

Hayhoe, K., & Stoner, A. (2019). *Alberta's climate future* (tech. rep.). ATMOS Research & Consulting.

Heavilin, J., & Powell, J. (2008). A novel method of fitting spatio-temporal models to data, with applications to the dynamics of mountain pine beetles. *Natural Resource Modeling*, *21*(4), 489–524.

Hocking, R. R., & Leslie, R. (1967). Selection of the best subset in regression analysis. *Technometrics*, *9*(4), 531–540.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hogg, E., Barr, A., & Black, T. (2013). A simple soil moisture index for representing multi-year drought impacts on aspen productivity in the western canadian interior. *Agricultural and Forest Meteorology*, *178*, 173–182.

Iverson, L. R., & Prasad, A. M. (1998). Predicting abundance of 80 tree species following climate change in the eastern united states. *Ecological Monographs*, *68*(4), 465–485.

Jenkins, J. L., Powell, J. A., Logan, J. A., & Bentz, B. J. (2001). Low seasonal temperatures promote life cycle synchronization. *Bulletin of mathematical biology*, *63*(3), 573–595.

Jiang, R., Gan, T. Y., Xie, J., Wang, N., & Kuo, C.-C. (2017). Historical and potential changes of precipitation and temperature of alberta subjected to climate change impact: 1900–2100. *Theoretical and applied climatology*, *127*, 725–739.

Kaufeld, K. A., Heaton, M. J., & Sain, S. R. (2014). A spatio-temporal model for mountain pine beetle damage. *Journal of agricultural, biological, and environmental statistics*, *19*, 437–450.

Kenkel, N., Hendrie, M., & Bella, I. (1997). A long-term study of pinus banksiana population dynamics. *Journal of Vegetation Science*, *8*(2), 241–254.

Klar, B., & Meintanis, S. G. (2012). Specification tests for the response distribution in generalized linear models. *Computational Statistics*, *27*, 251–267.

Kunegel-Lion, M., & Lewis, M. A. (2020). Factors governing outbreak dynamics in a forest intensively managed for mountain pine beetle. *Scientific Reports*, *10*(1), 7601.

Kurz, W. A., Dymond, C., Stinson, G., Rampley, G., Neilson, E., Carroll, A., Ebata, T., & Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, *452*(7190), 987–990.

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

Logan, J. A., & Bentz, B. J. (1999). Model analysis of mountain pine beetle (coleoptera: Scolytidae) seasonality. *Environmental Entomology*, *28*(6), 924–934.

Logan, J. A., & Powell, J. A. (2001). Ghost forests, global warming, and the mountain pine beetle (coleoptera: Scolytidae). *American Entomologist*, *47*(3), 160–173.

Lusebrink, I., Erbilgin, N., & Evenden, M. L. (2013). The lodgepole× jack pine hybrid zone in alberta, canada: A stepping stone for the mountain pine beetle on its journey east across the boreal forest? *Journal of chemical ecology*, *39*, 1209–1220.

Martin, G., Ringer, M., Pope, V., Jones, A., Dearden, C., & Hinton, T. (2006). The physical properties of the atmosphere in the new hadley centre global environmental model (hadgem1). part i: Model description and global climatology. *Journal of Climate*, *19*(7), 1274–1301.

Mccambridge, W. F. (1971). Temperature limits of flight of the mountain pine beetle, dendroctonus ponderosae. *Annals of The Entomological Society of America*, *64*, 534–535.

McCullagh, P. (1985). On the asymptotic distribution of pearson's statistic in linear exponential-family models. *International Statistical Review/Revue Internationale de Statistique*, 61–67.

Meng, S. X., Huang, S., Yang, Y., Trincado, G., & VanderSchaaf, C. L. (2009). Evaluation of population-averaged and subject-specific approaches for modeling the dominant or codominant height of lodgepole pine trees. *Canadian Journal of Forest Research*, *39*(6), 1148–1158.

Monserud, R. A., Yang, Y., Huang, S., & Tchebakova, N. (2008). Potential change in lodgepole pine site index and distribution under climatic change in alberta. *Canadian Journal of Forest Research*, *38*(2), 343–352.

Moss, E. (1949). Natural pine hybrids in alberta. *Canadian Journal of Research*, *27*(5), 218–229.

Nealis, V., & Peter, B. (2008). *Risk assessment of the threat of mountain pine beetle to canada's boreal and eastern pine forests* (tech. rep. BC-X-417). Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre. Victoria, BC.

Nelson, W. A., Potapov, A., Lewis, M. A., Hundsdörfer, A. E., & He, F. (2008). Balancing ecological complexity in predictive models: A reassessment of risk models in the mountain pine beetle system. *Journal of Applied Ecology*, *45*(1), 248–257.

Pacific Forestry Centre & Carroll, Allan L. (2006). *Impacts of climate change on range expansion by the mountain pine beetle* (Vol. 2006).

Preisler, H. K., Hicke, J. A., Ager, A. A., & Hayes, J. L. (2012). Climate and weather influences on spatial temporal patterns of mountain pine beetle populations in washington and oregon. *Ecology*, *93*(11), 2421–2434.

Raffa, K. F., Aukema, B. H., Bentz, B. J., Carroll, A. L., Hicke, J. A., Turner, M. G., & Romme, W. H. (2008). Cross-scale drivers of natural disturbances prone to anthropogenic amplification: The dynamics of bark beetle eruptions. *Bioscience*, *58*(6), 501–517.

Régnière, J., & Bentz, B. (2007). Modeling cold tolerance in the mountain pine beetle, dendroctonus ponderosae. *Journal of insect physiology*, *53*(6), 559–572.

Rice, A. V., Thormann, M. N., & Langor, D. W. (2007a). Mountain pine beetle associated blue-stain fungi cause lesions on jack pine, lodgepole pine, and lodgepole × jack pine hybrids in alberta. *Botany*, *85*(3), 307–315.

Rice, A. V., Thormann, M. N., & Langor, D. W. (2007b). Virulence of, and interactions among, mountain pine beetle associated blue-stain fungi on two pine species and their hybrids in alberta. *Botany*, *85*(3), 316–323.

Ringer, M., Martin, G., Greeves, C., Hinton, T., James, P., Pope, V., Scaife, A., Stratton, R., Inness, P., & Slingo, J. (2006). The physical properties of the atmosphere in the new hadley centre global environmental model (hadgem1). part ii: Aspects of variability and regional climate. *Journal of climate*, *19*(7), 1302–1326.

Robertson, C., Nelson, T. A., Jelinski, D. E., Wulder, M. A., & Boots, B. (2009a). Spatial–temporal analysis of species range expansion: The case of the mountain pine beetle, dendroctonus ponderosae. *Journal of Biogeography*, *36*(8), 1446–1458.

Robertson, C., Nelson, T. A., Jelinski, D. E., Wulder, M. A., & Boots, B. (2009b). Spatial–temporal analysis of species range expansion: The case of the mountain pine beetle, dendroctonus ponderosae. *Journal of Biogeography*, *36*(8), 1446–1458.

Rosenberger, D. W., Aukema, B. H., & Venette, R. C. (2017a). Cold tolerance of mountain pine beetle among novel eastern pines: A potential for trade-offs in an invaded range? *Forest Ecology and Management*, *400*, 28–37.

Rosenberger, D. W., Aukema, B. H., & Venette, R. C. (2017b). Cold tolerance of mountain pine beetle among novel eastern pines: A potential for trade-offs in an invaded range?. *Forest Ecology Management*, *400*, 28–37.

Rudolph, T., & Laidly, P. (1990). Pinus banksiana lamb. jack pine. *Silvics of North America*, *1*, 280–293.

Rweyongeza, D. M., Dhir, N. K., Barnhardt, L. K., Hansen, C., & Yang, R.-C. (2007). Population differentiation of the lodgepole pine (pinus contorta) and jack pine (pinus banksiana) complex in alberta: Growth, survival, and responses to climate. *Botany*, *85*(6), 545–556.

Safranyik, L. (1974). Management of lodgepole pine to reduce losses from the mountain pine beetle.

Safranyik, L. (1978). Effects of climate and weather on mountain pine beetle populations [25-27 April 1978]. In A. A. Berryman, G. D. Amman, & R. W. Stark (Eds.), *Proceedings, symposium: Theory and practice of mountain pine beetle management in lodgepole pine forests* (pp. 77–84). University of Idaho Forest, Wildlife; Range Experiment Station.

Safranyik, L. (1989). Mountain pine beetle: Biology overview. *Proceedings: Symposium on the Management of Lodgepole Pine to Minimize Losses to the Mountain Pine Beetle. USDA Forest Service, Intermountain Forest and Range Experiment Station, Gen. Tech. Rep. INT-262*, 9–12.

Safranyik, L., & Carroll, A. L. (2007). The biology and epidemiology of the mountain pine beetle in lodgepole pine forests. *The mountain pine beetle: a synthesis of biology, management and impacts on lodgepole pine*, 3–66.

Safranyik, L., Carroll, A. L., Régnière, J., Langor, D., Riel, W., Shore, T. L., Peter, B., Cooke, B. J., Nealis, V., & Taylor, S. W. (2010). Potential for range expansion of mountain pine beetle into the boreal forest of north america. *The Canadian Entomologist*, *142*(5), 415–442.

Safranyik, L., Shrimpton, D., & Whitney, H. (1975). An interpretation of the interaction between lodgepole pine, the mountain pine beetle and its associated blue stain fungi in western canada. *Management of lodgepole pine ecosystems*, *1*, 406–428.

Safranyik, L., & Wilson, B. (2006). The mountain pine beetle: A synthesis of biology, management and impacts on lodgepole pine. https://api.semanticscholar.org/CorpusID:83321357

Sambaraju, K. R., Carroll, A. L., Zhu, J., Stahl, K., Moore, R. D., & Aukema, B. H. (2012). Climate change could alter the distribution of mountain pine beetle outbreaks in western canada. *Ecography*, *35*(3), 211–223.

