Fall 12-15-2023

Integrating Remote Sensing and Machine Learning to Assess Forest Health and Susceptibility to Pest-induced Damage

Rajeev Bhattarai
University of Maine, rajeev.bhattarai@maine.edu

INTEGRATING REMOTE SENSING AND MACHINE LEARNING TO ASSESS FOREST HEALTH AND SUSCEPTIBILITY TO PEST-INDUCED DAMAGE

By Rajeev Bhattarai

B.S. Tribhuvan University, Nepal, 2017

M.S. University of Maine, 2020

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy

(in Forest Resources)

The Graduate School

The University of Maine

December 2023

Advisory Committee:

Parinaz Rahimzadeh-Bajgiran, Associate Professor of Remote Sensing of Natural Resources, School of Forest Resources, University of Maine, Orono, Maine, USA, Advisor

Aaron Weiskittel, Professor of Forest Biometrics and Modeling, School of Forest Resources, University of Maine, Orono, Maine, USA

Saeid Homayouni, Associate Professor, Institut National de la Recherche Scientifique, Centre Eau Terre Environment, Quebec, Canada

Ryan P. Hanavan, Forest Entomologist, USDA Forest Service, Missoula, Montana, USA

Angela Mech, Assistant Professor, School of Biology and Ecology, College of Natural Sciences, Forestry, and Agriculture, University of Maine, Orono, Maine, USA

INTEGRATING REMOTE SENSING AND MACHINE LEARNING TO ASSESS FOREST HEALTH AND SUSCEPTIBILITY TO PEST-INDUCED DAMAGE

By Rajeev Bhattarai

Dissertation Advisor: Dr Parinaz Rahimzadeh-Bajgiran

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Forest Resources)
December 2023

Spruce budworm (*Choristoneura fumiferana*; SBW) outbreaks are cyclically occurring phenomena in the northeastern USA and neighboring Canadian provinces. These outbreaks are often of landscape level causing impaired growth and mortality of the host species namely spruce (*Picea* sp.) and balsam fir (*Abies balsamea* (L.) Mill.). Acknowledging the recent SBW outbreak in Canadian provinces like Quebec and New Brunswick neighboring the state of Maine, our study devised comprehensive techniques to assess the susceptibility of Maine forests to SBW attack. This study aims to harness the power of remote sensing data and machine learning algorithms to model and map the susceptibility of forest in terms of host species availability and abundance (basal area per hectare; BAPH, and leaf area index; LAI), their maturity and the defense mechanism prevalent.

In terms of host species abundance mapping our study explores the integration of satellite remote sensing data to model BAPH and LAI of two economically vital SBW host species, red

spruce (*Picea rubens* Sarg.) and balsam fir, in Maine USA. Combining Sentinel-1 synthetic aperture radar (SAR), Sentinel-2 multispectral, and site variables, we used Random Forest (RF) and Multi-Layer Perceptron (MLP) algorithms for modeling LAI and BAPH. The results demonstrated the superiority of RF over MLP, achieving smaller normalized root mean square error (nRMSE) by 0.01 and 0.06 for LAI and BAPH, respectively. Notably, Sentinel-2 variables, especially the red-edge spectral vegetation indices, played a significant role in both LAI and BAPH estimation, with the minor inclusion of site variables, particularly elevation.

In addition, using various satellite remote sensing data such as Sentinel-1 C-band SAR, PALSAR L-band SAR and Sentinel-2 multispectral, along with site variables, the study developed large-scale SBW stand impact types and susceptibility maps for the entire state of Maine. The susceptibility of the forest was assessed based on the availability of SBW host species and their maturity. Integrating machine-learning algorithms, RF and MLP, the best model, utilizing site (elevation and aspect) and Sentinel-2 data achieved an overall accuracy of 83.4% to predict SBW host species. Furthermore, combining the host species data with age data from Land Change Monitoring, Assessment, and Projection (LCMAP) products we could produce the SBW susceptibility map based on stand impact types with an overall accuracy of 88.3%.

Moreover, the work builds upon the assessment of susceptibility of SBW host species taking into account the concentration of several canopy traits using remote sensing and site data. The study focused on various foliar traits affecting insect herbivory, including nutritive such as nitrogen (N), phosphorous (P), potassium (K), and copper (Cu), non-nutritive such as iron (Fe) and calcium (Ca), and defensive parameters such as equivalent water thickness (EWT) and leaf mass per area (LMA). Using Sentinel-2 and site data, we developed trait estimation models using machine-learning algorithms like Random Forest (RF), Extreme Gradient Boosting (XGB), and

Support Vector Machine (SVM). The accuracy of the developed model was evaluated based on the normalized root mean square error (nRMSE). Based on the model performances, we selected XGB algorithm to estimate Ca, EWT, Fe, and K whereas Cu, LMA, N, and P were estimated using RF algorithm. Regarding the variables used, almost all the best performing models included Sentinel-2 red-edge indices and depth to water table (DWT) as the most important variables. Ultimately, the study proposed a novel framework connecting the concentrations of foliar traits in SBW host foliage to tree susceptibility to the pest, enabling the assessment of host susceptibility on a landscape level.

To sum up, this study highlights the advantages and effectiveness of integrating satellite remote sensing data for enhanced pest management, providing valuable insights into tree attributes and susceptibility to spruce budworm outbreaks in Northeast USA. The findings offer essential tools for forest stakeholders to improve management strategies and mitigate potential forthcoming SBW outbreaks in the region.

DEDICATION

I dedicate this thesis to my parents (**Parameshwar Bhattarai** and **Sushila Wagle**), my advisor **Dr. Parinaz Rahimzadeh-Bajgiran**, and **Lord Pashupatinath**.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my advisor, Dr. Parinaz Rahimzadeh-Bajgiran, for her unwavering belief in me and her dedicated support throughout my academic journey. We have been working together with her since 2018 fall when I started my career in academia as a M.S. student. From the very beginning, I found her to be enthusiastic, motivating, kind and exceptionally knowledgeable in her field. Under her guidance, I have witnessed significant growth and progress in my academic pursuits over the courses of five years. Her mentorship has been a true blessing, and I consider myself incredibly fortunate to have had her as my advisor. Most importantly, I owe a substantial part of my accomplishments in academia to her invaluable contributions and guidance.

I extend my heartfelt gratitude to all the members of my dissertation committee, Dr. Aaron Weiskittel, Dr. Saeid Homayouni, Dr. Angela Mech, and Dr. Ryan Hanavan. Their presence and guidance in my committee were a great pleasure, and without their support, my thesis would not have reached its current form. I am also deeply thankful to all the agencies and researchers who generously shared their valuable field data with me.

I cannot forget to express my appreciation to the faculty and friends at the School of Forest Resources, who have become a wonderful family to me. I am especially grateful to my lab members for their constructive feedback throughout this journey.

Reflecting on my journey, I owe a debt of gratitude to all the teachers, fathers, and sisters of my high school, Notre Dame School in Bandipur, Nepal. Their inspiration instilled in me not only the pursuit of academic excellence but also the aspiration to be a compassionate and admirable

human being. Without Notre Dame School, I would not have had the chance to experience this vastly different world.

Lastly, my family holds a special place in my heart and memory, and it is impossible to quantify the value they have in every achievement of mine. To all the individuals who have contributed to shaping my path and standing today, my words cannot fully express my gratitude. I am forever grateful for your support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Insects and pests are one of the major historic disturbance agents in the forests of northeastern USA, particularly the forests in Maine, and neighboring Canadian provinces (Rahimzadeh et al., 2018; Bhattarai et al., 2022b). The cyclical outbreak of native forest pests such as spruce budworm (*Choristoneura fumiferana* Clem.; SBW) has been very destructive and has cost forest landowners hundreds of millions of dollars following every outbreak (Wagner et al., 2015). In addition, these episodic outbreaks also influence the composition of tree species across a large landscape. While fine spatial resolution forest composition data are not generally available at landscape scale for effective pest monitoring, constant change in the forest ecosystems, due to natural and human induced disturbances, make the existing forest type and composition maps less effective in aiding stakeholders to prepare themselves and strategize for the upcoming outbreak. The application of remote sensing technology in forestry has a long history (Curran, 1989; MacLean and MacKinnon, 1996; Holmgren and Persson 2004; Wolter and Townsend, 2008; Rahimzadeh et al., 2018). With the continuous advancement in this technology, producing products at landscape scale has become more feasible and cost effective (Wolter and Townsend, 2011; Rahimzadeh et al., 2018; Grabska et al., 2020; Rahimzadeh et al., 2020; Bhattarai et al., 2021). The application of remote sensing technology for effective forest health monitoring has even received more attention in particularly after the COVID-19 pandemic to be more integrated with other traditional pest monitoring methods such as annual Aerial Detection Survey (ADS) (Hanavan et al., 2022). Remote sensing technology can support pest monitoring and management through different ways from the detection of the exact location and extent of the damage (Fassnacht

et al., 2014; Rahimzadeh et al., 2018; Bhattarai et al., 2020; Donovan et al., 2021) to mapping the host species type and abundance (Wolter and Townsend, 2011; Bhattarai et al., 2021) and susceptibility to damage modeling (Bhattarai et al., 2022a; Bhattarai et al., 2023). In the case of SBW, the pest is host-specific and only defoliates balsam fir [*Abies balsamea* (L.) Miller; BF] and spruce (*Picea* spp.) needles. Considering the unique feeding nature of SBW, the host species composition map is considered a key layer of information to assess the susceptibility of the forests to SBW defoliation before the outbreak as well as detect the exact location of the damage. Given the complex interaction between SBW and its host species, our study investigates the diverse characteristics of the SBW host species that have potential impact on their overall susceptibility to SBW. For instance, host composition, their abundance and maturity, and the presence of defense mechanism in host species directly influence the susceptibility of the host species; however, for a forest to be defoliated, a highly susceptible forest should coincide with the elevated SBW population (please refer to Bhattarai (2020) for more information on SBW life cycle and their behavior). Regardless, it is always desirable for stakeholders to have assessments of their forest regarding their susceptibility to an anticipated future outbreak.

Evaluating the susceptibility of forests through traditional means involves labor-intensive and costly field inventories, making it impractical for landscape-level assessments. Additionally, effective monitoring of forest pest activities requires high temporal data regarding forest status, which is not achievable using conventional ground inventories. As a result, these traditional methods pose significant challenges in assessing and monitoring forest health and pest related vulnerability. On the other hand, remote sensing techniques have advanced tremendously in the past few decades due to which it is currently possible to acquire multi-scale and multi-temporal data essential for pest or forest disturbance monitoring at large scales. Moreover, remote sensing

technology offers a cost effective, rapid and technically superior alternative to the traditional forest disturbance monitoring methods (please refer to Bhattarai (2020) for details on the use of diverse remote sensing platforms for forest pest monitoring).

1.2 Research problems

Given the high potential of remote sensing techniques in forest health monitoring, there have been several recent studies to incorporate remote sensing data in forest health monitoring (Abdullah et al., 2019; Bhattarai et al., 2020) through different ways as explained earlier. However, studies incorporating physiological processes in both tree host species and the pest as well as host species characteristics for susceptibility to defoliation modeling are limited. There exists a complex relationship between the host species and the associated pests, and it changes over the time due to the coevolution between pests and hosts, which is essential to understand before formulating plans to monitor the insect/pest activities. Regarding the mechanism of the SBW defoliation outbreak, as mentioned earlier in the background section, outbreak occurs when an increased SBW population encounter the susceptible forest. Due to this reason, information on SBW population dynamics and forest susceptibility are equally important to determine the outbreak and severity of defoliation (Wagner et al., 2015; MacLean et al., 2019). However, forest owners do not have much control over the SBW dynamics, but they have the opportunity to assess the susceptibility of their forests against the forthcoming outbreak using remotely sensed susceptibility maps to manage their properties accordingly.

There have been three studies to evaluate the susceptibility of SBW host species primarily focusing on factors such as host species composition (Wolter et al., 2011; Bhattarai et al., 2021) and abundance (Wolter et al., 2008), however to the best knowledge of the author there haven't been any studies to incorporate other factors such as host maturity. In addition, the significance of

host-pest interaction, which plays a pivotal role in shaping the defensive strategies that hosts use against pests, has remained unexplored in the past studies. This dynamic interaction eventually influences the susceptibility of forests against specific pests, making it a fundamental aspect of monitoring overall forest health. This study presents and connects different parameters that could be practically used to determine SBW susceptibility namely, host composition, abundance (in terms of leaf area index; LAI, and basal area per hectare; BAPH), and their maturity using remote sensing techniques which we believe would be an efficient and cost-effective way for the stakeholders to monitor their forests. In addition, a novel approach to link foliar traits to the susceptibility of SBW host species based on their roles in SBW health and fitness (nutritive, nun-nutritive, and defense) is presented in this work. Ultimately, this study reinforces the current techniques for monitoring forest health and ecological processes in general incorporating modern remote sensing technologies and machine learning algorithms.

1.3 Research goals and objectives:

The specific objectives to meet our overall goal of developing and evaluating remote sensing-based techniques to assess the SBW host susceptibility to defoliation and damage are:

❖ Modeling SBW host species abundance based on new sensors and data availability:

- Modeling the LAI and BAPH of balsam fir and red spruce (*Picea rubens* Sarg.; RS) using the combined application of Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 optical satellite data.
- Evaluating the contribution of site variables to the estimation of LAI and BAPH.
- Improving the accuracy of LAI and BAPH modeling through incorporating deep artificial neural network (ANN) algorithm.
- Identifying key predictors for LAI and BAPH in BF and RS forests of Maine.

❖ Forest susceptibility modeling and mapping based on stand age and composition:

- Evaluate the effectiveness of PALSAR-2 L-band, and Sentinel-1 C-band SAR data individually and their integration for modeling SBW host species together with site variables.
- Employ deep ANN algorithm to improve SBW host species modeling and compare its performance with the widely tested Random Forest (RF) algorithm.
- Integrate Sentinel-2, Sentinel-1, PALSAR-2, and site variables to model and map the SBW host species over the state of Maine at a spatial resolution of 20 m.
- Estimate the maturity of host species using Land Change Monitoring, Assessment, and Projection (LCMAP) products.
- Mapping SBW susceptibility to defoliation based on stand age and composition in Maine forests.

❖ Forest susceptibility modeling and mapping based on the concentration of foliar traits in host canopies:

- Estimate and map both the nutritive (Nitrogen; N, Phosphorous; P, Potassium; K, and Copper; Cu) and non-nutritive (Iron; Fe, and Calcium; Ca) elements using Sentinel-2 bands and SVIs, along with site variables in SBW host species.
- Estimate and map mechanical defensive traits, equivalent water thickness (EWT) and leaf mass par area (LMA) of SBW host species using Sentinel-2 bands and SVIs, along with site variables.
- Compare performances of modern machine learning algorithms such as RF, support vector machine (SVM), and extreme gradient boosting (XGB) for modeling SBW host species foliar traits.

- Propose a framework for determining the susceptibility of SBW host species by assessing the relationship between host foliar trait values and SBW defoliation levels.

1.4 Thesis structure and overview of each chapter:

This thesis contains five chapters as depicted in Figure 1.1. The introductory chapter serves as the first chapter, while the final chapter comprises thoughts on integration, conclusion and future directions. Rest of the chapters consist of three independent research manuscripts. Chapter 1 provides an overall introduction and background of the research idea, past research approaches and problems, and our objectives and approaches to overcome the problems.

Chapter 2 explores the potential of Sentinel-1 C-band SAR, Sentinel-2 optical imagery and site variables to model BAPH and LAI of the major SBW host species (BF, and RS) using RF and ANN algorithms. This chapter aims to recommend the best variables combination and an algorithm to model BAPH and LAI of BF and RS.

Chapter 3 integrates multi-source satellite and site data to model and map the susceptibility of SBW host species based on the composition and maturity of the entire forested area in Maine. RF and ANN algorithms were used to evaluate the usefulness of Sentinel-2, Sentinel-1 (C-band SAR), and PALSAR-2 (L-band SAR) data along with site variables to model SBW host species. In addition, the time series of Land Change Monitoring, Assessment, and Projection (LCMAP) products were used to determine the age of the SBW host species to ultimately assess the susceptibility of SBW host stands. A susceptibility to SBW defoliation map for the entire state of Maine was produced as the final product from this chapter.

Chapter 4 attempts to link diverse SBW host species canopy traits with their susceptibility to SBW. This chapter evaluates the performance of several machine learning algorithms namely, RF, SVM, and XGB to model eight different SBW host canopy traits (nutritive, non-nutritive, and

defense) using Sentinel-2 and site data. This chapter proposes the best algorithm and variables' combination to estimate these foliar canopy traits and a framework to link them with the SBW susceptibility.

Finally, Chapter 5 presents the conclusion of the entire study. It summarizes and sheds light on how we can integrate the ideas from three different chapters (Chapters 2-4) for effective SBW management. It also discusses about the major takeaways from our work, limitations and future directions.

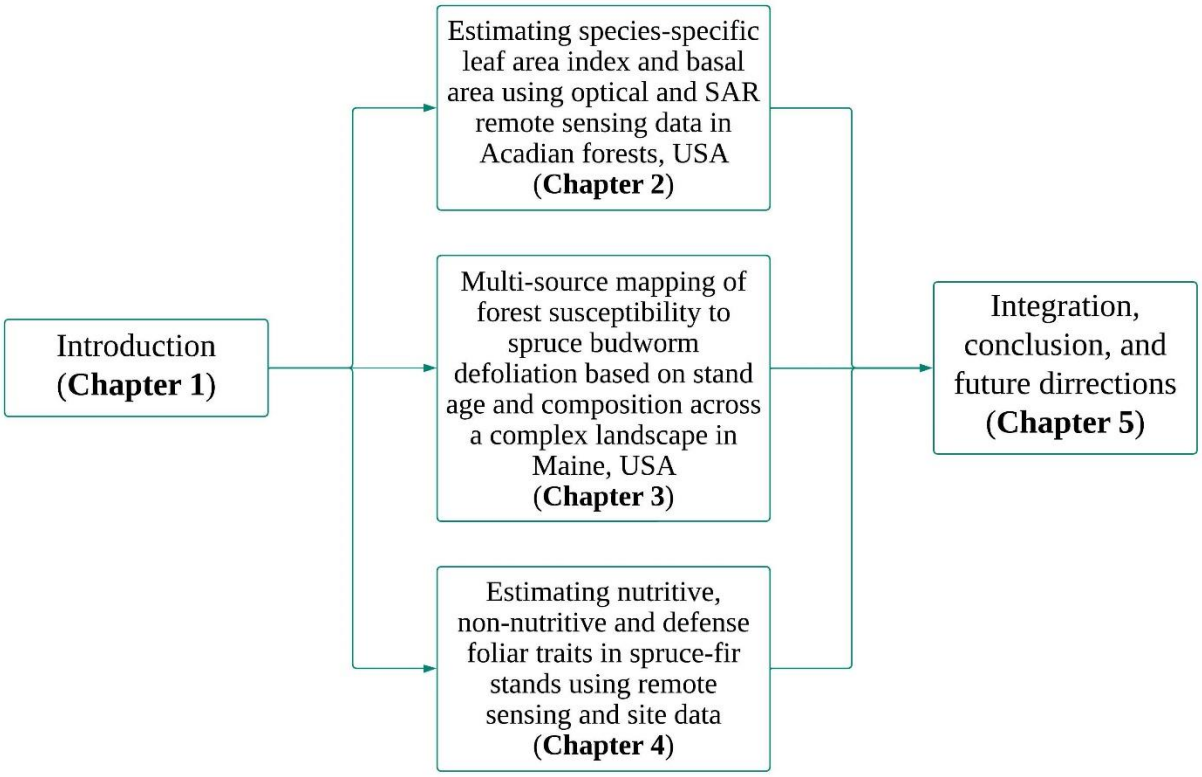

Figure 1. 1. Structure of the thesis

CHAPTER 2

ESTIMATING SPECIES-SPECIFIC LEAF AREA INDEX AND BASAL AREA USING OPTICAL AND SAR REMOTE SENSING DATA IN ACADIAN FORESTS, USA

"The contents of this chapter have been published in Bhattarai et al. (2022b)."

2.1 Introduction

Forest resources play a vital role in the long-term fulfillment of societal needs and ensuring the continued availability of ecological services; however, the forest systems around the world are under increasing threats from several stressors such as pests and diseases, climate change, and over-exploitation. Therefore, providing comprehensive information about key forest attributes and dynamics is essential for strategic planning and sustainable forest ecosystem management. Forest inventory attributes such as leaf area index (LAI) and basal area per ha (BAPH) are primarily of interest to both the research community and land managers because they are significant indicators of forest condition, structure, and function (Fassnacht et al., 1994; Mensah et al., 2020).

LAI is considered as a determinant of silvicultural characteristics like vegetation growth, stocking, and wood productivity and largely correlates with soil water storage and transpiration (Running, 1992), photosynthetic capacity (Mensah et al., 2020), the magnitude of carbon and energy fluxes (Fassnacht et al., 1994), and many more biological and ecological phenomena on the stand, regional, and global scales. BAPH as a measure of forest density, is another central attribute of a forest stand to be considered for sustainable forest management (Ahmadi et al., 2020).

LAI and BAPH measurement and mapping hold a long history and have been widely executed using traditional field measurement methods (Fang et al., 2019). Direct measurement of these parameters using destructive sampling in the field is neither feasible nor desirable on a regional or global scale (Fassnacht et al., 1994; Dube et al., 2019). With the advent of freely

available satellite data, remote sensing techniques are gaining popularity among researchers for a rapid, accurate, and efficient estimation of forest stand attributes and their temporal evolutions on a landscape level (Fang et al., 2019).

There exists a suite of non-commercial remote sensing data to derive the LAI and BAPH in forests. Single-date optical data such as Landsat (Dube et al., 2019; Neinavaz et al., 2019), Moderate Resolution Imaging Spectroradiometer (MODIS) (Sprintsin et al., 2007; Qiao et al., 2019), and Sentinel-2 (Korhonen et al., 2017; Mananze et al., 2018; Darvishzadeh et al., 2019b) have been widely used for the estimation of LAI. However, a recent trend depicts the preference of Sentinel-2 data over other satellite data sources attributed to their decent and even finer resolution (spatial, spectral, and temporal) at no acquisition cost (Korhonen et al., 2017; Darvishzadeh et al., 2019b; Meyer et al., 2019). In addition, the use of synthetic aperture radar (SAR) data has also been reported for LAI modeling (Manninen et al., 2005; Stankevich et al., 2017), which can be a complementary data source for the optical data. However, the combined use of SAR and optical data for the LAI modeling across diverse or mixed forest areas has not been reported. In contrast, BAPH modeling has been done primarily using either optical data such as Landsat (Wolter et al., 2008), and Sentinel-2 (Astola et al., 2019; Ahmadi et al., 2020) or SAR data (Townsend, 2002). To our knowledge only one study was reported that used a combination of the two types of data (Wolter and Townsend, 2011). Given that the optical and SAR data represent signals from different regions of the electromagnetic spectrum, and are complementary to each other, the incorporation of SAR data is expected to add structural information of tree canopy to the optical data that mainly reflect biochemical properties (Wolter and Townsend, 2011; Bhattarai et al., 2021).

While BAPH estimation is widely performed using statistical modeling utilizing its relationship with remote sensing variables, particularly spectral vegetation indices (SVIs) (Wolter et al., 2008; Ahmadi et al., 2020), the estimation of LAI is generally based on either of the following methods: 1) empirical methods that statistically relate LAI to canopy reflectance or SVIs; or 2) physical or radiative transfer models where LAI is retrieved through inversion of a spectral library created in forward mode (Delegido et al., 2011; Fang et al., 2019; Darvishzadeh et al., 2019b). Empirical methods are often biome- and/or time-specific and can suffer from saturation effect (Delegido et al., 2011); but this can be addressed by including appropriate site variables and optimized band combinations for creating SVIs, in particular red-edge and three-band combination SVIs (Mananze et al., 2018; Wang et al., 2018**c**). In addition, using a combination of suitable SVIs and robust machine learning algorithms, empirical models can outperform physical ones due to their statistical superiority (Cui and Kerekes, 2018; Manzane et al., 2018). Furthermore, the empirical methods are simple and straightforward for both interpretation and future development (Delegido et al., 2011; Fang et al., 2019).

Our study focused on the estimation of both LAI and BAPH for two commercially and ecologically important tree species in eastern North America, balsam fir (*Abies balsamea* (L.) Mill.; BF) and red spruce (*Picea rubens* Sarg.; RS), which are also threatened by a periodic outbreak of eastern spruce budworm (*Choristoneura fumiferana* Clem.; SBW) (Rahimzadeh-Bajgiran et al., 2018). The susceptibility of the forest stands to SBW is largely impacted by the distribution and stand characteristics of host species (Chen et al., 2021; Wolter et al., 2008). A comprehensive inventory of host species can guide us to identify areas with high host-species abundance, perform effective silvicultural operations, and management approaches on both stand and landscape levels to mitigate the susceptibility and eventually future damage from the SBW.

In general, we aimed to accurately estimate the abundance of two primary SBW host species in Maine, USA. Our specific objectives were:

- Modeling the LAI and BAPH of BF and RS using the combined application of Sentinel-1 SAR and Sentinel-2 optical satellite data.
- Evaluating the contribution of site variables to the estimation of LAI and BAPH.
- Improving the accuracy of LAI and BAPH modeling through incorporating deep learning algorithms.
- Identifying key predictors for LAI and BAPH in BF and RS forests of Maine.

2.2 Materials and Methods

The workflow presented in Figure 2.1 summarizes the entire procedure of our study. Explanations of each step in the workflow are described in later sections.

2.2.1 Study area and field data

This study was conducted in the state of Maine, USA (Figure 2.1). Maine has 89.1 percent of the total land area covered with forest, where BF and RS are the most abundant tree species in terms of number (36 percent of the entire stems) and volume (83.2 percent of the total tree volume), respectively (Butler, 2018).

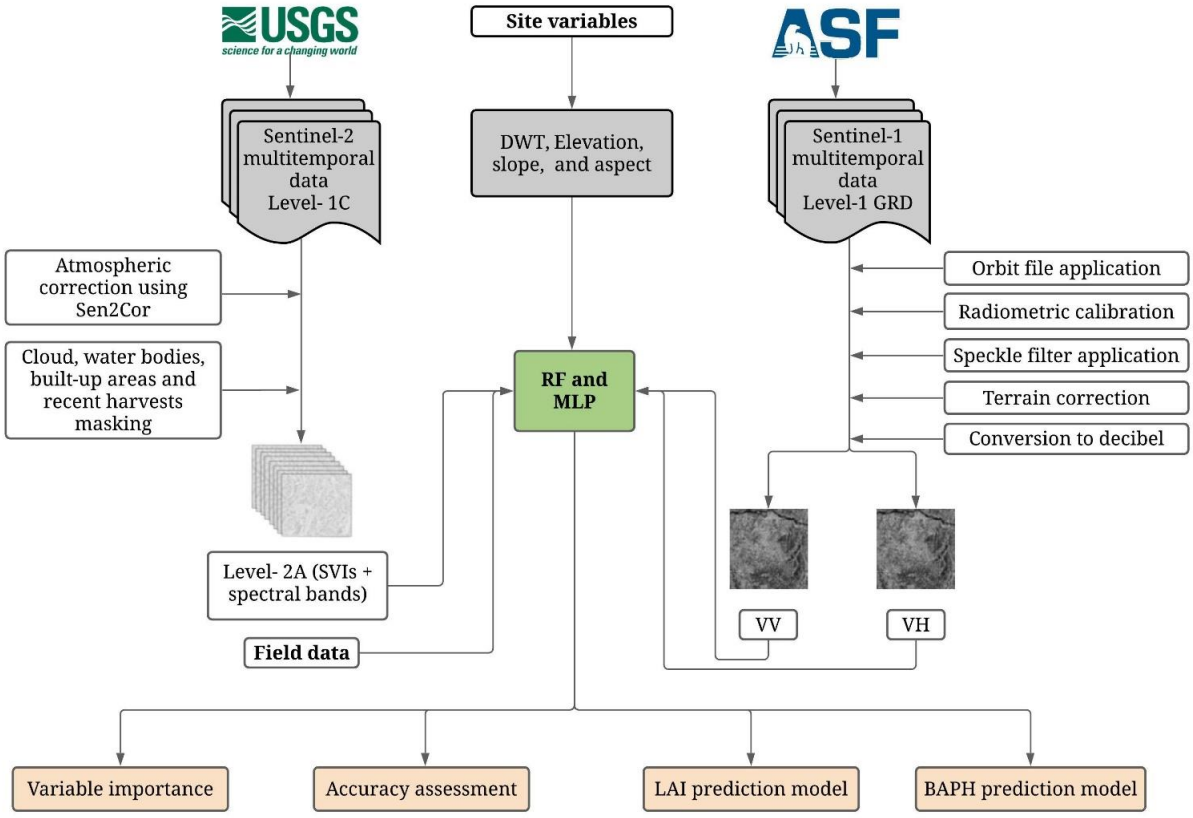

Figure 2. 1. General workflow of the methodology adopted to model leaf area index (LAI), and basal area per hectare (BAPH). VV and VH are the Vertical-Vertical, and Vertical-Horizontal polarization modes, respectively. RF: Random Forest; MLP: Multi-Layer Perceptron

The ground plot data were collected from 12 University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites (Figure 2.2) in northern and central Maine in summer 2018. We used a total of 73 measurement plots (26.6×30.5 m^2) with 40 dominant BF and 33 dominant RS plots. The tree-level measurements: diameter at breast height (DBH), crown length (CL), and tree height (HT) were measured in each plot (DeRose and Seymour, 2010). Tree leaf area (TLA) was calculated using a species- and site-specific allometric equation (Weiskittel et al., 2009):

$$TLA = b_0 (DBH^{(b_1 + b_2 CL + b_3 (HT/DBH))}) \quad (1)$$

where, b_0, b_1, b_2, and b_3 are the unique parameters of the site-specific, tree-level leaf area equation for different tree species with remaining variables already defined above. Finally, LAI was derived using the following equation:

LAI=TLA/plot area (2)

BAPH for each plot was estimated using the DBH. The minimum/maximum values for the LAI and BAPH of inventory plots were 0.22 m^2 m^{-2}/5.85 m^2 m^{-2} and 2.22 m^2 ha^{-1} /48.33 m^2 ha^{-1}, respectively.