Sauchyn, D. J., Stroich, J., & Beriault, A. (2003). A paleoclimatic context for the drought of 1999–2001 in the northern great plains of north america. *Geographical Journal*, *169*(2), 158–167.

Seybert, J. P., & Gara, R. (1970). Notes on flight and host-selection behavior of the pine engraver, ips pini (coleoptera: Scolytidae). *Annals of the Entomological Society of America*, *63*(4), 947–950.

Shepherd, R. (1966). Factors influencing the orientation and rates of activity of dendroctonus ponderosae hopkins (coleoptera: Scolytidae) 1. *The Canadian Entomologist*, *98*(5), 507–518.

Srivastava, V., & Carroll, A. L. (2023). Dynamic distribution modelling using a native invasive species, the mountain pine beetle. *Ecological Modelling*, *482*, 110409.

Stage, A. R. (1976). An expression for the effect of aspect, slope, and habitat type on tree growth. *Forest Science*, *22*(4), 457–460.

Sustainable Resource Development. (2007). Mountain pine beetle : Management strategy [Accessed: December 2007]. https://open.alberta.ca/dataset/68b62d2d-98f3-4750-815a-ba5dac44fb68/resource/4dd889d3-441e-43fd-b39e-4c90ec4bb123/download/mpb-managementstrategy-dec2007.pdf

Suter II, G. W. (2016). *Ecological risk assessment*. CRC press.

Taylor, S. W., & Safranyik, L. (2003). Effect of climate change on range expansion by the mountain pine beetle in british columbia.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Wang, H., Li, R., & Tsai, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, *94*(3), 553–568.

Wang, Z., Ma, S., Zappitelli, M., Parikh, C., Wang, C.-Y., & Devarajan, P. (2016). Penalized count data regression with application to hospital stay after pediatric cardiac surgery. *Statistical methods in medical research*, *25*(6), 2685–2703.

Westfall, J., & Ebata, T. (2007). *2007 summary of forest health conditions in british columbia*. Ministry of Forests; Range, Forest Practices Branch.

Wulder, M. A., Ortlepp, S. M., White, J. C., Coops, N. C., & Coggins, S. B. (2009). Monitoring the impacts of mountain pine beetle mitigation. *Forest ecology and management*, *258*(7), 1181–1187.

Wulder, M. A., White, J., Bentz, B., Alvarez, M., & Coops, N. (2006). Estimating the probability of mountain pine beetle red-attack damage. *Remote Sensing of Environment*, *101*(2), 150–166.

Yang, R.-C., Ye, Z., & Hiratsuka, Y. (1999). Susceptibility of pinus contorta-pinus banksiana complex to endocronartium harknessii: Host-pathogen interactions. *Canadian journal of botany*, *77*(7), 1035–1043.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*(2), 894–942.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Appendix A: Appendices to chapter 2

A.1 Spatial Auto-correlation

Spatial autocorrelation is a common phenomenon in ecological data. In the context of the mountain pine beetle, the influence of dispersal on infestations cannot be neglected. Although utilizing grid cells of 500m in size leads us to expect reduced spatial autocorrelation, it remains essential to examine how spatial autocorrelation manifests in each year's data. We employ the Moran's I statistic along with hypothesis tests to reveal the autocorrelation in each year's data within the area depicted in Figure A.1. Based on Carroll et al. (2017), we assume that the influence of an infestation is up to 4 km. We conduct two-sided tests with the null hypothesis stating that there are no spatial autocorrelations in the adjacent cells. As shown in Table A.6, the value of Moran's I statistic is relatively small. However, the hypothesis tests have a value 0 indicating that spatial auto-correlation plays a significant role in this area.

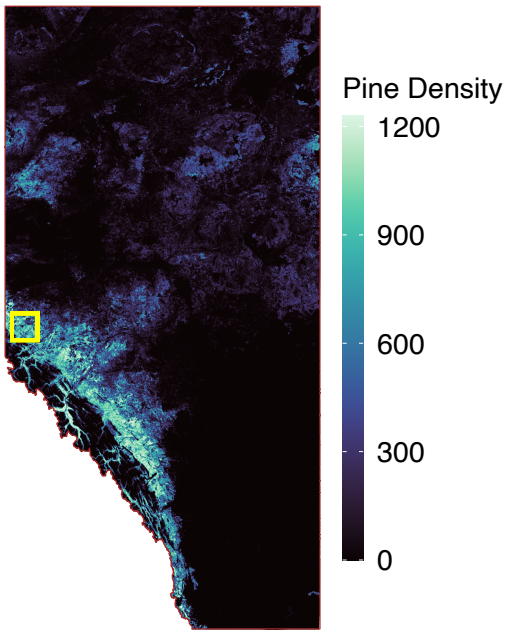


Figure A.1: The position of the selected area in Alberta map along with pine density of lodgepole pine and jack pine (stems/hectare).

Year	Moran's I	P(I>Z)
2007	0.14	0
2008	0.16	0
2009	0.32	0
2010	0.42	0
2011	0.42	0
2012	0.27	0
2013	0.27	0
2014	0.3	0
2015	0.28	0
2016	0.28	0
2017	0.27	0
2018	0.16	0
2019	0.14	0
2020	0.11	0

Table A.1: Moran's I statistic with p-value

A.2 Age-height model

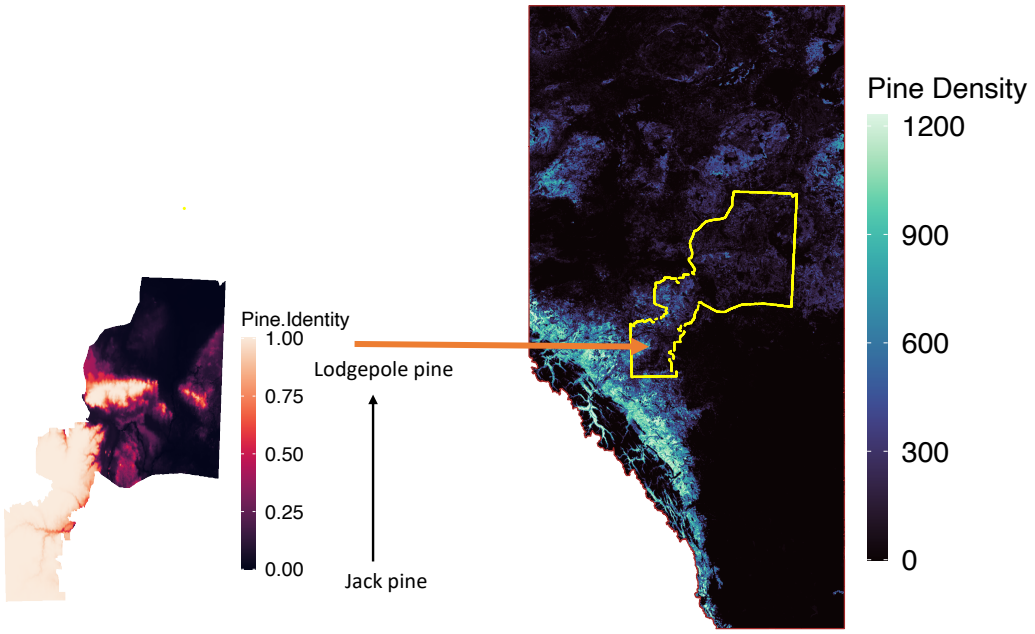


Figure A.2: Selected area in the eastern Alberta

To develop age-height models, we utilize the available age and height data from AVI, with a specific focus on the geographical area discussed in the chapter 3 as shown in Figure A.2. In this region, we can observe the genetic transition of the dominant pine species in Alberta, shifting from lodgepole pine to jack pine as we move from south to north. We conduct hypothesis testing to ascertain if there are any important differences among the pattern of pine growth among three species. The population-averaged age-height model we considered is (Meng et al., 2009) :

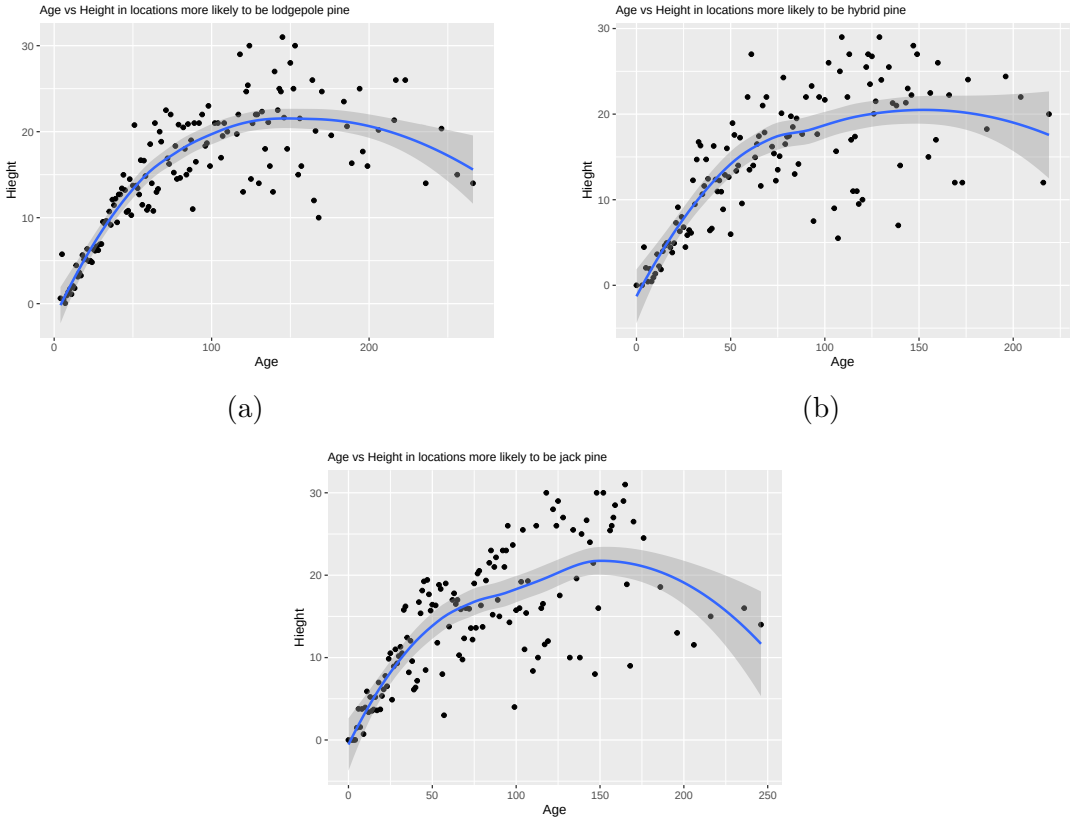
$$\text{Height} = \frac{b1}{1 + e^{b2 + b3\log(\text{Age}+1)}} \tag{A.1}$$

Based on the pine distributions shown in Burns et al. (2019), we group the pines into three species: lodgepole pine, lodgepole-jack hybrid pine, and jack pine. We then calculate the average height for each age group within each species and fit the logistic model to analyze the relationship between age and mean height. Plotting age against

mean height, we observe that the variance of height tends to increase with the age of the pine trees across all three species as depicted in Figure A.3. Thus, to make reliable estimation, we try the following two assumptions :

$$\text{height}_i \sim N(\frac{b1}{1+e^{b2+b3\log(\text{Age}_i+1)}}, \sigma^2 = b4 + b5Age_i) \tag{A.2}$$

$$\log(\text{height}_i) \sim N(\log(\text{b1}) - \log(1+\exp(b2+b3*x)), \sigma^2 = b4) \tag{A.3}$$



(a) (b)

(c)

Figure A.3: Age-Height plots for three species ((a)lodgepole pine, (b)hybrid pine and (c)jack pine).

The coefficients obtained from the models fitted to the three individual species show no important differences compared to the model fitted to all the pines (presented in Table A.2, Table A.3, Table A.4, and Table A.5). Additionally, all of the coefficients fall within the 95% confidence interval, as illustrated in Figure A.4 and Figure A.5. The R-Squared value indicates that combining all the species into a single model only

slightly reduced the model's performance, as shown in Table A.6. Although the R-Squared value is slightly lower than that of the lodgepole pine model, it still exceeds 0.7. Therefore, we use the model fitted to all pines to estimate height.

Hypothesis	b1	b2	b3	b4	b5	log-Lik
equation(2.13)	22.5972114	7.2104013	-1.9434286	1.2073589	0.1191352	-369.7866
equation(2.14)	22.0838185	7.1539065	-1.9202749	0.1433188	0	-427.197

Table A.2: Table of Parameters fitted to lodgepole pine models.

Hypothesis	b1	b2	b3	b4	b5	log-Lik
equation(2.13)	21.62706194	6.61550130	-1.84729303	1.34628243	0.04048958	-373.9831
equation(2.14)	22.1312733	5.9732666	-1.6034224	0.4086919	0	-412.5634

Table A.3: Table of Parameters fitted to the hybrid pine models.

Hypothesis	b1	b2	b3	b4	b5	log-Lik
equation(2.13)	22.70865440	5.53890111	-1.53347438	0.96496813	0.04969931	-396.9747
equation(2.14)	18.0504509	7.4675307	-2.2637383	0.5157982	0	-454.1574

Table A.4: Table of Parameters fitted to the jack pine models.

Hypothesis	b1	b2	b3	b4	b5	log-Lik
equation(2.13)	23.30350240	6.38320129	-1.69921327	0.39726618	0.03541758	-439.2946
equation(2.14)	23.0833140	5.9968450	-1.5804953	0.2607932	0	-507.2572

Table A.5: Table of Parameters for the models fitted to all pines in this area.

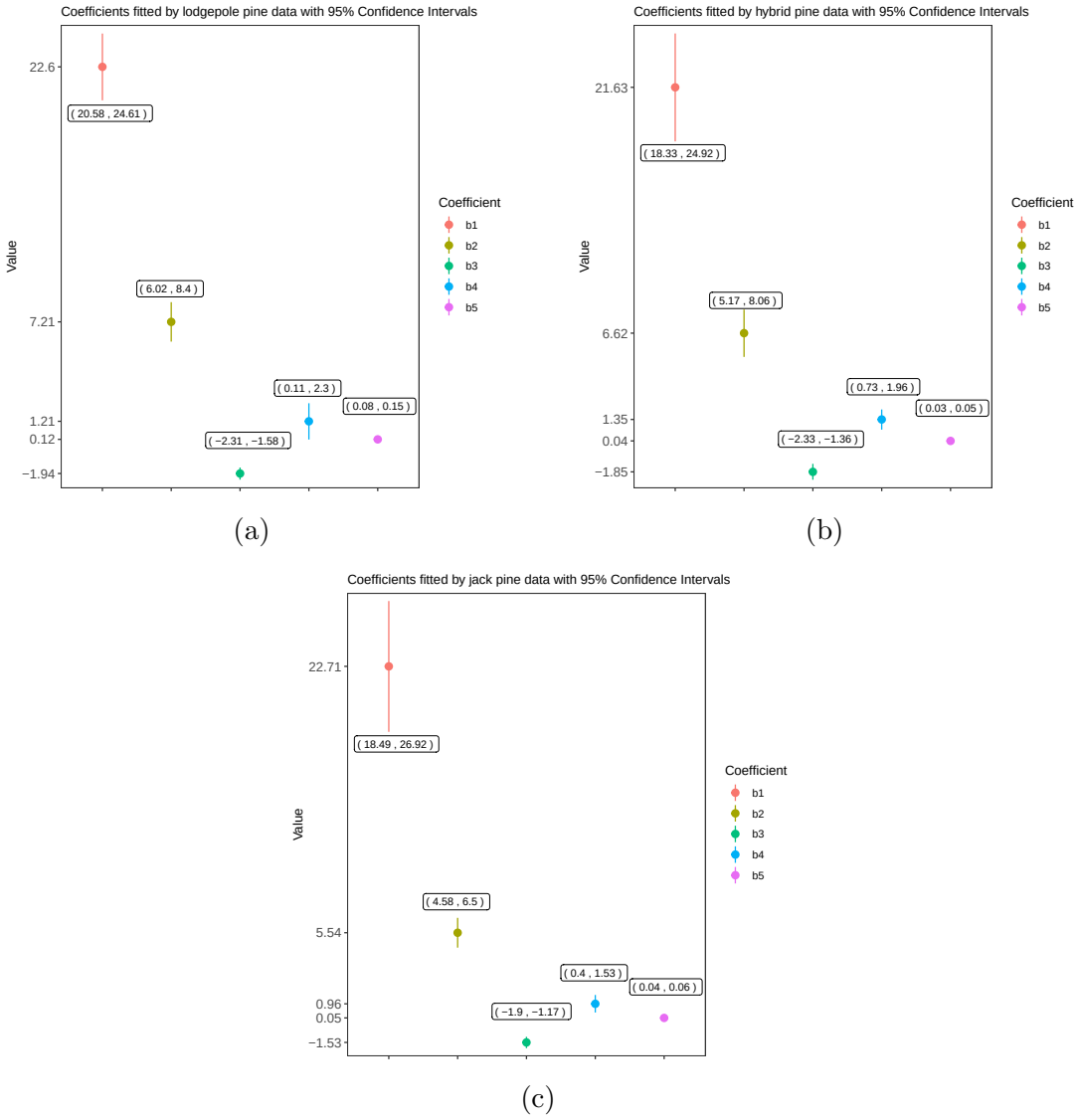


Figure A.4: Coefficients with 95% confidence interval under hypothesis equation(2.13)((a)lodgepole pine, (b)hybrid pine and (c)jack pine).