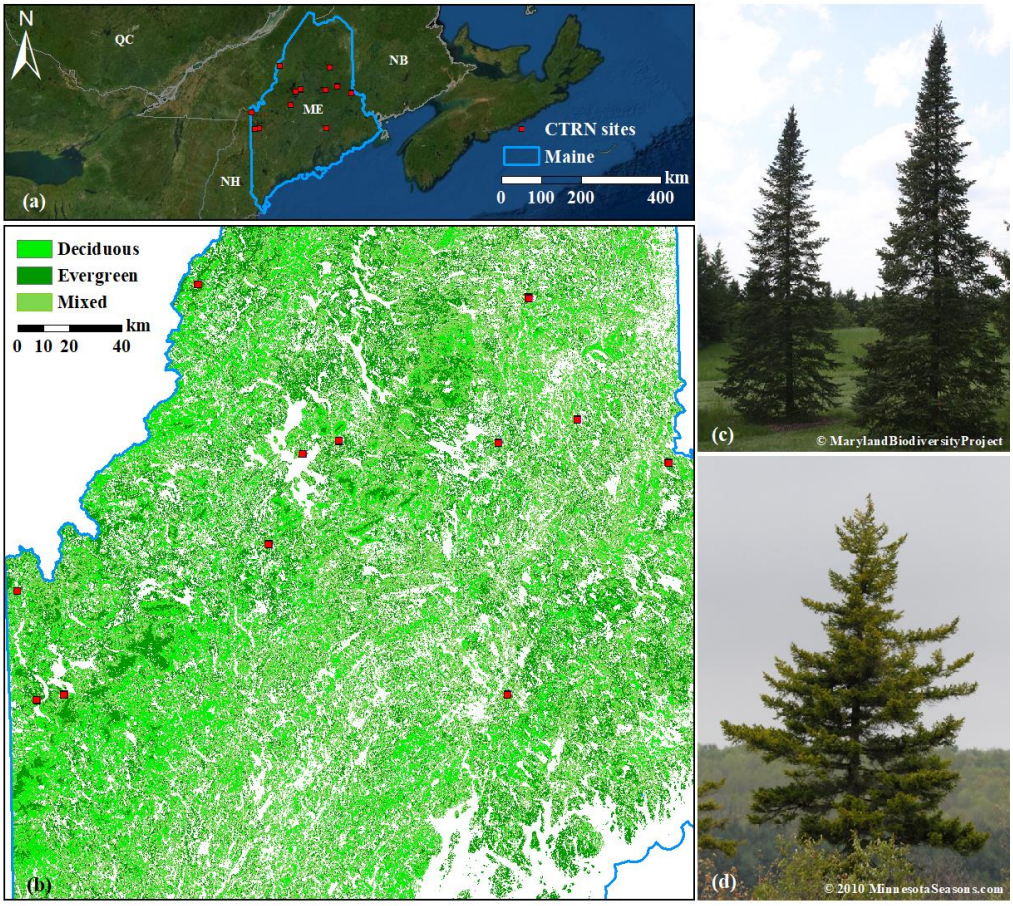

Figure 2. 2. Location of University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites (a) and forest type map retrieved from the National Land Cover Database (NLCD 2016) (b) where the evergreen forest type is dominated by balsam fir (c) and red spruce (d). QC: Quebec; NB: New Brunswick; ME: Maine; NH: New Hampshire.

2.2.2 Satellite data collection and preprocessing

We downloaded 21 multi-temporal Sentinel-2A and 2B Level 1C Top-Of-Atmosphere (L1C-TOA) reflectance images (https://earthexplorer.usgs.gov/) between 2017 and 2020 and synthetic aperture radar (SAR) Sentinel-1 images (https://asf.alaska.edu/) for 2018. Multi-temporal images for three seasons of mid-spring (May, MSp), mid-summer (June and July, MSm) and fall (September through November, Fall) were used for modeling BAPH, whereas LAI estimation was conducted using single-date imagery from the peak growing season (June-July).

We used the Sen2Cor plugin in Sentinel Application Platform (SNAP) to convert Sentinel-2 L1C-TOA products to L2A surface reflectance products and then to calculate several SVIs at 20 m spatial resolution (Korhonen et al., 2017; Dube et al., 2019; Bhattarai et al., 2021). Our variables include both spectral bands and SVIs, resulting in 118 Sentinel-2 variables from three dates for BAPH estimation, and 36 variables (single-date) for LAI estimation.

The dual-polarized (VH: Vertical-Horizontal, and VV: Vertical-Vertical) Sentinel-1 data were preprocessed in SNAP, then resampled to 20 m spatial resolution (Bhattarai et al. 2021). In addition to single-date speckle filtering, we also incorporated multi-temporal speckle filtering in the preprocessing workflow of SAR imagery, considering the promising results from previous studies (Yuan et al., 2018). The multi-temporal speckle filtering method was expected to better suppress the speckles in SAR images without compromising the spatial resolution, taking advantage of temporal information from the images. However, in our study, multi-temporal speckle filtering did not produce better results (results not included) than the single-date speckle filtering; consequently, models were built using the data from single-date speckle filtering. We ultimately derived two sets of six (three-date; VV_MSp, VH_MSp, VV_MSm, VH_MSm,

VV_Fall, VH_Fall) and two (single-date; VV_MSm, VH_MSm) variables from Sentinel-1 for BAPH and LAI modeling, respectively.

2.2.3 Modeling LAI and BAPH

Artificial neural networks (ANN) (Jain et al., 1996) were used and compared with the widely used machine learning algorithm, Random Forest (RF) (Breiman, 2001), for modeling LAI and BAPH estimation. We used a feedforward ANN known as Multi-Layer Perceptron (MLP), which is trained with a backpropagation technique (Jain et al., 1996). From the application point of view, MLP is becoming popular among researchers in various realms, including analyzing satellite data (Foody, 2004; Astola et al., 2019). On the other hand, RF is an ensemble of decision trees proven suitable in remote sensing data analysis due to its interpretation simplicity and usability with different data types (Astola et al., 2019; Bhattarai et al., 2021).

The VSURF R-package was used for Sentinel-2 variable reduction based on the recursive variables elimination technique (Bhattarai et al., 2021). In addition, the Pearson correlation coefficient (ranging between -1 and +1) between the variables returned after running the VSURF algorithm was evaluated, and highly correlated variables (correlation coefficient approaching -1 or +1) were eliminated. Eventually, the SAR and site variables (elevation, slope, aspect, and depth to water table (DWT) at 10 m spatial resolution) (White et al., 2013) were incorporated with the Sentinel-2 remote sensing variables to reinforce the model. MLP and RF models were run with the final sets of variables in R 3.5.1 using the Caret package, compared and the best model was used for the final prediction.

We selected the two final individual models with the best accuracies to estimate LAI and BAPH for RS and BF. Final models were validated based on the repeated k-fold cross-validation technique, where five-fold cross-validation was repeated five times. External validation was

carried out on the test data (30% of total data), which were never used for model training. Furthermore, the performance of the models was assessed based on the coefficient of determination (R^2), root mean squared error (RMSE), and normalized RMSE (nRMSE: normalized with range), along with their standard deviations (sd).

Table 2. 1. Sentinel-2 derived SVIs used for RS and BF LAI and BAPH modeling.

SVIs	Equation
Structure	
Atmospherically Resistant Vegetation Index (ARVI)	(B8A-2*B4+ B2)/(B8A+2* B4+ B2)
Enhanced Vegetation Index7 (EVI7)	2.5*(B7-B4)/(1+B7+6*B4-7.5*B2)
Enhanced Vegetation Index8 (EVI8)	2.5*(B8A-B4)/(1+B8A+6* B4-7.5*B2)
Modified Simple Ratio (MSR)	((B7/B4)-1)/sqrt((B7/B4)+1)
Normalized Difference Vegetation Index (NDVI)	(B8A-B4)/(B8A+B4)
Soil Adjusted Vegetation Index (SAVI)	1.5*(B8A-B4)/(B8A+B4+0.5)
Wide Dynamic Range Vegetation Index (WDRVI)	((0.01*B7)-B4) /((0.01*B7+B4)+((1-0.01)/(1+0.01)))
Physiology/stress	
Anthocyanin Reflectance Index1 (ARI1)	(1/B3)-(1/B5)
Anthocyanin Reflectance Index2 (ARI2)	(B8A/B3)-(B8A/B5)
Carotenoid Reflectance Index1 (CRI1)	(1/B2)-(1/B3)
Carotenoid Reflectance Index2 (CRI2)	(1/B2)-(1/B5)
Normalized Difference Infrared Index11 (NDII11)	(B8A-B11)/(B8A+B11)
Normalized Difference Infrared Index 12 (NDII12)	(B8A-B12)/(B8A+B12)
Plant Senescence Reflectance Index (PSRI)	(B4−B3)/B8A
Biochemistry	
Chlorophyll Green Index (GCI)	(B8A/B3) –1
Chlorophyll Red Edge (Clre)	(B7/B5)-1
Green Atmospherically Resistant Index (GARI)	(B8A-B3-(B2-B4))/(B8A+B3-(B2-B4))
Green NDVI (GNDVI)	(B8A-B3)/(B8A+B3)
Inverted Red Edge Chlorophyll Index (IRECI)	(B7-B4)*(B6/B5)
MERIS Terrestrial Chlorophyll Index (MTCI)	(B6-B5)/(B5-B4)
Modified Chlorophyll Absorption in Reflectance Index (MCARI)	1-((0.2)*(B5-B3)/(B5-B4))
Normalized Difference Vegetation Index45 (NDVI45)	(B5-B4)/(B5+B4)
Normalized Difference Vegetation Index65 (NDVI65)	(B6-B5)/(B6+B5)

Table 2.1 continued…

Red-Edge Normalized Difference Vegetation Index (NDVIRE)	(B8A-B6)/(B8A+B6)
Sentinel-2 Red Edge Position (S2REP)	705+35*((((B7+B4)/2)-B5)/(B6-B5))
Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	3∗((B5-B4)-0.2∗(B5-B3)∗(B5/B4))
Triangular Vegetation Index (TVI)	0.5∗(120∗(B6−B3)−200∗(B4−B3)

2.3 Results

2.3.1. Selection of Sentinel-1, Sentinel-2, and site variables for modeling

The final set of variables (including Sentinel-1, Sentinel-2, and site variables) as shown in Figure 2.3 depicted the dominance of Sentinel-2 variables for the RS, and BF LAI and BAPH modeling. Most of the Sentinel-2 variables selected for the prediction were red-edge variables and represented the category of SVIs expressing physiology/stress (CRI2, PSRI, CRI1) and leaf chemistry (S2REP, TCARI, IRECI, and NDVIRE). Specifically, S2REP and IRECI were prominent for modeling BAPH, while NDVIRE and IRECI were significant for LAI. In addition, the variables selected for BAPH estimation mainly consisted of mid-summer as compared to mid-spring and fall variables.

The incorporation of Sentinel-1 variables did not benefit any of the final models, although these variables demonstrated some promise on integration with Sentinel-2 variables during the intermediate formulations of LAI models (Table 2.2). However, the site variables were significant for all models. Elevation was among the best variables for modeling LAI and BAPH of both species, while DWT was important for modeling LAI for RS only. Slope and aspect were not influential for any of the models.

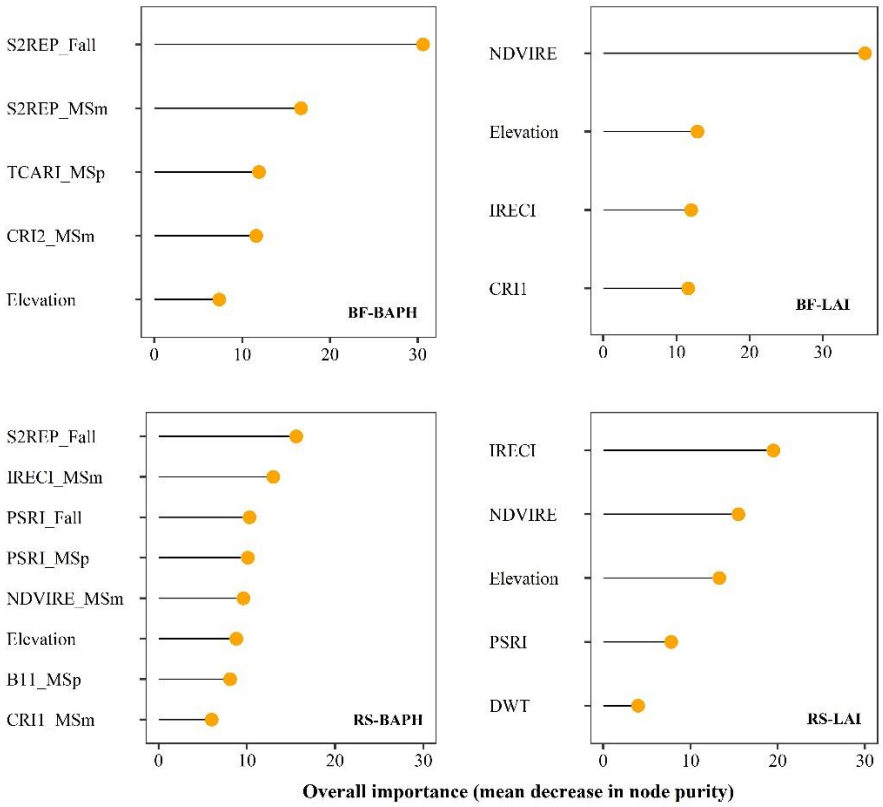

Figure 2. 3. Variable importance plots for four final models using Sentinel-1, Sentinel-2, and site variables. Variable abbreviations follow the pattern of, variable name_season of acquisition. MSp: mid-spring; MSm: mid-summer; Fall: fall (refer to Table 2.1 for the variables names and information).

2.3.2. Model performance and validation

Several combinations of Sentinel-1, Sentinel-2, and site variables were tested for LAI and BAPH modeling using the RF algorithm, and the outcomes are presented in Table 2.2. In this table, the rows in bold indicate our best model combinations based on Figure 2.3. However, to be comparable with available literature we also evaluated single best SVI and SAR variables for model building.

The RF models developed for LAI estimation attained the lowest nRMSE of 0.14 and 0.18 for RS and BF, respectively, with Sentinel-2 multi-variable combination. The superiority of the multi-variable model was observed over models based on a single SVI for LAI estimation in our study (Table 2.2). Further addition of SAR variables marginally decreased the estimation error

(0.72 m^2 m^{-2} for RS and 1.07 m^2 m^{-2} for BF) with nRMSE almost the same; however, the combination of Sentinel-2 and site variables generated better results. The combination of Sentinel-2 and site variables (Elevation and DWT) attained an overall lowest LAI estimation nRMSE of 0.12 for RS, whereas Sentinel-2 variables plus elevation yielded the lowest estimation nRMSE of 0.16 for BF.

BAPH for both BF and RS was estimated with nRMSE of 0.12, using Sentinel-2 variables alone. The addition of Sentinel-1 variables did not improve the accuracy further. Nevertheless, incorporating site variables (elevation) decreased the RMSE slightly to 5.69 m^2 ha^{-1} for BF-BAPH and 5.18 m^2 ha^{-1} for RS-BAPH with the nRMSE unchanged at 0.12. Besides elevation, none of the site variables (slope, aspect, or DWT) helped to improve the model.

Sentinel-1 variables alone performed poorer than expected with RF (Table 2.2). However, the addition of site variables, especially elevation, improved the model accuracy marginally for all models (RS-LAI, BF-LAI, and RS-BAPH) except BF-BAPH, where the model error increased after the addition of elevation. After incorporating elevation, the lowest nRMSEs achieved with Sentinel-1 variables were 0.27, 0.32, 0.28, and 0.30 for RS-LAI, BF-LAI, RS-BAPH, and BF-BAPH models, respectively.

Under further evaluation of the performance of final variables obtained after variable reduction (Figure 2.3) using MLP regression, none of the formulated models could yield better accuracy than their RF regression counterparts (Figure 2.4). A single hidden layer returned the best results for all four MLP models. RS-LAI and BF-LAI attained the best result with six and five neurons, respectively, while RS-BAPH and BF-BAPH required seven individual neurons in the hidden layer. The nRMSEs with the MLP algorithm for RS-LAI, BF-LAI, RS-BAPH, and BF-BAPH models were 0.13 (0.04), 0.18 (0.06), 0.14 (0.03), and 0.18 (0.04), respectively.