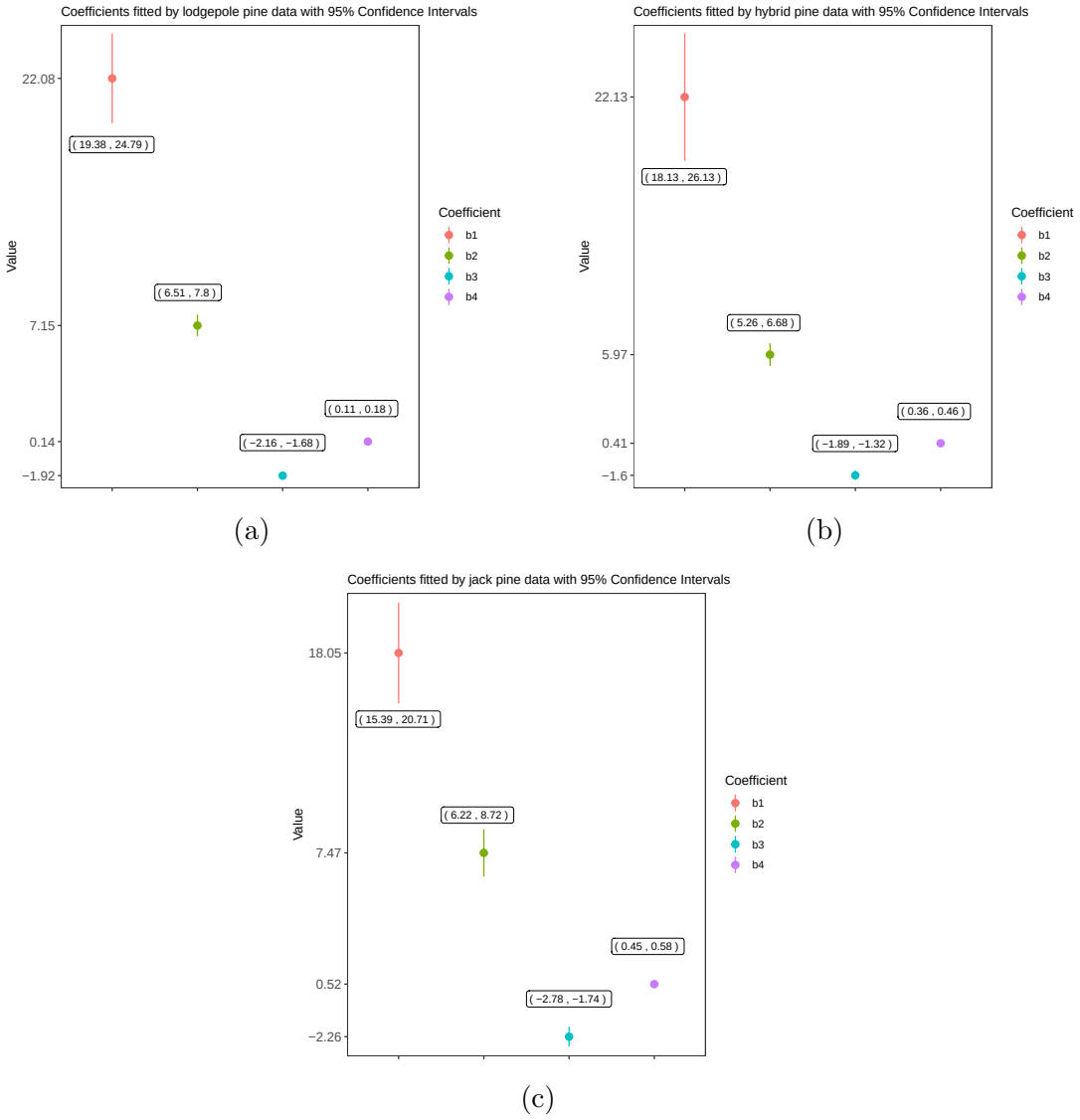


Figure A.5: Coefficients with 95% confidence interval under hypothesis equation(2.14) ((a)lodgepole pine, (b)hybrid pine and (c)jack pine).

Model	R Square
lodgepole pine	0.7409338
hybrid pine	0.6271369
jack pine	0.6271369
all three pines	0.7122566

Table A.6: R-Squared value of four fitted models.

Appendix B: Appendices to chapter 3

B.1 Results of the endemic stage

The endemic stage presents less significant covariates, with most of the confidence intervals including zero in both models. The confidence intervals for pine identity also include zero, suggesting a minimal influence of pine species on infestations. This minimal influence leads us not to estimate the absolute and relative risks, as well as the reproduction rate. The optimal penalties, coefficients, and their 95% confidence intervals are shown below.

	$\alpha_{\text{abundance}}$	α_{presence}
endemic stage	0.25	1

Table B.1: The optimal weights of penalties for the outbreak stage and endemic stage.

Covariates	Presence Model	Abundance Model
Northness	0	-0.074
Easterness	0	-0.019
Slope	1.09	-0.008
Pine identity	2.4	0.539
Pine density	0.053	0.155
Age	0.264	0.326
Minimum temperature in summer	0	-0.056
Maximum temperature in summer	0	0.93
Relative Humidity	0	0.191
Wind Speed	0	-0.394
Soil Moisture Index	0	0.302
Degree Days	0.521	-1.014
Overwinter Survival Rate	0	0.805
Last year infestations	0	16.311
Dispersal impacts within 4 km	0	11.718

Table B.2: The coefficients of the standardized covariates in the endemic data model.

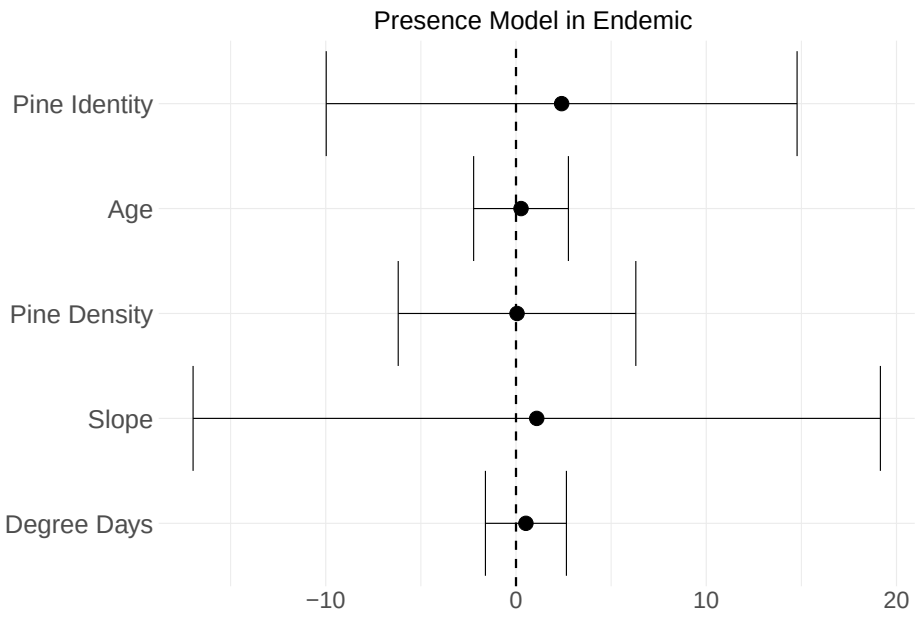


Figure B.1: The 95% confidence intervals of the standardized non-zero coefficients of our covariates in the presence model in the endemic stage. Last year infestations and dispersal impacts within 4 km are 0 in the presence model.

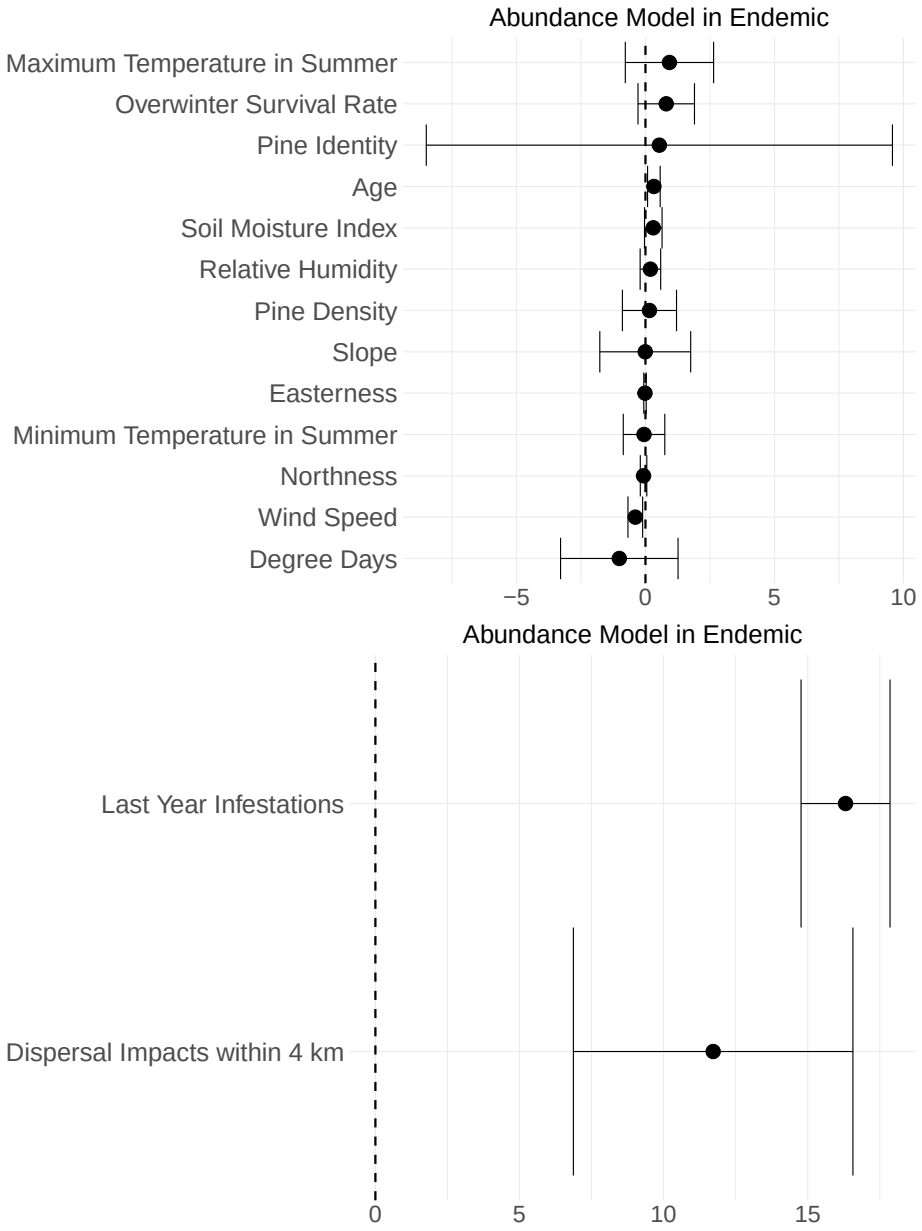
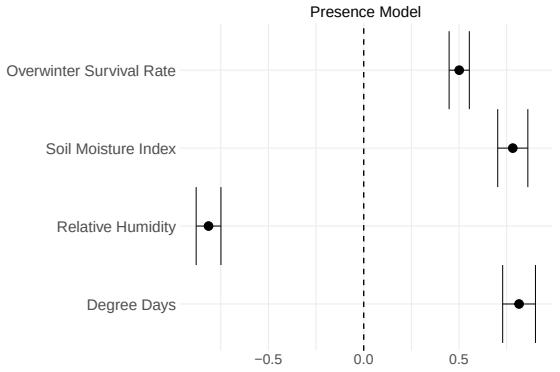


Figure B.2: The 95% confidence intervals of the standardized non-zero coefficients of our covariates in the abundance model in the endemic stage.