The four best RF models presented in bold in Table 2.2 were selected to predict the independent test dataset. The model performance on the testing dataset (Figure 2.5) confirms the stability of the models we created using our training data. We obtained a robust nRMSE as low as 0.13 and 0.15 for BF-BAPH, and RS-BAPH whereas these values were 0.18 and 0.17 for BF-LAI, and RS-LAI, respectively. Figure 2.5 demonstrates the relationship between the predicted and actual LAI and BAPH values, which highlights our models' capacity on a relatively independent dataset.

Table 2. 2. Performance of the models formulated to estimate LAI and BAPH using RF model. DWT: depth to the water table. SAR includes six variables (VV_MSp, VH_MSp, VV_MSm, VH_MSm, VV_Fall, VH_Fall) for BAPH while two variables (VV_MSm, VH_MSm) for LAI modeling.

Models	**Variables used**	**R^2 (sd)**	**RMSE (sd)**	**nRMSE (sd)**
RS-LAI	IRECI	0.45 (0.15)	0.95 (0.15)	0.18 (0.03)
	NDVIRE	0.45 (0.21)	1.00 (0.28)	0.19 (0.05)
	PSRI	0.25 (0.16)	1.10 (0.20)	0.21 (0.04)
	IRECI+Elevation	0.68 (0.18)	0.75 (0.21)	0.15 (0.04)
	IRECI+NDVIRE+PSRI	0.67 (0.17)	0.74 (0.17)	0.14 (0.03)
	IRECI+NDVIRE+PSRI+Elevation	0.74 (0.16)	0.65 (0.16)	0.13 (0.03)
	IRECI+NDVIRE+PSRI+Elevation+DWT	**0.75 (0.16)**	**0.63 (0.19)**	**0.12 (0.04)**
	IRECI+NDVIRE+PSRI+SAR	0.69 (0.17)	0.72 (0.20)	0.14 (0.04)
	SAR	0.11 (0.13)	1.00 (0.20)	0.29 (0.06)
	SAR + Elevation	0.10 (0.13)	0.93 (0.17)	0.27 (0.05)
BF-LAI	NDVIRE	0.59 (0.15)	1.31 (0.23)	0.22 (0.04)
	IRECI	0.36 (0.15)	1.66 (0.15)	0.28 (0.04)
	CRI1	0.26 (0.18)	1.84 (0.36)	0.31 (0.06)
	NDVIRE+Elevation	0.71 (0.13)	1.06 (0.23)	0.18 (0.04)
	NDVIRE+IRECI+CRI1	0.71 (0.13)	1.09 (0.21)	0.19 (0.04)
	NDVIRE+IRECI+CRI1+Elevation	**0.78 (0.10)**	**0.95 (0.19)**	**0.16 (0.03)**
	NDVIRE+IRECI+CRI1+Elevation+DWT	0.77 (0.09)	0.97 (0.19)	0.17 (0.03)
	NDVIRE+IRECI+CRI1+SAR	0.72 (0.11)	1.07 (0.20)	0.18 (0.03)
	SAR	0.13 (0.13)	1.08 (0.19)	0.32 (0.06)
	SAR+Elevation	0.12 (0.14)	1.05 (0.17)	0.32 (0.05)
RS-BAPH	S2REP_Fall+IRECI_MSm+PSRI_Fall +PSRI_MSp+NDVIRE_MSm+B11_MSp +CRI1_MSm	0.79 (0.12)	5.27 (1.29)	0.12 (0.03)
	S2REP_Fall+IRECI_MSm+PSRI_Fall +PSRI_MSp+NDVIRE_MSm+B11_MSp +CRI1_MSm+Elevation	**0.80 (0.11)**	**5.18 (1.23)**	**0.12 (0.03)**

Table 2.2 continued…

	S2REP_Fall+IRECI_MSm+PSRI_Fall +PSRI_MSp+NDVIRE_MSm+B11_MSp +CRI1_MSm+Elevation+DWT	0.80 (0.12)	5.20 (1.32)	0.12 (0.03)
	S2REP_Fall+IRECI_MSm+PSRI_Fall +PSRI_MSp+NDVIRE_MSm+B11_MSp +CRI1_MSm+SAR	0.79 (0.10)	5.47 (1.33)	0.13 (0.03)
	SAR	0.11 (0.14)	8.77 (1.87)	0.29 (0.06)
	SAR+Elevation	0.14 (0.15)	8.55 (1.84)	0.28 (0.06)
BF-BAPH	S2REP_Fall+S2REP_MSm+TCARI_MSp +CRI2_MSm	0.87 (0.06)	5.71 (1.45)	0.12 (0.03)
	S2REP_Fall+S2REP_MSm+TCARI_MSp +CRI2_MSm+Elevation	**0.87 (0.06)**	**5.69 (1.44)**	**0.12 (0.03)**
	S2REP_Fall+S2REP_MSm+TCARI_MSp +CRI2_MSm+Elevation+DWT	0.88 (0.06)	5.85 (1.27)	0.13 (0.03)
	S2REP_Fall+S2REP_MSm+TCARI_MSp +CRI2_MSm+SAR	0.84 (0.08)	6.40 (1.49)	0.14 (0.03)
	SAR	0.17 (0.15)	7.99 (1.11)	0.30 (0.04)
	SAR+Elevation	0.15 (0.14)	8.11 (1.17)	0.30 (0.04)

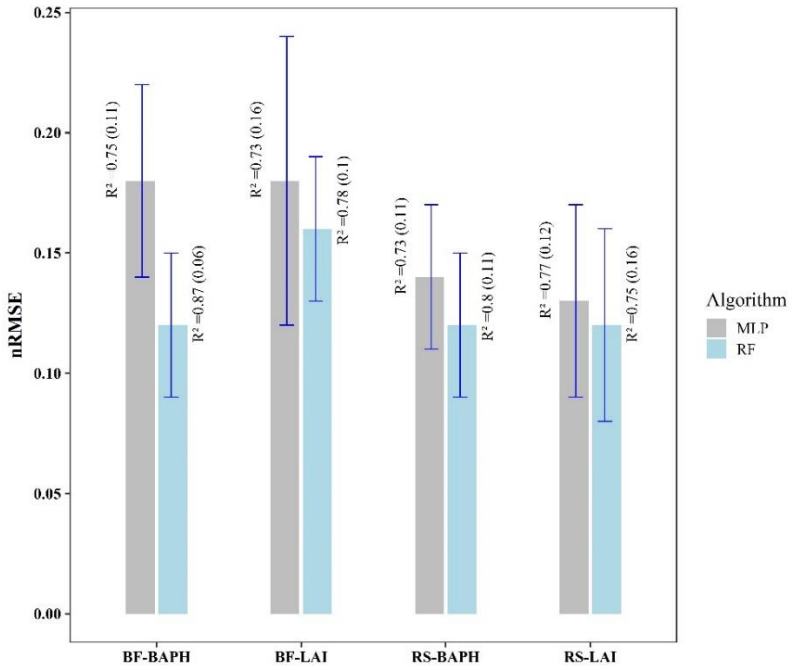

Figure 2. 4. LAI, and BAPH model accuracy for the RF and MLP models using the best performing variables obtained after the variable reduction procedure (Table 2.2 and Figure 2.3). Accuracy metrics are presented with standard deviation (sd) inside the parenthesis. Error bars represent one sd.

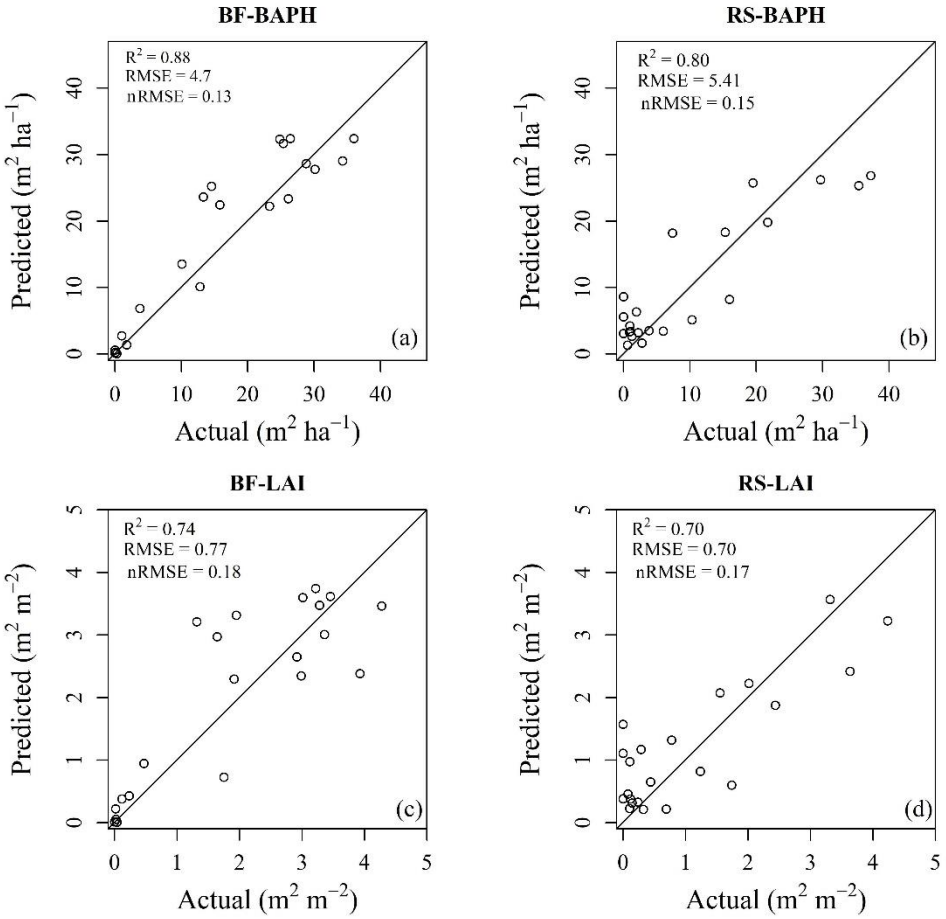

Figure 2. 5. Actual vs. predicted values of LAI and BAPH for four best RF regression models (Table 2.2). The diagonal line running across each subplot is a 1:1 relationship between the actual and predicted values.

2.4 Discussion

Our study attempted to construct models to estimate the species-specific LAI and BAPH of two economically important tree species in Maine, namely, red spruce (RS) and balsam fir (BF. Previous studies on the estimation of LAI (Korhonen et al., 2017; Neinavaz et al., 2019) and BAPH (Ahmadi et al., 2020) of forest cover were not species-specific. These species-specific models to estimate LAI and BAPH are essential to determine the susceptibility of forests to host-specific defoliators like SBW and to evaluate the basal area of the hosts under risk, respectively. In addition, we evaluated the combined use of SAR and optical data for LAI and BAPH modeling and included several site variables in estimating both LAI and BAPH in our study.

We used combined Sentinel-2 multispectral, Sentinel-1 SAR, and site data with widely used machine learning algorithms (RF and MLP). Our results demonstrate that incorporating Sentinel-1 SAR variables did not improve models based on Sentinel-2 data using either single-date or a multi-temporal speckle filtering procedure. Our results in terms of the performance of Sentinel-1 SAR data for LAI estimation are in contrast to the findings from Manninen et al. (2005). They reported the superiority of radar (Environmental Satellite (ENVISAT) Advanced Synthetic Aperture Radar (ASAR) C-band) data over optical (SPOT high-resolution visible and infrared (HRVIR1) imagery for the estimation of LAI in the boreal forest of Finland (estimation error of 0.28 $m^2 m^{-2}$ vs. 0.39 $m^2 m^{-2}$ respectively). Stankevich et al. (2017) also demonstrated a promising relationship (R^2=0.81) between the relative difference polarization index (RDPI) (derived from Sentinel-1 bands) and the LAI of temperate deciduous and mixed forest in Ukraine, which could not be observed in our study using VV, and VH polarized bands. The inferior result from SAR data in our study compared to other studies could be attributed to the differences in variables used (RDPI vs raw bands) and sensors (SPOT-HRVIR1 vs Sentinel-2) as well as the type of forest under consideration. More importantly, our models are species specific which might not be directly comparable with the above studies.

The modeling of BAPH also demonstrated the superiority of optical variables (Sentinel-2) over the SAR variables. A similar observation for tree species BAPH modeling in Minnesota was reported by Wolter and Townsend (2011), where the performances of Radarsat (C-band SAR) and PALSAR (L-band SAR) were inferior to that of Landsat-5 data. In agreement with our findings, Shamsoddini (2012) reported the superior performance of SPOT-5 optical data for BAPH estimation as compared to the Advanced Land Observation Satellite (ALOS) radar data for radiata pine (*Pinus radiata* L.) plantation in Australia.

We formulated Sentinel-2 multi-variable SVI, and individual spectral band models for both LAI and BAPH estimation to include the combined effect of different variables responsible for the determination of unique canopy traits. The aforementioned approach (multi-variable modeling) has already been practiced in several studies from the past for both LAI (Korhonen et al., 2017; Dube et al., 2019; Meyer et al., 2019) and BAPH (Wolter and Townsend, 2011; Astola et al., 2019; Ahmadi et al., 2020) estimation with a great promise which can be observed in our results as well (Table 2.2). The nRMSE was improved by 0.06 for RS-LAI and BF-LAI, using multi-variables in place of the single best predicting variable. Single variables were not tested for BAPH models since we used multi-temporal data for modeling.

The dominance of red-edge variables for modeling LAI (NDVIRE, IRECI, PSRI, and CRI1) and BAPH (S2REP, NDVIRE, IRECI, TCARI, PSRI, CRI1, and CRI2) in our study are in agreement with the findings from the previous investigations (Korhonen et al., 2017; Wang et al., 2018c; Darvishzadeh et al., 2019b). Majasalmi and Rautiainen (2016) found a moderate relationship between S2REP and/or IRECI and LAI, which does not align fully with our results (S2REP not selected). Nevertheless, NDVIRE (highly correlated with S2REP) was one of the most influential variables for the LAI models we developed. While the majority of the available studies on LAI estimation are focused on the broader spectrum of forest types (Majasalmi and Rautiainen, 2016; Korhonen et al., 2017; Meyer et al., 2019), our study proposes an algorithm that can estimate the LAI of specific tree species.

For BAPH estimation, the major role of red-edge variables in our study could be supported with several other studies with similar findings. Astola et al. (2019) compared the potential of Landsat-8 and Sentinel-2 variables for predicting several forest variables, including the BAPH in the boreal forest in southern Finland, where they concluded the advantage of red-edge bands in

Sentinel-2, which resulted in better performance compared to Landsat-8. Furthermore, the study on mapping BAPH and other stand variables in northern Iran conducted by Ahmadi et al. (2020) also emphasizes the prime role of red-edge bands in estimating species-specific BAPH.

Site variables played a significant role in estimating both LAI and BAPH in our study. There have been very few investigations regarding the contribution of site variables to forest biophysical parameters estimation; however, their incorporation with remote sensing variables has produced improved results for similar studies (Bhattarai et al., 2020; Rahimzadeh-Bajgiran et al., 2020; Bhattarai et al., 2021). A recent study from Ahmadi et al. (2020) on estimating forest BAPH integrating site variables with Sentinel-2 data highlights the importance of elevation, which aligns with our findings. We found elevation to be one of the best variables for the estimation of LAI and relatively noteworthy in BAPH modeling (Figure 2.3 and Table 2.2). In general, the inclusion of site variables in the final model (Sentinel-2 variables) improved the RMSE by 0.11-0.14 m^2 m^{-2} (nRMSE rose by 0.02-0.03) for LAI models, whereas by 0.02-0.09 m^2 ha^{-1} (nRMSE unchanged) for BAPH models.