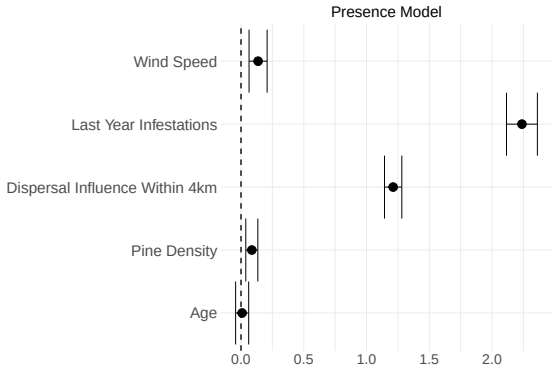
Appendix C: Appendices to chapter 4

C.1 Plots of the growth peak declining stage in the selected area

The results in the growth peak declining stage differ from those in the outbreak stage, showing that the soil moisture index is significant in the presence model, and degree days exhibit a negative sign in the abundance model, as shown in Figures C.1 and C.2. We also present the historical values of four climate-related covariates during the growth peak declining stage in Figure C.3. The overall trends under climate change also show differences. As temperature rises, relative humidity negatively impacts the probability of occurrence but increases the number of infestations, as demonstrated in Figures C.4 and C.5. A similar trend is observed for the soil moisture index and degree days. As illustrated in Figures C.6, C.7, C.8, and C.9, these covariates decrease the probability of infestation and increase the numbers. The overwinter survival rate positively affects infestations by increasing the probability of occurrence, although its impact on abundance is minimal, as shown in Figures C.10 and C.11. Overall, the trend in the growth peak declining stage indicates a decrease in probability and an increase in numbers, with local variations depicted in Figures C.12 and C.13.

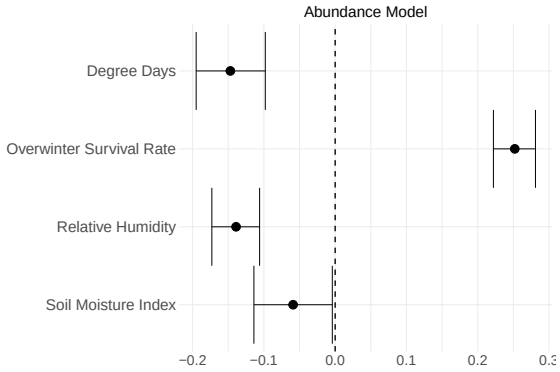


(a) The coefficients of the four standardized climate covariates in the presence model in the growth peak declining stage.

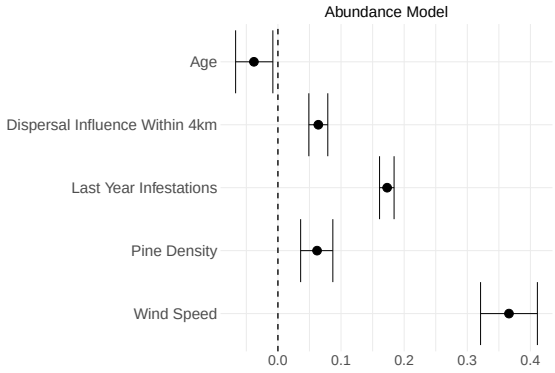


(b) The coefficients of the five standardized non-climate covariates in the presence model in the growth peak declining stage.

Figure C.1: Presence model coefficients for the selected area (Figure 4.1) in the growth peak declining stage.

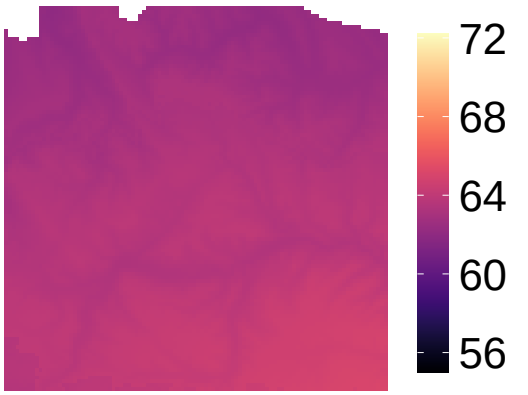


(a) The coefficients of the four standardized climate covariates in the abundance model in the growth peak declining stage.

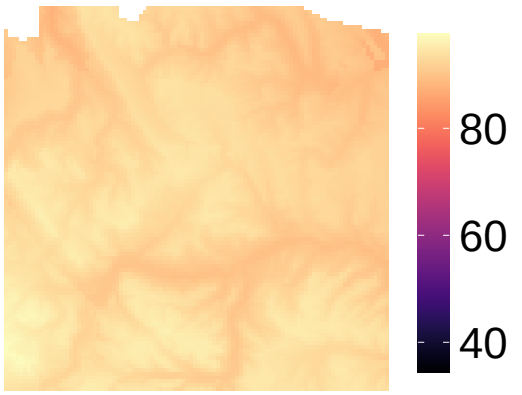


(b) The coefficients of the five standardized non-climate covariates in the abundance model in the growth peak declining stage.

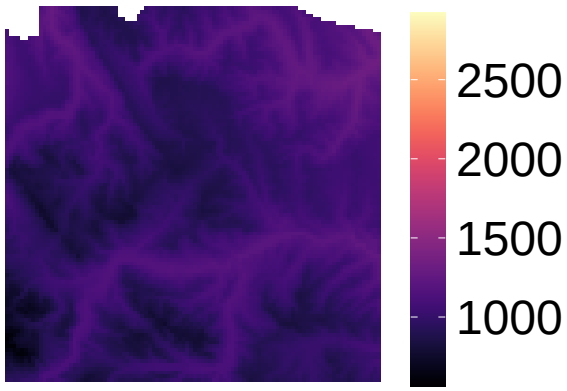
Figure C.2: Abundance model coefficients for the selected area (shown in Figure 4.1) in the outbreak stage.



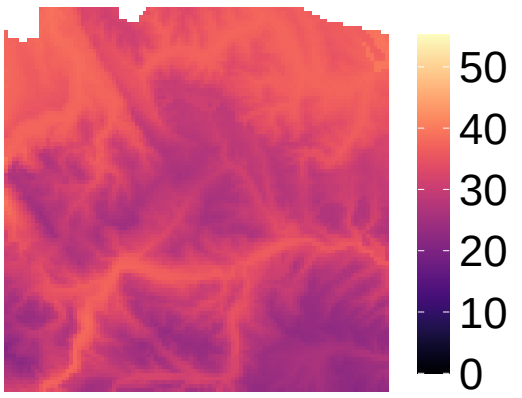
(a) Mean of the relative humidity from 2015 to 2020 in the selected area (shown in Figure 4.1).



(b) Mean of the soil moisture index from 2015 to 2020 in the selected area (shown in Figure 4.1).

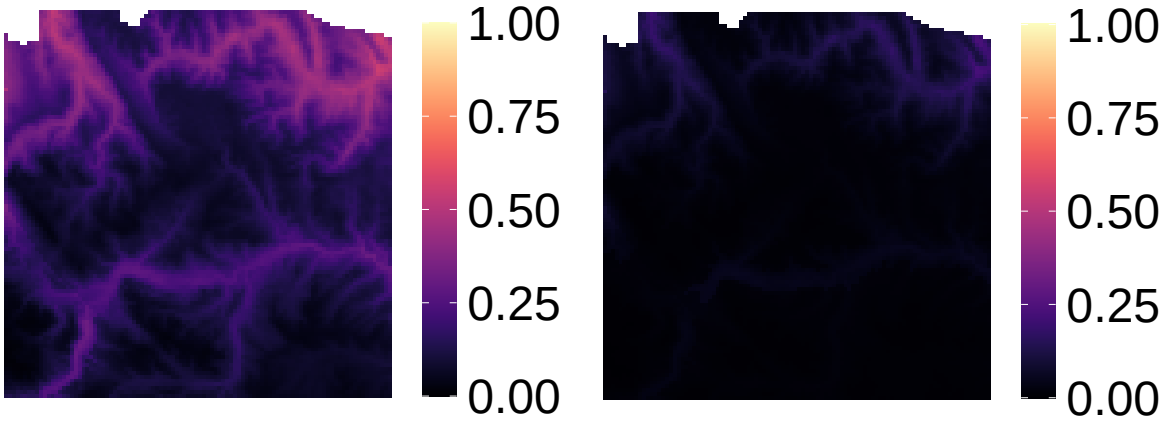


(c) Mean of the degree days from 2015 to 2020 in the selected area (shown in Figure 4.1).



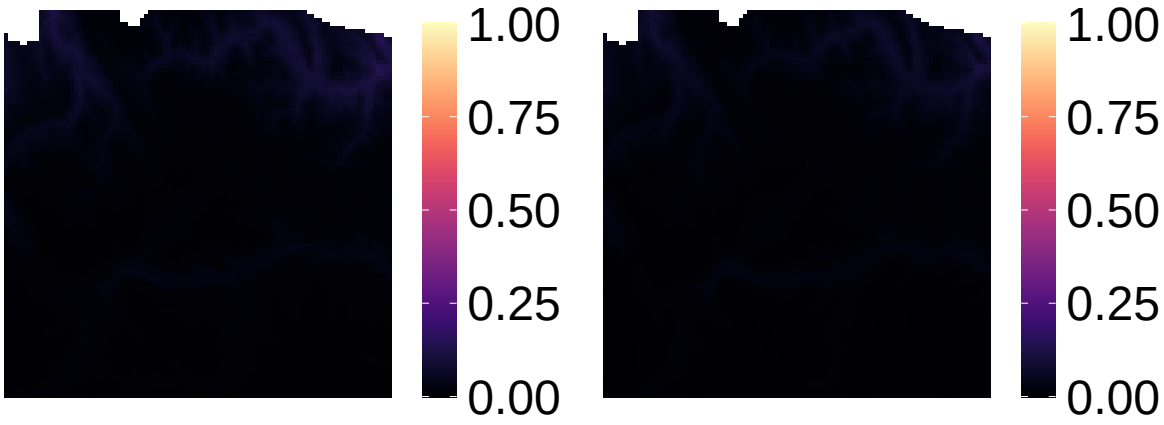
(d) Mean of the overwinter survival rate from 2015 to 2020 in the selected area (shown in Figure 4.1).

Figure C.3: Historical values of four climate-related covariates in the growth peak declining stage.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.

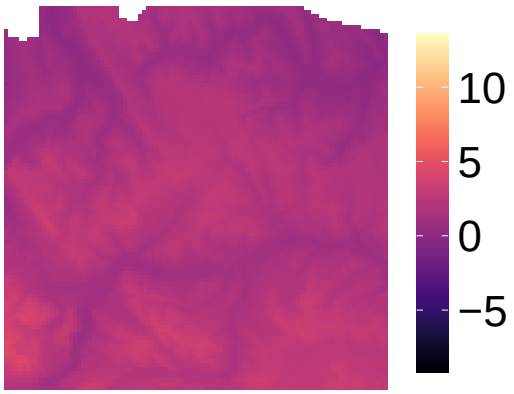
(b) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



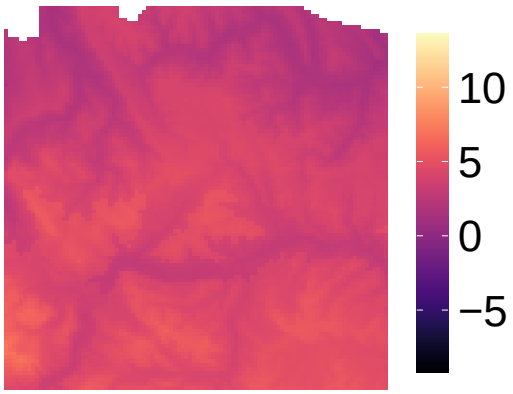
(c) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

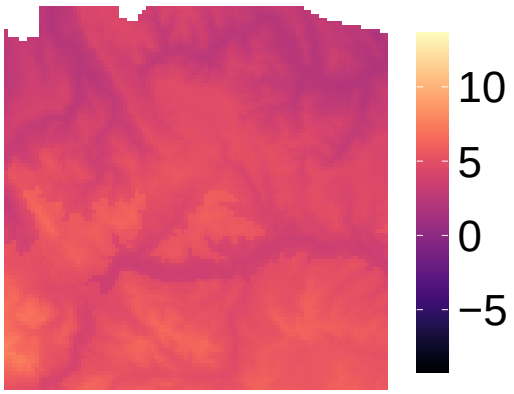
Figure C.4: Comparative analysis of current and future presence under three RCP scenarios, illustrating a decreasing of probability of occurrence as temperatures increase when varying relative humidity. Only relative humidity is varied under four scenes.



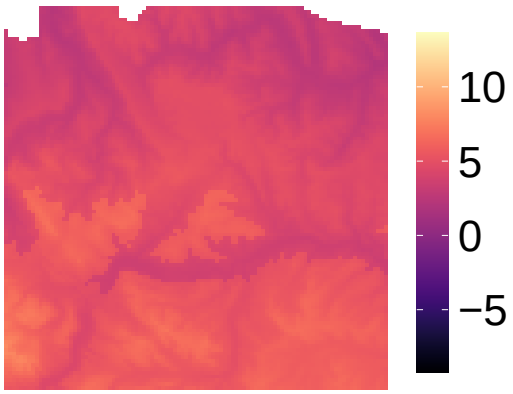
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.

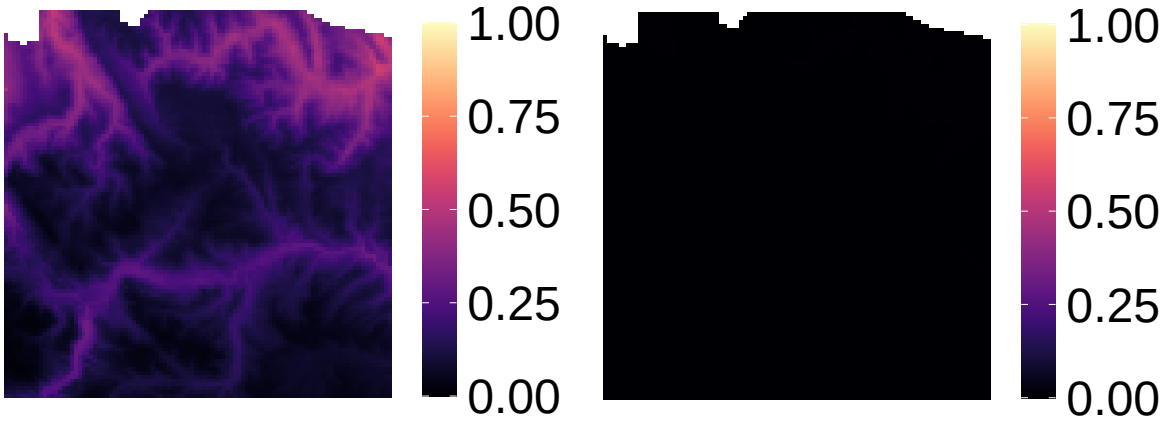


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.



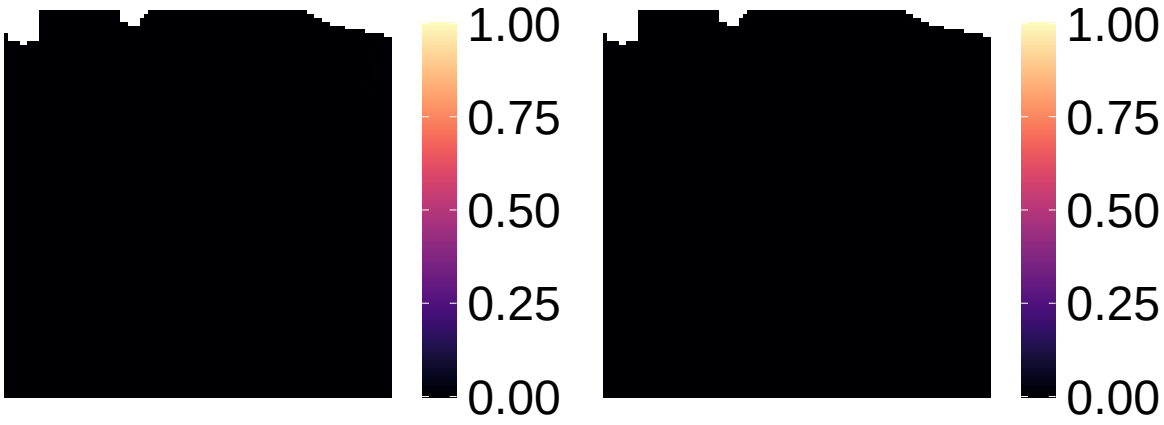
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except relative humidity in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

Figure C.5: Comparative analysis of current and future abundance under RCP scenarios, illustrating an increasing of number of infestations as temperatures increase when varying relative humidity. Only relative humidity is varied under four scenes.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.

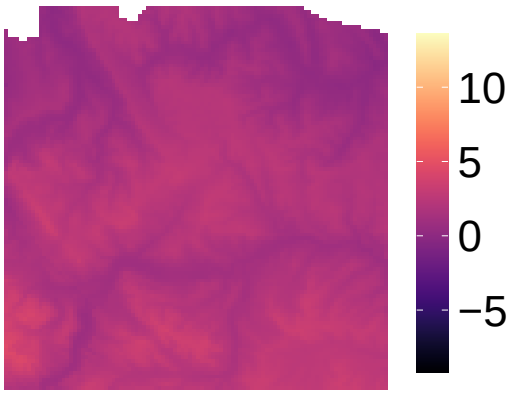
(b) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



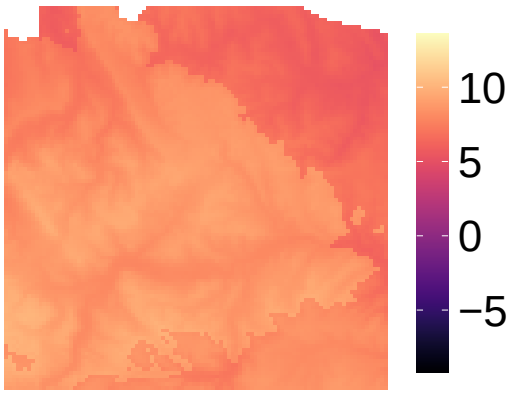
(c) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

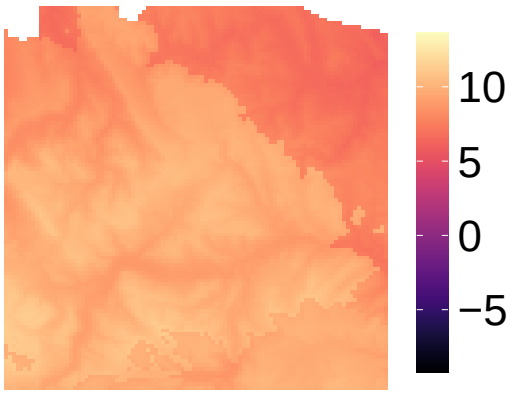
Figure C.6: Comparative analysis of current and future presence under three RCP scenarios, illustrating a decreasing of probability of occurrence as temperatures increase when varying soil moisture index. Only soil moisture index is varied under four scenes.



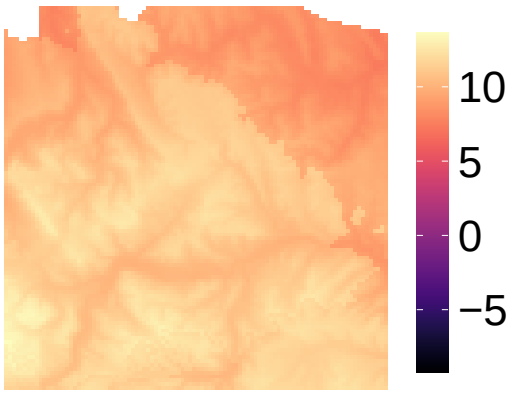
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.