Comparing the performances of RF and MLP algorithms using the final sets of reduced variables (Figure 2.3), RF model outperformed MLP in terms of the estimation error (Figure 2.4); however, the RS-LAI model performed slightly better with MLP algorithm in terms of R^2. The results suggest that RF regression can model tree inventories (LAI and BAPH) comparatively better than ANN. Similar to our findings, several other researches have suggested using other algorithms like Gaussian process regression (Verrelst et al., 2015) and RF (Wei et al., 2017; Wang et al., 2018a) over the ANN algorithm for the estimation of biophysical properties of vegetation. The better performance of RF in our study can be attributed to the robustness of the RF algorithm

in handling high variability in the field data, which was evident in previous findings as well (Wang et al., 2018a; Dube et al., 2019).

Both LAI and BAPH models were successfully cross-validated. Our R^2 values of 0.80-0.87 for BAPH models (Table 2.2) are relatively higher than some previous studies and our RMSE values of 5.18-5.69 m^2 ha^{-1} are lower. For example, Townsend (2002) used a combination of radar images (ERS-1, JERS-1, and Radarsat) and Landsat imagery to achieve an overall validation R^2 of 0.59-0.87 and RMSE of 6.71-13.40 m^2 ha^{-1} for modeling BAPH of a mixed forest in North Carolina. Further, Wolter and Townsend (2011) used multi-sensor optical and SAR data to estimate the abundance of various coniferous and broadleaved species where the BAPH model for balsam fir attained an R^2 value of 0.78, which is comparatively lower than the R^2 value we achieved (0.87); nonetheless, the RMSE value they reported (2.26 m^2 ha^{-1}; nRMSE unknown) is better than that of ours. Ahmadi et al. (2020) also achieved the highest R^2 of only 0.48 and RMSE of 8.12 m^2 ha^{-1} (inferior performance as compared to our model) for BAPH modeling using Bayesian additive regression trees technique.

The LAI models we developed attained a cross-validated R^2 of 0.75-0.78 and RMSE of 0.63- 0.95 m^2 m^{-2}. Our accuracy metrics were comparable and even better at times than other similar models reported in the literature (Korhonean et al., 2017; Dube et al., 2019; Meyer et al., 2019; Neinavaz et al., 2019; Qiao et al., 2019). Meyer et al. (2019) attained the best R^2 of 0.59 and RMSE of 0.88 m^2 m^{-2} for the LAI estimation of a temperate forest in Germany that is lower in terms of R^2 while close to ours in terms of the RMSE. Furthermore, Korhonean et al. (2017) used Sentinel-2 and Landsat-8 to compare their potential to estimate LAI in the boreal forest of Finland with the highest cross-validated R^2 of 0.73, RMSE of 0.59, and NRMSE of 0.19, which is slightly better than our models in terms of RMSE but not in terms of R^2 (0.75-0.78) or nRMSE (0.12-0.16).

The empirical method we used for LAI and BAPH estimation using RF is promising for future reference; however, the comparisons we made between different studies acknowledges that they have used unique data collected from diverse sites and sensors. Given that there are limited studies on species-specific inventory parameter estimation using remote sensing techniques, our models can be useful for the estimation of BF and RS LAI and BAPH in Maine and similar geographical regions. Nevertheless, we highly recommend that future studies evaluate the strength of other machine learning algorithms as well as SAR L-band and P-band data for LAI and BAPH modeling.

2.5 Conclusions

Sentinel-1 and -2 data are invaluable resources for remote sensing applications in forest resources, whose potential has been studied in this analysis for BAPH and LAI estimation for two economically and ecologically conifers, RS and BF, in Maine, USA. Specifically, the diverse SVIs derived from Sentinel-2 imagery represent a multitude of canopy properties. Sentinel-1 SAR imagery was not found to be helpful in our study and needs additional evaluation in future studies. In particular, site variables were found promising in prediction models and should be considered more in future work.

Our study further demonstrated the superiority of RF over the MLP algorithm for estimating canopy biophysical parameters, implying the adequacy of the commonly used machine learning algorithms over deep learning algorithms for this particular application. However, the use of deep learning algorithms should still be considered for other datasets and variable combinations. The accuracy produced by our statistical models is encouraging. The models developed could be used by stakeholders to predict BF and RS abundance for future forest management activities as they are simple to build and cost-effective. In light of SBW outbreak, evaluation of host abundance

in terms of LAI and BAPH is a simplified concept that can assist the forest managers in adopting a desirable silvicultural practice to increase the resistance of the broader forest ecosystem against possible outbreaks in the future.

CHAPTER 3

MULTI-SOURCE MAPPING OF FOREST SUSCEPTIBILITY TO SPRUCE BUDWORM DEFOLIATION BASED ON STAND AGE AND COMPOSITION ACROSS A COMPLEX LANDSCAPE IN MAINE, USA

"The contents of this chapter have been published in Bhattarai et al. (2022a)."

3.1 Introduction

Species composition mapping with their detailed spatial information is crucial for stand as well as landscape-level forest management. Species composition over a region serves as an indicator of habitat quality and biodiversity (Riedler et al. 2015), forest health (Terhonen et al. 2019), and pest susceptibility (Riihimaki et al. 2004). Therefore, up-to-date information on the constituent forest species of a landscape is beneficial for the forest managers in framing robust management or conservation plans to attain a healthy and sustainably growing forest.

In eastern North America, spruce (*Picea* spp)-balsam fir (*Abis balsamea* (L.) Mill.) forests have important ecological and economical values such as production of industrial timber and pulpwood, serving as habitat to wildlife, sustaining water resources and providing recreational services to the society; however, their growth and survival are severely affected by cyclic outbreak (recurrence period of about 30-40 years) of the eastern spruce budworm (*Choristoneura fumiferana* Clem.; SBW) in the region. SBW is one of the historic forest disturbance agents in the Northeastern United States and Canada, which has been a significant threat to these forests (MacLean et al. 2019; Rahimzadeh-Bajgiran et al. 2018; Rose and Lindquist 1994). The recurrent outbreak pattern of SBW and its potential to invade a large area within a few years renders challenges to its control. The current SBW outbreak in the region started in 2006 from Quebec, Canada, and defoliated more than 13 million hectares of forests in Quebec by 2020 and is spreading

towards the northern forests of Maine, USA, and other bordering territories (MacLean et al. 2019; Ministère des Forêts, de la Faune et des Parcs 2020). In this context, the state of Maine (neighboring Quebec) recorded a loss of hundreds of millions of dollars during the last SBW outbreak in the 1970s-1980s (Wagner et al. 2015) when both forest composition and SBW defoliation data were coarse and incomplete for effective outbreak management. This warrants preparedness to respond to the upcoming outbreak effectively, informed by spatially comprehensive products that incorporate forest composition-based susceptibility maps produced through modern remote sensing technology.

The host-specific SBW primarily feeds on the needles of balsam fir, red spruce (*Picea rubens* Sarg.), and white spruce (*Picea glauca*), but damages to other conifer species can also occur when these are growing in a mixture with the favored tree species (Rose and Lindquist 1994). In general, the available SBW host species in the Northeastern USA and Canada (from high to low SBW susceptibility; the tendency of being attacked) are balsam fir, white spruce, red spruce, and black spruce (*Picea mariana* (Mill.) (Bhattarai et al. 2020; Hennigar et al. 2008; Wolter et al. 2008), which are also the most important commercial tree species in the region. There exists a complex interaction between the population of SBW and the availability of host species for a spruce-fir forest to be infected. The forest experiences defoliation when an elevated SBW population coincides with the host species; however, the population of SBW is controlled by the speed of outbreak spread as well as the application of chemical treatments to suppress them (MacLean et al. 2019).

Nevertheless, the host availability in the forest plays a major role in severing the defoliation. In addition, the mortality and growth reduction after the defoliation of host species depends upon their maturity, where mature hosts are more impacted than young ones (Hennigar et

al. 2011). Based on the premise of the host-maturity-dependent nature of SBW defoliation severity, a precise and periodically updated host species composition product incorporating maturity (stand impact types map) is of great significance for regular assessment of the forests, modifying management approaches, and preparing the landowners for the future outbreaks in general (Bhattarai et al. 2021; Wagner et al. 2015).

Forest-based species composition assessment holds a long history; however, the methods involved in the past such as in-situ field and aerial surveys, were associated with high resource consumption. With the development of satellite-based remote sensing data acquisition technologies over the past few decades, a diverse range of spatial, spectral, and temporal information has become available to researchers for mapping tree species on large scale and high resolution at minimum cost. From the suite of optical remote sensing data openly available, Landsat (Attarchi and Gloaguen 2014; Soleimannejad et al. 2019; Thompson et al. 2015), and Sentinel-2 (Bhattarai et al. 2021; Grabska et al. 2019; Grabska et al. 2020; Immitzer et al. 2019; Persson et al. 2018; Wessel et al. 2018) data are widely used for species composition mapping; yet, Sentinel-2 data is more desirable attributed to their suitable spatial, spectral, and temporal resolutions (Bhattarai et al. 2021; Grabska et al. 2019). Very high resolution commercial optical imagery such as those from Worldview satellites, as well as hyperspectral imagery are also available and perhaps better suited for tree-level identification but they come with higher cost and are only locally accessible or only collected on-demand with lower coverage (Furniss et al. 2021; Immitzer et al. 2018). Therefore, data from sensors such as Sentinel-2, which provide fine spectral and spatial resolution with reasonable temporal resolution (every 5 days), are ideal for forest composition mapping.

Active remote sensing data like light detection and ranging (LiDAR) are also used either alone (Holmgren and Persson 2004) or in combination with other optical data (Hartling et al. 2019) to map tree species. Nonetheless, their high cost of acquisition and local availability limit their applicability. Sentinel-1 C-band Synthetic aperture radar (SAR) provides global coverage at no cost to the end users with high temporal resolution of six days, and it has become more popular for vegetation classification (Bhattarai et al. 2021; Dostalova et al. 2021; Erinjery et al. 2018). Longer wavelength SAR data like L-band SAR are generally commercially available and have been recommended (Bhattarai et al. 2021; Erinjery et al. 2018) and used (Attarchi and Gloaguen 2014; Liesenberg and Gloaguen 2013; Wolter and Townsend 2011) largely for forest composition mapping due to their superiority over C-band SAR data. Furthermore, in recent years, the integrated use of cross-sensor data such as radar-radar (Lardeux et al. 2011; Wolter and Townsend 2011), or optical-radar (Attarchi and Gloaguen 2014; Bhattarai et al. 2021; Wolter and Townsend 2011; Yu et al. 2018) has garnered interest among the researchers in the field of vegetation mapping. In particular, the integration of optical and radar data can characterize the vegetation from different perspectives as they are believed to be complementary to each other (Bhattarai et al. 2021; Jones and Vaughan 2010).

Pest host species composition mapping on a landscape level is a relatively newer approach, and very few investigations have been reported in this area so far (Bhattarai et al. 2021; Wolter et al. 2008). A recent study from Bhattarai et al. (2021) modeled SBW tree host species in New Brunswick, Canada, using freely available remote sensing (Sentinel-1, Sentinel-2), and site data. However, other than the availability of the host trees, for pests like SBW, the maturity of the forest is also a key factor (Hennigar et al. 2011) to determine overall vulnerability. In this work, we incorporated additional and improved remote sensing and site variables and compared the accuracy

of two widely used classification algorithms to classify SBW host species into eight classes. We also produced a SBW susceptibility to defoliation product based on stand impact types (host species composition product incorporating maturity) in Maine, USA. Additionally, this study also devises a technique to estimate the forest stand maturity utilizing the archive of annual land cover and land cover change data. In short, this work leverages and improves upon our prior work (Bhattarai et al. 2021) by the incorporation of additional remote sensing products, accounting for both stand composition as well as age, and extending the number of impact types mapped, while evaluating an alternative nonparametric modeling approach.

Our specific objectives were to: 1) evaluate the effectiveness of PALSAR-2 L-band, and Sentinel-1 C-band SAR data individually and their integration for modeling SBW host species together with site variables, 2) employ deep artificial neural network (ANN) algorithm to improve SBW host species modeling and compare its performance with the widely tested Random Forest (RF) algorithm, 3) integrate Sentinel-2, Sentinel-1, PALSAR-2, and site variables to model and map the SBW host species over the state of Maine at a spatial resolution of 20 m, 4) estimate the maturity of host species using Land Change Monitoring, Assessment, and Projection (LCMAP) products, and 5) mapping SBW susceptibility to defoliation based on stand impact types in Maine forests.

3.2 Materials and Methods

The overall schematic diagram of our entire study and the methods involved is presented in Figure 3.1. Elaboration of each step is provided in the subsequent sections.

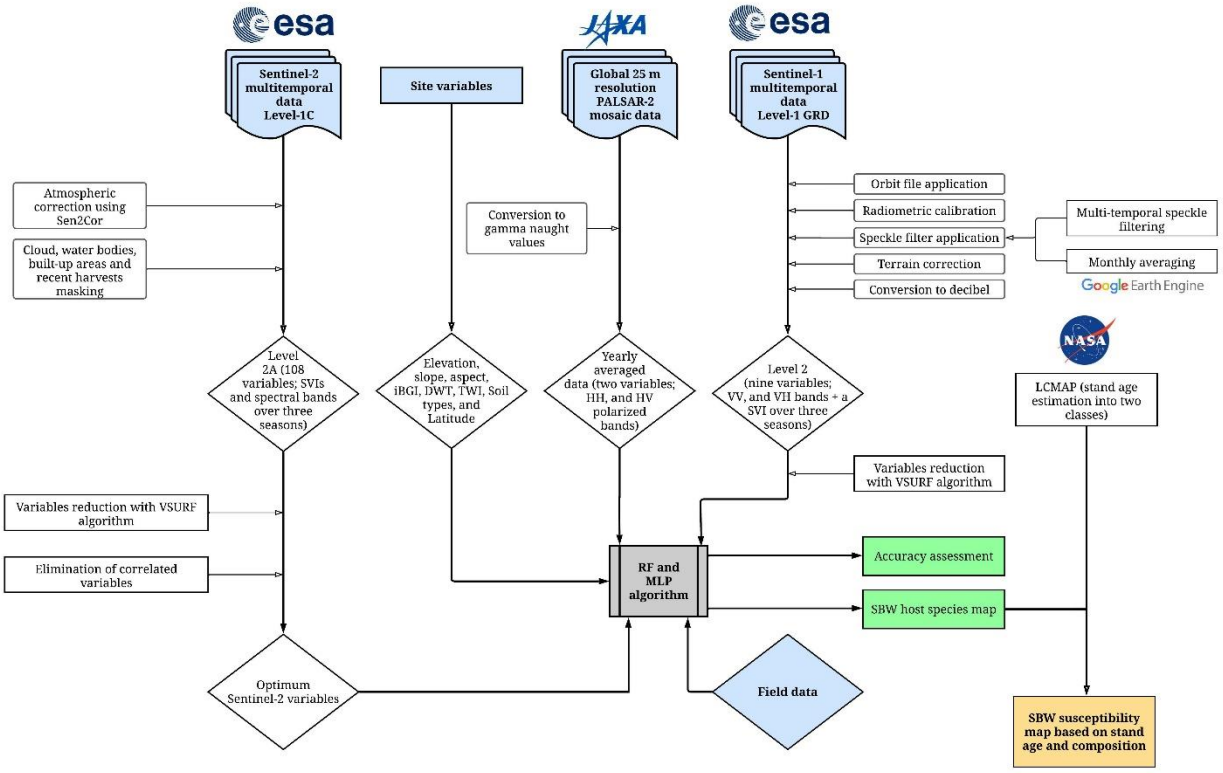

Figure 3. 1. Methodological flowchart of the study. The whole study was broadly segmented into four parts: (1) field and multi-sensor remote sensing data acquisition and preprocessing, (2) data modeling using random forest (RF), and multi-layer perceptron (MLP) algorithms and model evaluation, (3) host species map creation using the best model, and (4) age adjusted SBW host species and susceptibility mapping using Land Change Monitoring, Assessment, and Projection (LCMAP) products. HH (Horizontal-Horizontal), and HV (Horizontal-Vertical) are the polarization modes available for PALSAR-2 images, while VV, and VH are the polarization modes for Sentinel-1 images. DWT: Depth to Water Table; TWI: Topographic Wetness Index; iBGI: improved Biomass Growth Index; SVIs: spectral vegetation indices.

3.2.1 Study area

The study area comprises most of the state of Maine, USA where SBW susceptible forests exist and satellite and field data were available (Figure 3.2). The extrapolation of study area outside the range of training data was based on the assumption that the training data used in the study are representative of the forest types prevalent throughout the study area. Maine is a maritime state with nine diverse climatic zones running from northern uplands to the coastal south of the state (Briggs and Lenin 1992). Forests in Maine cover approximately 89.1 percent of the total land area,

with the dominance of balsam fir and red spruce in terms of stem number and volume, respectively (Butler 2018). The major tree species found in Maine other than balsam fir and red spruce are red maple (*Acer rubrum* L.), yellow birch (*Populus alleghanensis* Britt.), paper birch (*Populus papyrifera* Marshall), American beech (*Fagus grandifolia*), quaking aspen (*Populus tremuloides* Michex.), white spruce, black spruce, northern white cedar (*Thuja occidentalis* L.), eastern hemlock (*Tsuga canadensis* (L.) Carr.), and eastern white pine (*Pinus strobus* L.) (Butler 2018). Almost all the forest in Maine, except a small portion in the southwest, belongs to the Acadian forest, a transitional forest between northern coniferous boreal and southern broadleaved forest (Rowe 1972).

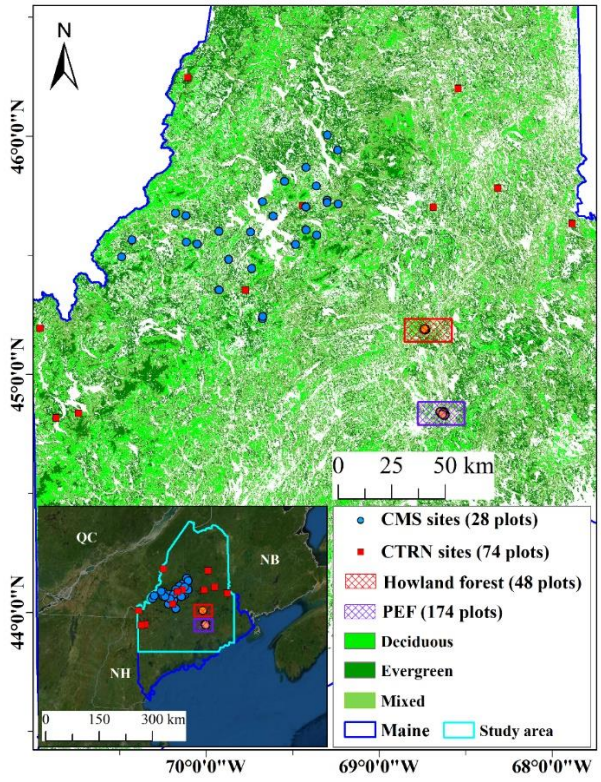

Figure 3. 2. Study area and the location of sample plots. Training data were obtained from multiple sites, namely, University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites, Carbon Monitoring System (CMS) sites, Howland forest, and Penobscot Experimental Forest (PEF). The forest type map presented was acquired from the National Land Cover Database (NLCD 2016). QC: Quebec, Canada; NB: New Brunswick, Canada; NH: New Hampshire, USA; ME: Maine, USA.

3.2.2 Field data collection for model training and validation

Ground reference data were obtained from four different field campaigns in Maine completed between 2015 and 2018 from a total of 324 sample plots. The four inventory data sources were: 1) University of Maine Cooperative Forestry Research Unit's Commercial Thinning Research Network (CTRN) sites (74 plots; Kuehne et al., 2020), 2) Carbon Monitoring System (CMS; Deo et al., 2017) sites (28 plots), 3) Howland Research Forest database (48 plots; Teets et al., 2018), and 4) Penobscot Experimental Forest database (174 plots; Puhlick et al., 2020). Plot sizes were approximately 800 m^2 in all sites. Plot-wise basal area was calculated for all the constituent tree species, and a threshold of 70% combined basal area was applied to assign each plot to one of the eight SBW host species classes developed for our study: balsam fir (BF; 47 plots), broadleaved (BL; 17 plots), BF and BL (BF/BL; 32 plots), conifers other than BF and spruce (Con; 32 plots), Con and BF (Con/BF; 66 plots), Con and BL (Con/BL; 42 plots), spruce (SP; 39 plots), and Con and SP (Con/SP; 49 plots). In addition to the host species composition field data, we also obtained the SBW defoliation data (333 plots classified into nil, light and moderate defoliation classes) for the entire state of Maine (2020) to evaluate our final SBW susceptibility based on stand impact types map.

3.2.3 Satellite data acquisition and preprocessing

3.2.3.1 Sentinel-2 optical data

We downloaded 30 Sentinel-2 (A, and B) Level 1C Top-Of-Atmosphere (L1C-TOA) reflectance images (https://earthexplorer.usgs.gov/) to include three seasons. Data were collected from mid-spring (MSp): May, mid-summer (MSm): June, and July, and fall (Fall): September, October, and November) (Table 3.1) to represent the unique biological events taking place on the tree canopies throughout the year. Essentially, the leaf-on (MSm, and MSp images) and leaf-off

(Fall images) conditions along with the distinct changes in leaf chemistry occurring towards the end of the fall season can be beneficial to discriminate the canopies of tree species from each other. All L1C-TOA products were converted into L2A surface reflectance products and subsequently resampled to 20 m spatial resolution using the Sen2Cor plugin available in Sentinel Application Platform (SNAP) software. Eventually, nine L2A spectral bands (B2: Blue (490 nm), B3: Green (560 nm), B4: Red (665 nm), B5: Red-Edge (705nm), B6: Red-Edge (740 nm), B7: Red-Edge (783 nm), B8A: Near Infrared (865 nm; NIR), B11: Shortwave Infrared (1610 nm; SWIR), B12: SWIR (2190 nm)) were used to derive spectral vegetation indices (SVIs) for further analysis. We used nine spectral bands and 27 Sentinel-2 based SVIs (Table A.1) from three seasons (108 variables altogether) as optical variables for modeling. Finally, all the spectral bands and SVIs from the respective seasons were masked for clouds/haze and non-forested areas (water bodies, developed areas, and agricultural lands) and mosaicked (reflectance values averaged at overlapping areas).

3.2.3.2 Sentinel-1 and PALSAR-2 SAR data

Sentinel-1 Ground Range Detected (GRD) C-band SAR images collected using interferometric wide swath (IW) mode were downloaded from two different sources to compare the effectiveness of multi-temporal speckle filtering vs monthly averaging during SAR data preprocessing. These preprocessing techniques are capable of suppressing the radar noises better, incorporating the temporal information (Bhattarai et al. 2022b), which were not tested in previous study conducted by Bhattarai et al. (2021). Data were collected for three seasons (MSp, MSm, and Fall) of 2018: one using the website of Alaska Satellite Facility (ASF; https://asf.alaska.edu/) (Table 3.1) and the other from Google Earth Engine (GEE) platform. The images downloaded from the ASF website were preprocessed in SNAP (Filipponi 2019), where the speckle filtering

step during the preprocessing workflow was performed using the multi-temporal speckle filtering technique (Bhattarai et al. 2022b). In contrast, the images downloaded from the GEE (already preprocessed except for the speckle filtering) were averaged monthly, which would ultimately suppress speckles. Finally, we acquired two separate sets of dual-polarized amplitude bands: VV (Vertical-Vertical) and VH (Vertical-Horizontal) resampled to 20 m spatial resolution. We used these amplitude bands to create Normalized Ratio Procedure between Bands (NRPB; Filgueiras et al. 2019) (Table A.1), resulting in two sets of nine C-band SAR variables from three seasons.

Freely available global 25 m resolution L-band PALSAR-2 yearly mosaic data (Table 3.1) from Japan Aerospace Exploration Agency were downloaded for 2018 from GEE platform. This yearly averaged world mosaic is in dual-polarization mode: HH (Horizontal-Horizontal), and HV (Horizontal-Vertical). The analysis-ready image data values were converted into the backscatter intensity values (gamma naught) in decibel units by using the calibration equation provided:

$$\gamma^{\circ} = 10 * log_{10}(DN)^2 + CF \qquad (1)$$

where, γ° is the gamma naught value, DN is the image product's digital number, and CF is the calibration factor provided. The HH and HV amplitude bands with gamma naught values obtained thereafter were resampled to 20 m spatial resolution for final analysis.

Table 3. 1. Multi-sensor remote sensing data used for our study. Multi-temporal images (from three seasons: mid-spring (MSp), mid-summer (MSm), and fall (Fall)) were collected for both Sentinel-1 and Sentinel-2 sensors, whereas an annual world mosaic was obtained for PALSAR-2.

Sensor	**Imagery date**	**Tile numbers/product ID**	**Resolution**
Multispectral optical instrument (Sentinel-2)	May 18 2018	TCL, TDN	Resampled to 20 m
	May 05 2018	TDK, TEK	
	May 24 2020	TDL	
	May 10 2018	TDM, TEM, TEN	
	May 13 2018	TCK	
	May 19 2020	TEL	

Table 3.1 continued…

	July 07 2018	TCK, TCL, TDN	
	July 09 2018	TDK, TEL, TEK	
	June 18 2020	TDL, TDM	
	July 19 2018	TEM, TEN	
	October 05 2020	TCK, TCL	
	October 17 2017	TDK, TDL, TEK	
	November 10 2020	TDM, TEM, TEN	
	September 30 2019	TDN	
	October 31 2020	TEL	
C-band SAR (Sentinel-1)	May 04 2018	D74F, 6067, 210C	Resampled to 20 m
	May 09 2018	937A	
	October 12 2018	DDC7	
	October 07 2018	1BA5, 517A, 9B05	
	July 08 2018	9C13	
	July 03 2018	0B57, 65C6, 31E2	
L-band SAR (PALSAR-2)	2018	----------	Resampled to 20m

3.2.4 Variable selection and modeling approaches

Table A.1 presents all Sentinel-1, and Sentinel-2 based SVIs used for our study. Altogether, we used 27 Sentinel-2 based SVIs and nine spectral bands, a single Sentinel-1 based SVI, and two bands (VV and VH) for three different seasons. The dimension of optical (Sentinel-2: 108 variables), and SAR (Sentinel-1: nine variables; PALSAR-2: two variables) data were reduced separately and they were supplemented with site data (elevation, slope, aspect, latitude, depth to water table (DWT; White et al. 2013), topographic wetness index (TWI; Hennigar et al. 2017), improved biomass growth index (iBGI; Rahimzadeh et al. 2020), and soil types (U.S. Department of Agriculture 2016). We used VSURF R package (Genuer et al. 2015) for the variables' dimensionality reduction, which selects the best predictor variables in three basic steps: 1) removal of irrelevant variables, 2) selection of all variables contributing to the prediction, and 3) elimination of the redundant variables obtained from Step 2. In addition, the Pearson correlation coefficient was estimated among the variables selected by the VSURF algorithm, and highly

correlated variables were removed. One of the variables in a pair under correlation evaluation with the correlation coefficient value smaller than - 0.7 and greater than + 0.7 (subjective evaluation) was removed based on the overall performance of the model with and without that particular variable. Ultimately, the optimum variables obtained from both remote sensing data (optical and radar) were modeled individually and combined with each other and site variables to obtain a best possible model.

For the classification of SBW host species, Random Forest (RF), one of the most widely used machine learning algorithms for remote sensing data analysis, was used. RF is an ensemble of decision trees that makes classification decisions based on the majority votes from the decision trees developed. It is a non-parametric machine-learning algorithm and returns out-of-bag (OOB) accuracy metrics every time the model is run, preventing the necessity of additional ground data for external model validation (Breiman 2001). We also formulated an ANN algorithm (Jain et al. 1996) to examine its performance against the RF algorithm. Our study used Multi-Layer Perceptron (MLP), a feedforward ANN well known for remote sensing data analysis (Astola et al. 2019; Foody 2004), which uses the backpropagation technique for model training. In addition, 10-fold cross-validation was performed for accuracy assessment in the case of the MLP algorithm, whereas the OOB accuracy metrics produced from the internal validation technique were adopted for RF. All the models were run in R 3.5.1 using the Caret package. Essentially, we ran our RF and MLP models in three different scenarios: 1) using Sentinel-2 variables only; 2) combining SAR (Sentinel-1 and PALSAR-2) with Sentinel-2 and site variables; and 3) combining SAR, and site variables only. In advance to running any of our model formulations, the training data was corrected for class-wise imbalances using R-package SMOTE as performed by Bhattarai et al. (2022b) to avoid the potential classification bias in the models.

3.2.5 Age determination of host species

Forest age determination using remote sensing technique is generally performed in combination with in-situ data (Maltamo et al. 2020; Zhang et al. 2014); however, we used the historic LCMAP products (spatial resolution of 30 m) from 1985 to 2018 (https://www.usgs.gov/core-science-systems/eros/lcmap) for the stand age estimation. We selected two products from the LCMAP database: 1) primary land cover (LCPRI) and 2) annual land cover change (LCACHG) data. The thematic change detection method was applied over two land cover classified images from 1985 and 2018 to detect the land cover transitions from one class to another. Among the eight land cover classes (Barren, Ice/Snow, Wetlands, Water, Tree Cover, Grass/Shrub, Cropland, and Developed) available in the LCPRI data, we were only concerned about the changes taking place among the Barren, Tree Cover, and Grass/Shrub cover types. All forest areas with the transitions; Barren to Tree Cover, Barren to Grass/Shrub, Tree Cover to Grass/Shrub, and Grass/Shrub to Tree Cover within the period of 33 years (1985-2018) were classified as young stands while the Tree Cover class that did not change throughout was classified as mature stands.

Additionally, we created a cumulative LCACHG image intersecting all the LCACHG images throughout the period of 33 years to examine probable intermediate disturbances on the Tree Cover class. This procedure avoids false classification of Tree Cover class as mature in 2018, which might have been harvested between 1985 and 2018 and regrown as forest later. Eventually, the LCMAP derived stand age data were used to classify the SBW host species in terms of their maturity into two classes (mature: > 33 years, and young: < 33 years). The age threshold for SBW host species maturity has been widely adopted as 40 years (Hennigar et al. 2008); nevertheless, we used a threshold age of 33 years given that the LCMAP data were

unavailable beyond the window of 1985-2018. Finally, SBW susceptibility to defoliation map was produced based on the host-maturity-specific ranking suggested by Hennigar et al. (2011) (mature BF > immature BF > mature SP > immature SP) and evaluated using SBW defoliation field data available for Maine.