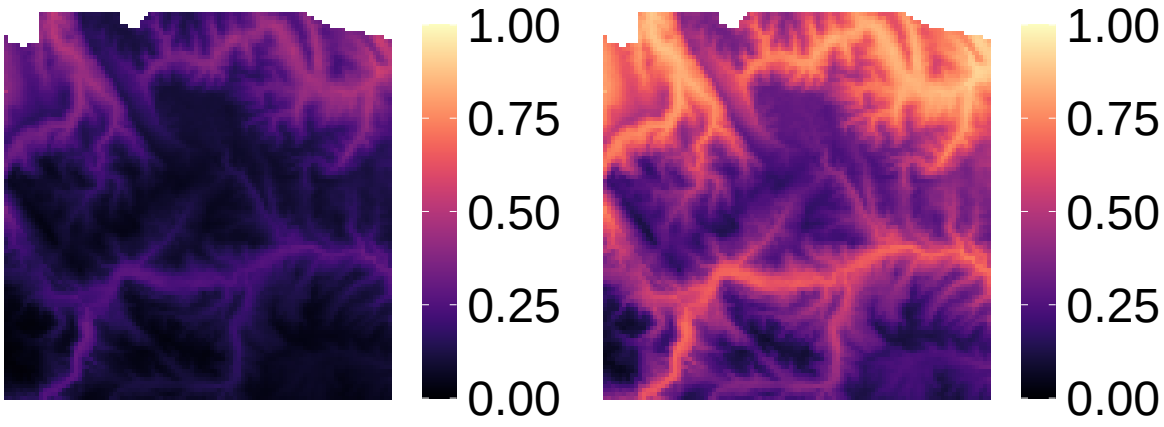


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.



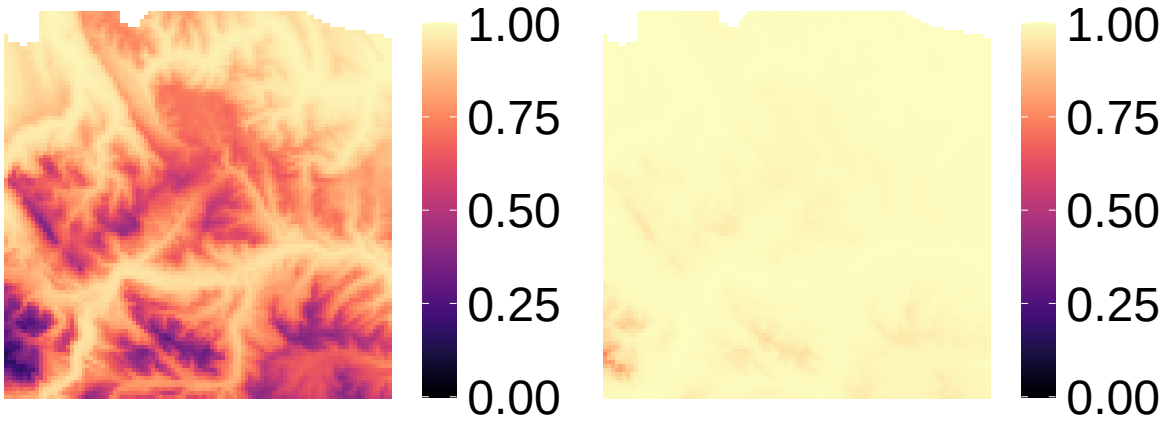
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except soil moisture index in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

Figure C.7: Comparative analysis of current and future abundance under three RCP scenarios, illustrating an increasing of number of infestations as temperatures increase when varying soil moisture index. Only soil moisture index is varied under four scenes.



(a) Probability of occurrence estimated in the presence model in selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.

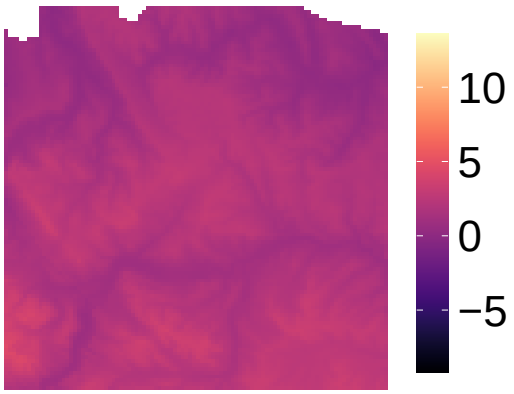
(b) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



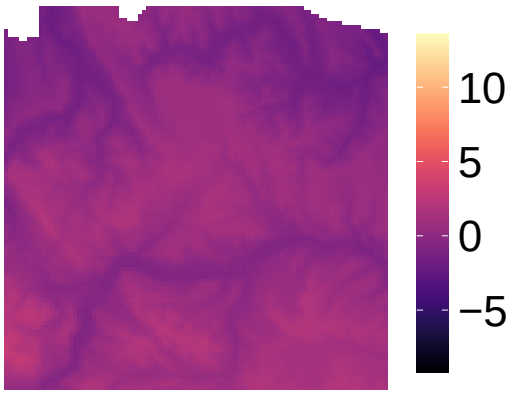
(c) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

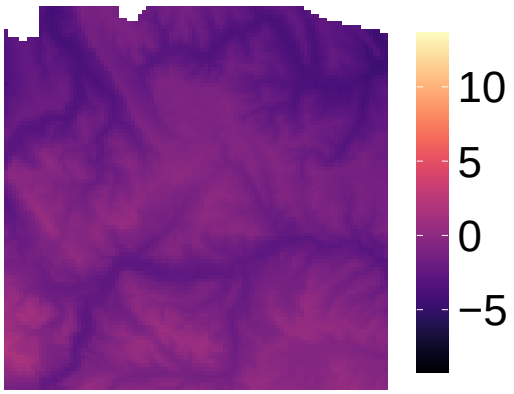
Figure C.8: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperature increase when varying degree days. Only degree days is varied under four scenes.



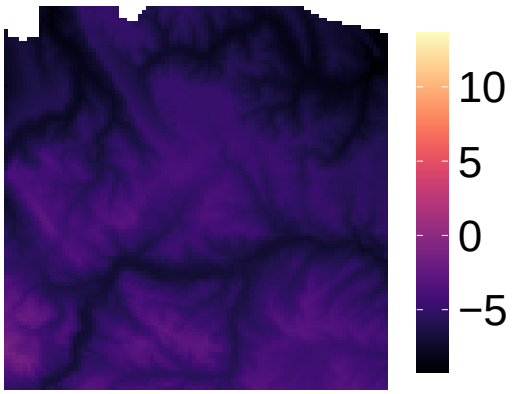
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.

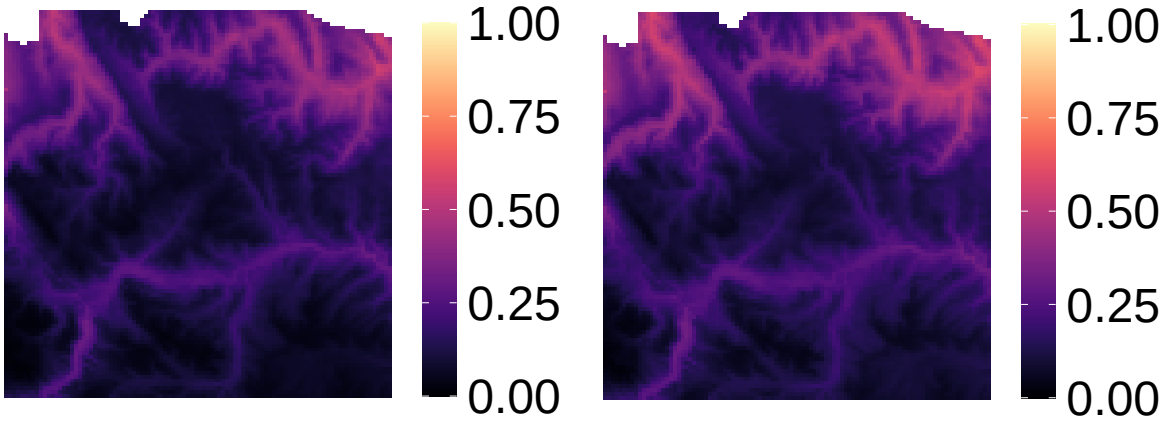


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.



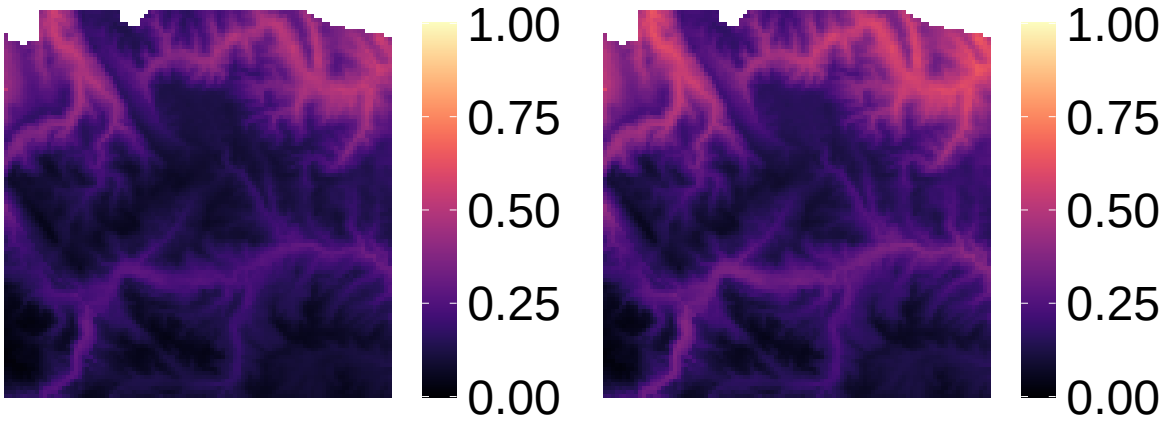
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except degree days in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

Figure C.9: Comparative analysis of current and future abundance under three RCP scenarios, illustrating a decreasing of numbers of infestations as temperatures increase when varying degree days. Only degree days is varied under four scenes.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.