3.3 Results

3.3.1 Variables' dimensionality reduction for modeling

The optimum Sentinel-2 variables selected after using the VSURF algorithm and eliminating correlated variables based on their correlation coefficients are presented in Figure 3.3. The variable size was reduced from 108 to eight, and from 11 to three for Sentinel-2, and SAR data, respectively. The final three variables selected from SAR data were VH_Fall and VH_MSm from Sentinel-1 and HV from PALSAR-2. Among the SAR variables selected after the variable reduction procedure, HV band from PALSAR-2 (L-band SAR) was the most important variable followed by VH_MSm, and VH_Fall variables from Sentinel-1 (C-band SAR) Sensor. Regarding Sentinel-1 variables, the monthly averaged image variables (obtained from GEE) performed slightly better than the single date image variables obtained from ASF (results not included), due to which single date SAR variables were dropped from further analysis. Similarly, the important Sentinel-2 variables selected after the variable elimination step comprised the bands and indices from all three seasons (MSp: B11_MSp, Sentinel-2 Red Edge Position (S2REP_MSp), Anthocyanin Reflectance Index1 (ARI1_MSp), MSm: S2REP_MSm, B8A_MSm, Normalized Difference Vegetation Index45 (NDVI45_MSm), and Fall: Red-Edge Normalized Difference Vegetation Index (NDVIRE_Fall), B11_Fall), demonstrating the key role of multi-temporal imagery in tree species classification. Among the selected variables, the majority of them were

from the red-edge region, with SVIs expressing biochemistry (NDVIRE, S2REP, and NDVI45) and physiology/stress (ARI1).

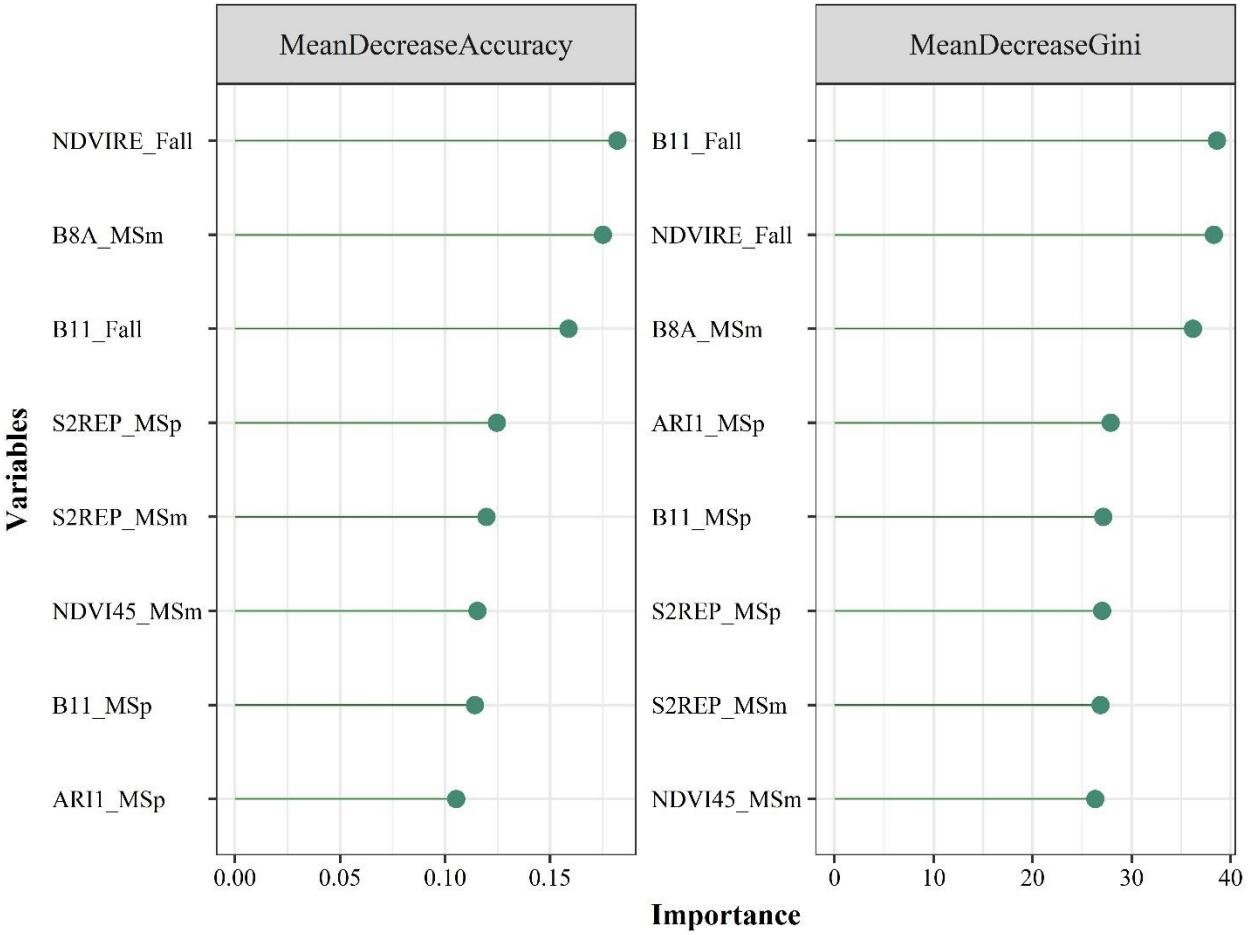

Figure 3. 3. Sentinel-2 variables obtained after the variables' dimensionality reduction using the Variable Selection Using Random Forest (VSURF) technique along with their importance in SBW host species classification modeling. The variable abbreviations were based on the convention: name of the variable_season of image acquisition (refer to Table A.1 for further information on variables). MSp: mid-spring; MSm: mid-summer; Fall: fall.

3.3.2 Model formulation, performance evaluation, and map production

We formulated several models with Sentinel-1, Sentinel-2, PALSAR-2, and site data using the RF and MLP algorithms. Among multiple models evaluated, the results for the best-performing models from the three scenarios mentioned in Section 2.4 are presented in Figure 3.4. The RF models yielded better outcomes for all the model formulations than MLP. Based on preliminary

comparisons of model performance, the hyper-parameters, number of trees grown (ntree) and variables at each split (mtry) for all the RF models formulated were 500 and two, respectively. In the case of MLP models, a single hidden layer attained the highest accuracy. MLP models with Sentinel-2 variables only, Sentinel-2, SAR and site variables, and SAR and site variables returned the best results (overall accuracy (OA) of 64.4%, 70.1%, and 49.5%) with nine, 14, and eight neurons in the hidden layer, respectively.

Using RF, the OA obtained with the best Sentinel-2 variables (Figure 3.3) was 81.7%. The model's class-specific accuracy for BF (PA: 90%, UA: 98%) and SP (PA: 94%, UA: 97%) classes were even better. On the other hand, the combination of Sentinel-2 variables and site variables (elevation, and aspect) produced the best overall accuracy (OA: 83.4%); addition of other site variables could not contribute to increasing the model accuracy. Nevertheless, this combination could not improve the accuracies for BF (PA: 90%, UA: 98%) and SP classes (PA: 94%, UA: 97%). The incorporation of SAR with Sentinel-2 and site variables did not improve the Sentinel-2 + Site model further. The best model with SAR and site variables (SAR + Site) included Sentinel-1 (VH_Fall, VH_MSm), PALSAR-2 (HV), and site variables (elevation, aspect, and slope) and resulted in an OA of 77.9% (Figure 3.4). The model also demonstrated a reasonable accuracy for BF (PA: 88%, UA: 88%) and SP (PA: 89%, UA: 89%) classes.

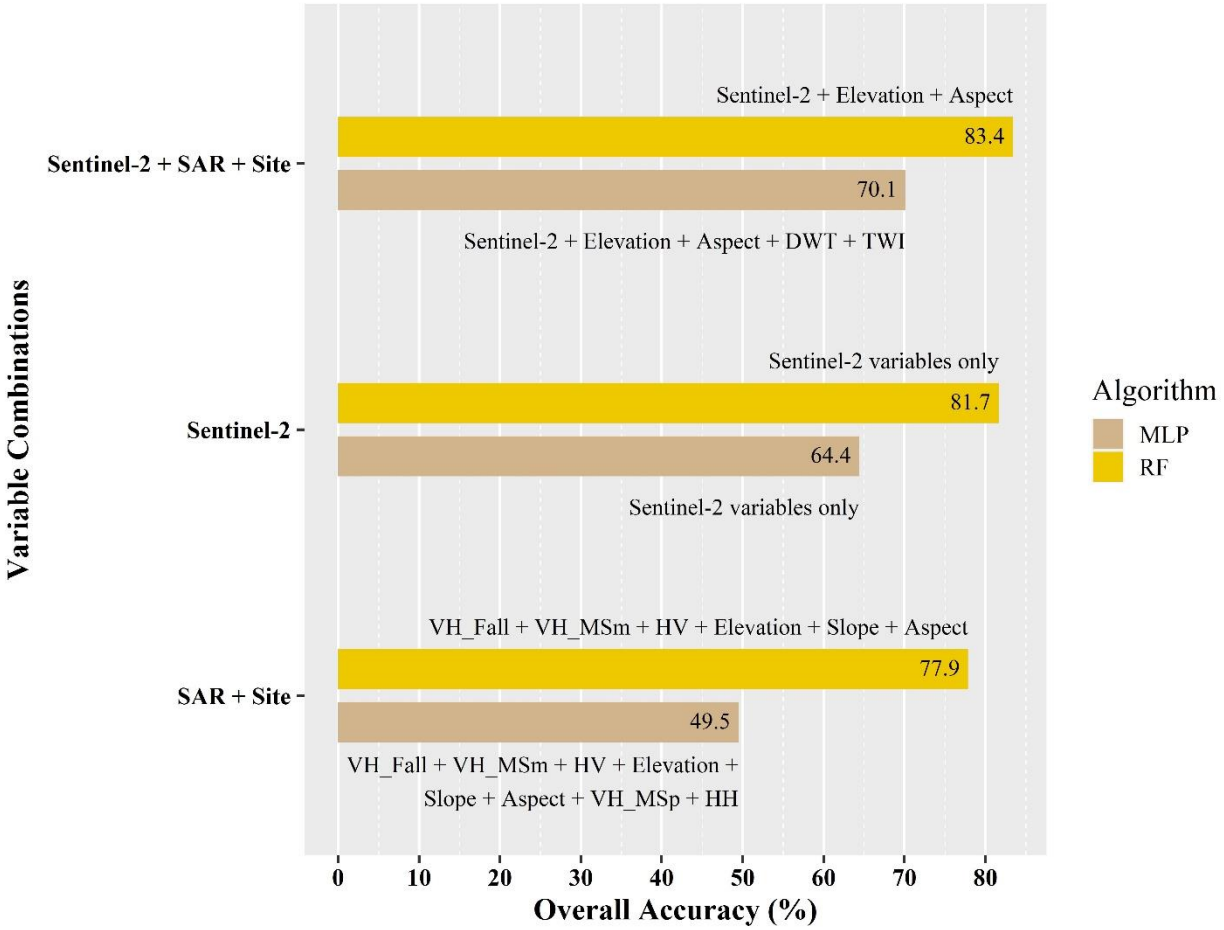

Figure 3. 4. Comparison of the best performing RF and MLP classification models based on their overall accuracy (OA) presented as numbers over the bars. Sentinel-2 variables were the variables obtained after running the VSURF algorithm for variables reduction. RF accuracies were based on the OOB error estimates, while MLP accuracies were obtained after 10-fold cross-validation. DWT: Depth to Water Table; TWI: Topographic Wetness Index; MSp: mid-spring; MSm: mid-summer; Fall: fall.

Evaluating the classification accuracy and confusion associated with the two different algorithms tested in this study (Figure 3.4), the accuracy was relatively higher for separating broadleaved from conifer classes compared to the separation between the conifer classes. The overall best RF model (Sentinel-2 + Elevation + Aspect), classified the tree species into eight classes with good producer's accuracy (PA), and user's accuracy (UA), especially for BF (PA: 90%, UA: 98%) and SP (PA: 94%, UA: 97%) classes. Similarly the best model with MLP and site variables (Sentinel-2 + Elevation + Aspect + DWT + TWI) also indicated promising class-wise

accuracies for BF (PA: 87%, UA: 94%), and SP (PA: 89%, UA: 88%) classes. However, the class-wise accuracies for other classes declined considerably when we used MLP in place of RF algorithm (Table 3.2).

Table 3. 2. Confusion matrices for the overall best performing RF (out of bag error based) and MLP (10-fold cross-validated) models and their accuracy estimation for the SBW host species classification. The entries are the % occurrences in each class. UA is the user's accuracy, and PA is the producer's accuracy. BF: balsam fir; BF/BL: balsam fir and broadleaved; BL: broadleaved species; Con: conifers other than balsam fir and spruce majority; Con/BF: coniferous and balsam fir; Con/BL: coniferous and broadleaved; Con/SP: coniferous and spruce; SP: spruce.

Models	Species	BF	BF/ BL	BL	Con	Con/ BF	Con/ BL	Con/ SP	SP	Total	PA (%)
Best RF (Sentinel-2 + Elevation + Aspect)	BF	**90**	2	6	0	0	0	0	2	100	90
	BF/BL	0	**86**	0	0	14	0	0	0	100	86
	BL	0	3	**91**	3	0	0	3	0	100	91
	Con	0	0	9	**61**	4	13	13	0	100	61
	Con/BF	2	2	0	0	**88**	3	5	0	100	88
	Con/BL	0	0	3	17	14	**59**	7	0	100	59
	Con/SP	0	0	0	7	6	6	**81**	0	100	81
	SP	0	0	3	0	0	0	3	**94**	100	94
	UA (%)	98	89	82	64	83	71	71	97		
Overall accuracy: 83.4%											
Best MLP (Sentinel-2 + Elevation + Aspect + DWT + TWI)	BF	**87**	5	6	0	0	0	0	2	100	87
	BF/BL	1	**39**	8	2	46	2	2	0	100	39
	BL	3	4	**84**	6	0	2	0	1	100	84
	Con	1	1	4	**50**	5	24	15	0	100	50
	Con/BF	1	5	0	2	**74**	10	7	1	100	74
	Con/BL	0	3	5	14	29	**40**	8	1	100	40
	Con/SP	0	3	2	6	13	4	**63**	9	100	63
	SP	1	0	0	0	1	1	8	**89**	100	89
	UA (%)	94	54	78	54	65	44	59	88		
Overall accuracy: 70.1%											

The two best models, one from each RF, and MLP algorithm (Table 3.2) category, were used to produce host species composition maps of the study area at 20 m spatial resolution (Figure 3.5). The individual map classified the study area into eight tree species classes (five hosts: BF, BF/BL, Con/BF, Con/SP, and SP; three non-hosts: Con, BL, and Con/BL). Host species composition maps created using both algorithms demonstrate the dominance of SBW host species,

mainly BF and SP, in the north, whereas broadleaved species (non-hosts) are more frequently found towards the southern part. In addition, we presented detailed classification maps in Figure 3.5c and 3.5d for a subset of our study area to facilitate a better comparison between the maps created using RF, and MLP algorithms.

3.3.3 Classifying tree species by maturity and mapping stand impact types and susceptibility to SBW defoliation based on stand impact types

The species composition map produced from our overall best model (Figure 3.5a) was used further to classify tree species classes based on their maturity (Figure 3.6), also named as SBW stand impact types map in our study. Based on the available data, the forest age classes derived from the LCMAP products were applied to the species composition map to classify each of those eight classes into two maturity classes, young (< 33 years) and mature (> 33 years).