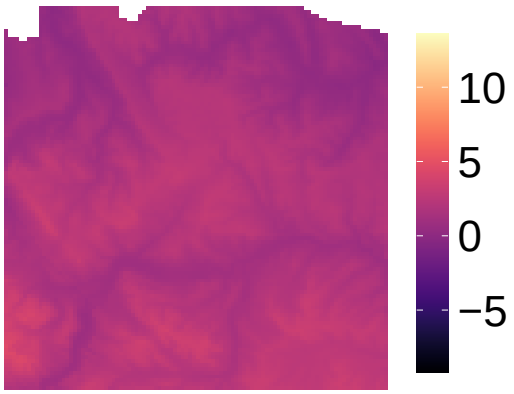
(b) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



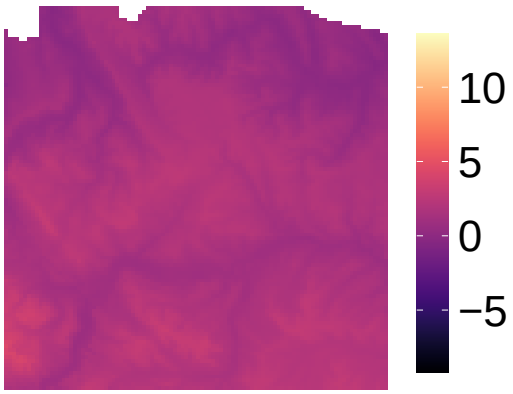
(c) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

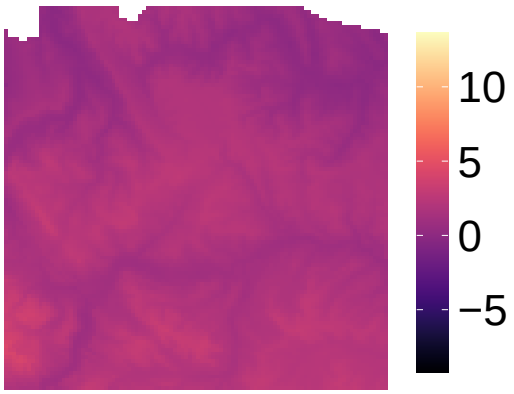
Figure C.10: Comparative analysis of current and future presence under three RCP scenarios, illustrating an increasing of probability of occurrence as temperature increase when varying overwinter survival rate. Only overwinter survival rate is varied under four scenes.



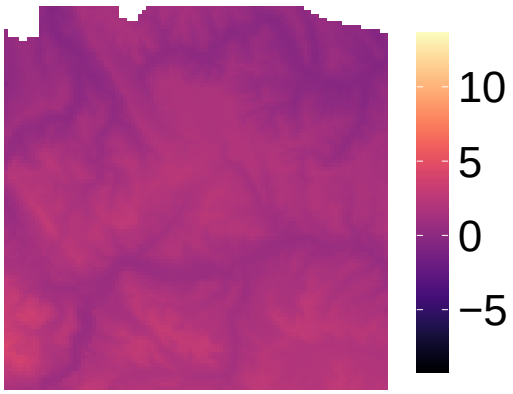
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate value in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.

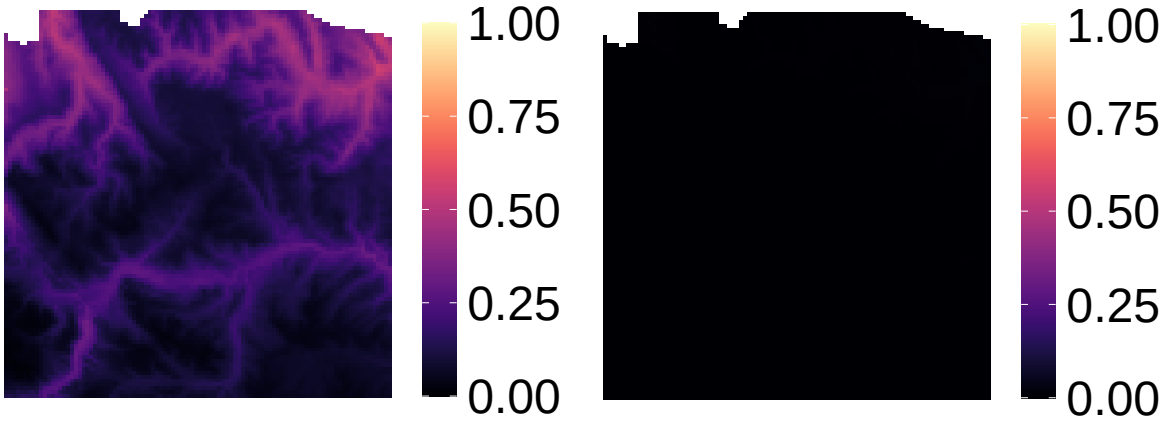


(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.



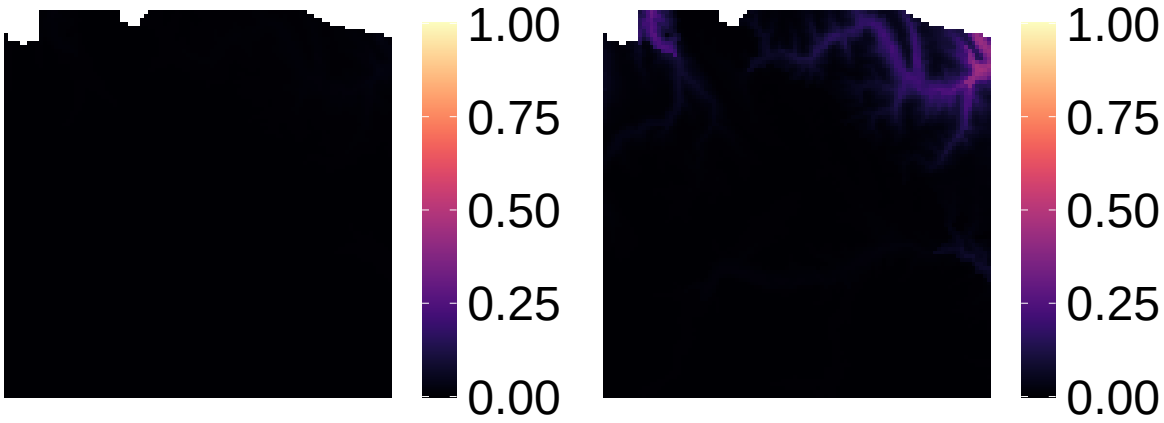
(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except overwinter survival rate in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

Figure C.11: Comparative analysis of current and future abundance under three RCP scenarios, illustrating a decreasing of the number of infestations as temperatures increase when varying overwinter survival rate. Only overwinter survival rate is varied under four scenes.



(a) Probability of occurrence estimated in the presence model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.

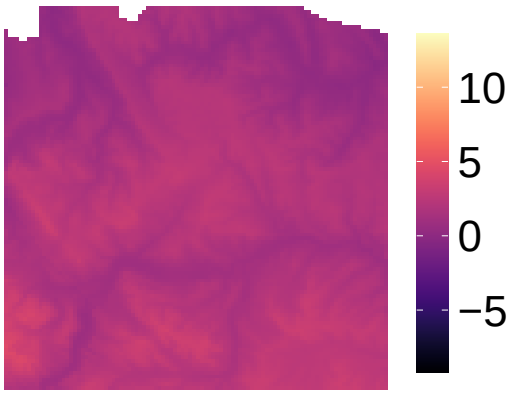
(b) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



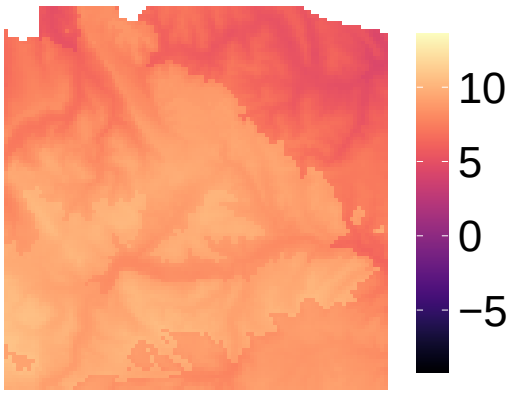
(c) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.

(d) Probability of occurrence estimated in the presence model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5 in 2100.

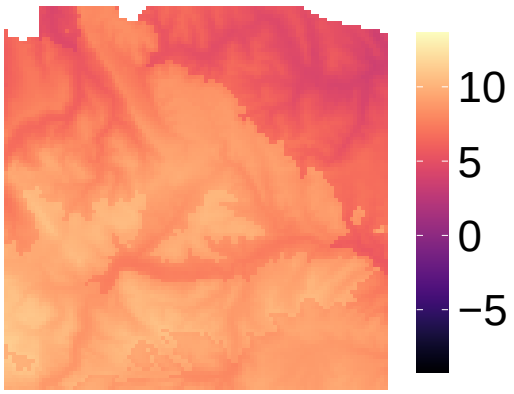
Figure C.12: Comparative analysis of current and future presence under three RCP scenarios, illustrating a decreasing of probability of occurrence with local variations as temperature increase when varying four covariates. Only four climate covariates are varied under four scenes.



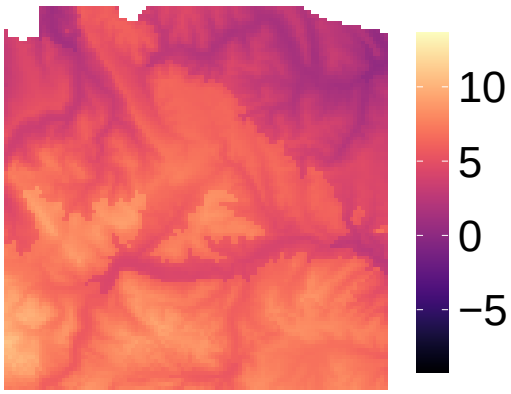
(a) log scale of the expected infestations estimated in the abundance model in the selected area (shown in Figure 4.1) using growth peak declining model and fixing all covariates except beetle pressure by median for cells.



(b) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 2.6 in 2100.



(c) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 4.5 in 2100.



(d) log scale of the expected infestations estimated in the abundance model while fixing all covariates except four climate covariates in the selected area (shown in Figure 4.1) using growth peak declining model under RCP 8.5.

Figure C.13: Comparative analysis of current and future abundance under three RCP scenarios, illustrating an increasing of the number of infestations with local variations as temperatures increase when varying four covariates simultaneously. Only four climate covariates are varied under four scenes.