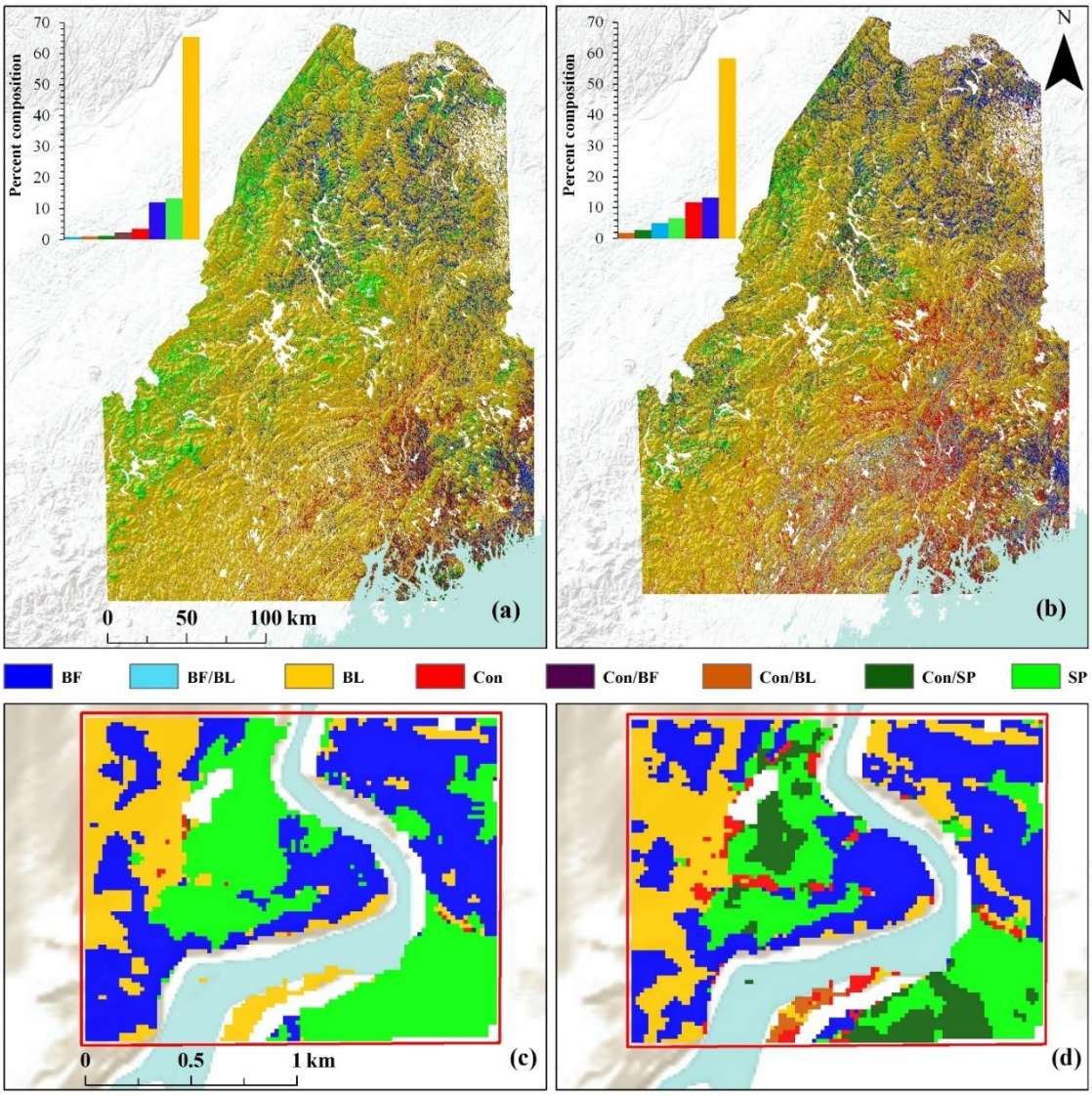

Figure 3. 5. SBW host species composition maps at 20 m spatial resolution produced using the overall best RF (a) and MLP (b) models with their respective subsets (c, and d) presenting the species classes in detail (refer to Table 3.2 for more information on variables used and model performances). The central coordinates of the selected subsets are -69.48°W and 46.87°N. BF: balsam fir; BF/BL: balsam fir and broadleaved; BL: broadleaved species; Con: conifers other than balsam fir, and spruce majority; Con/BF: coniferous, and balsam fir; Con/BL: coniferous, and broadleaved; Con/SP: coniferous, and spruce; SP: spruce.

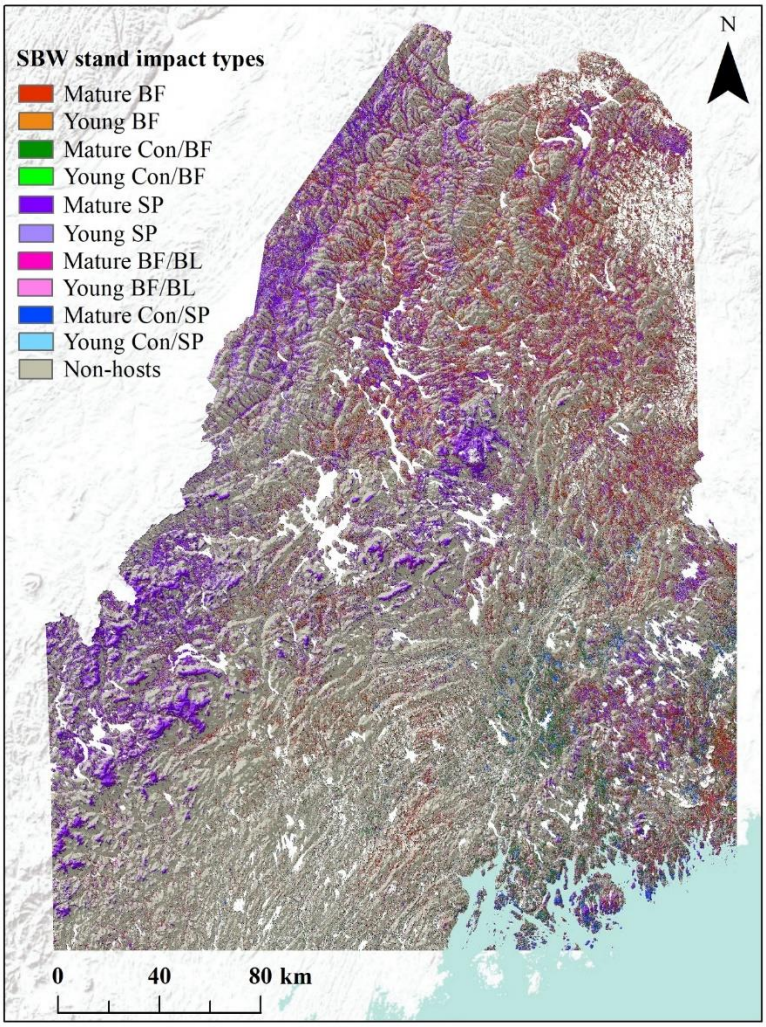

Figure 3. 6. Age adjusted SBW stand impact types map created coupling the stand age information with the host species composition map. BF: balsam fir; BF/BL: balsam fir and broadleaved; BL: broadleaved species; Con: conifers other than balsam fir and spruce majority; Con/BF: coniferous, and balsam fir; Con/BL: coniferous, and broadleaved; Con/SP: coniferous, and spruce; SP: spruce.

We used the SBW stand impact types map (Figure 6) to produce SBW susceptibility map (Figure 3.7a). The SBW susceptibility map was based on the defoliation impact rating of our host classes as: Mature BF > Young BF > Mature SP > Young SP > Mature Con/BF >Young Con/BF >Mature Con/SP > Young Con/SP > Mature BF/BL > Young BF/BL (Hennigar et al., 2011). Evaluating the SBW susceptibility based on stand impact types map for the state of Maine, it is evident that northern forests are at much higher risk to the SBW as compared to the southern forests.

Figure 3.7 illustrates the current status of SBW defoliation in Maine (2020) overlaid on the SBW susceptibility based on stand impact types map created from our study. The overall trend of defoliation across the state exhibits higher defoliation in the northern part. Moreover, the severity of SBW defoliation on the ground from 2020 coincides with the SBW susceptibility based on stand impact types predicted by our study (better seen in zoomed windows in Figure 3.7c and 3.7d). Among 333 comparison points (nil: 155; light: 144; moderate: 34) with percent SBW defoliation data available, 88.3% were in agreement with the host species classes predicted by our model (Table 3.3). Furthermore, 89.7% of the validation points with nil (< 10%) ground defoliation fell within the first five high SBW stand impact type classes (Mature BF, Young BF, Mature SP, Young SP, and Mature Con/BF) (Table 3.3 and Figure 3.7). However, for light (11% - 30%) and moderate (31% - 70%) defoliation classes, 87.5%, and 85.3% of points, fell within the first four high SBW stand impact type classes (Mature BF, Young BF, Mature SP, and Young SP), respectively. Evaluating the 11.7% defoliation points falling outside our susceptibility based on stand impact types map, 87.2% were nil to light in terms of defoliation severity whereas only 12.8% were moderately defoliated (Table 3.3).

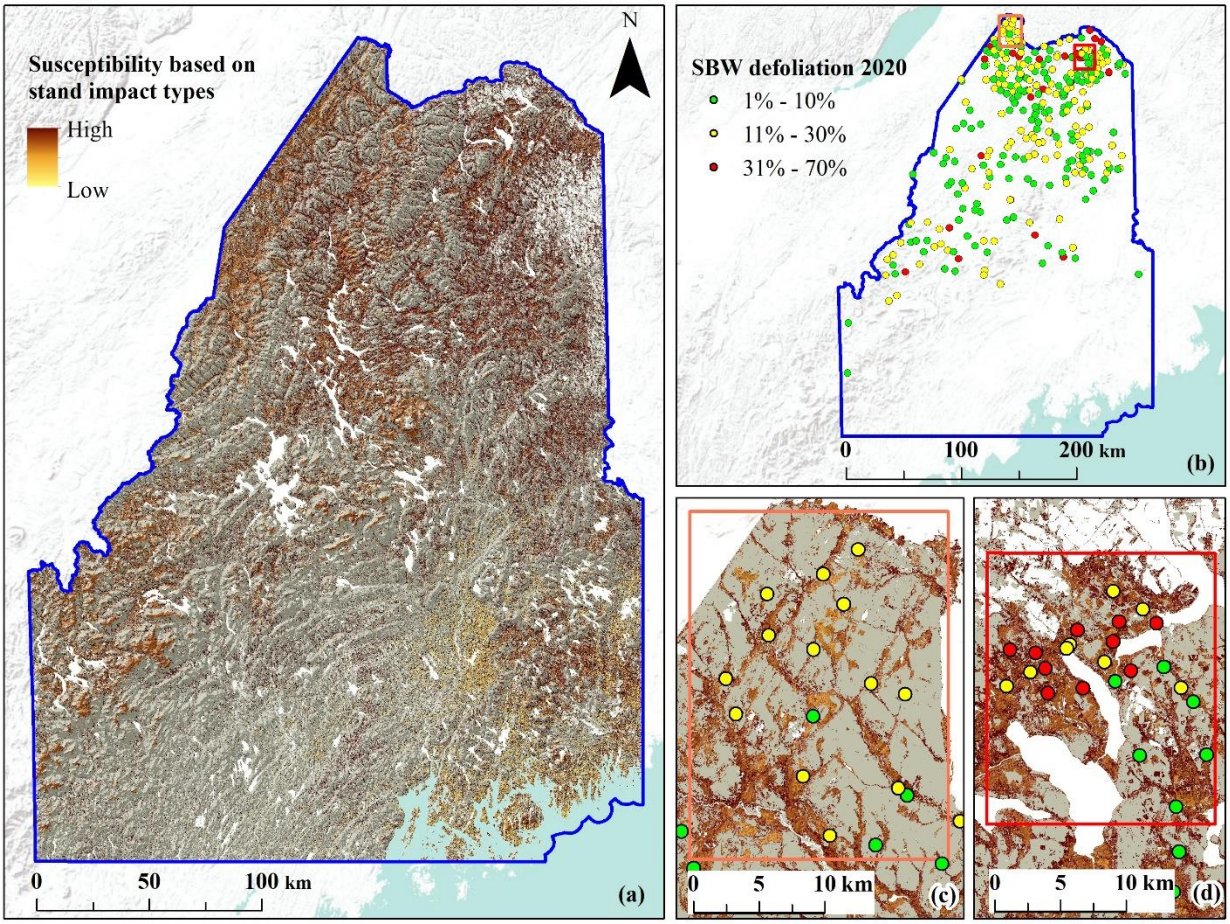

Figure 3. 7. The SBW susceptibility based on stand impact types map created from our study (a) alongside the SBW defoliation recorded in Maine for the year 2020 (b; 333 plots). Two subsets of study area represent the areas with the majority of light defoliation (c; 11% - 30%) and moderate defoliation (d; 31% - 70%). BF: balsam fir; BF/BL: balsam fir and broadleaved; Con/BF: coniferous and balsam fir; Con/SP: coniferous and spruce; SP: spruce.

Table 3. 3. Evaluation of the agreement between the SBW susceptibility based on stand impact types and the percent SBW defoliation recorded over multiple sites throughout the state of Maine for 2020 (333 plots). The SBW defoliation data used for validation are not necessarily from pure host species plots; rather they indicate the severity of SBW defoliation in selected trees in those sites.

SBW defoliation level 2020	**Points falling within host classes (%)**	**Points falling outside host classes (%)**
Nil (0 – 10)%	89.7	10.3
Light (11 – 30)%	87.5	12.5
Moderate (31 – 70)%	85.3	14.7
Overall	88.3	11.7

3.4 Discussion

Classification of tree species from a pest-susceptibility point of view is a relatively novel concept and very few studies have attempted to explore the potential of remote sensing techniques for this purpose (Bhattarai et al. 2021; Wolter and Townsend 2011). Our study modeled and mapped tree species for nearly the entire state of Maine (~6.5 million ha of forest) and assessed their potential susceptibility to the imminent SBW outbreak. To the best of our knowledge, potential susceptibility assessment of forests to SBW defoliation based on age and composition on a landscape level using remote sensing data has never been reported in the literature. Moreover, the host species composition map produced in this study is by itself a valuable resource and a key layer for SBW defoliation detection using either remote sensing data or aerial surveys. We used an array of optical (Sentinel-2), radar (Sentinel-1, and PALSAR-2), and several topographic variables (elevation, slope, aspect, latitude, TWI, DWT) as well as soil types and iBGI for modeling tree species using RF, and MLP algorithms. We also utilized the LCMAP data to produce the age adjusted SBW susceptibility map based on stand impact types.

3.4.1 Random Forest and Multi-Layer Perceptron performances for forest composition mapping

The MLP algorithm used for modeling host species did not perform better than the RF algorithm (Table 3.2, and Figure 3.4) in our study. The cross-validated OA (70.1%) for the best MLP model were considerably lower than those obtained with the best RF model (OA: 83.4%). Our findings align with the results from Zagajewski et al. (2021) where they obtained a better accuracy using RF than ANN model (OA: 85% vs 77%) working with Landsat-8 imagery for tree species classification in Czech-Polish border. According to the authors, the use of ANNs like MLP might be inefficient to classify forests with the moderate resolution pixel information derived from

Landsat-8 and Sentinel-2 imagery. Furthermore, several studies on the performance comparison between ANN and RF models outside the discipline of tree species or vegetation classification suggest the better performance of RF (Sevgen et al. 2019; Wang et al. 2018a) compared to the performance of ANN models similar to what was found here.

In contrast, Raczko and Zagajewski (2017) produced better results with ANN than RF (OA: 77% vs 62%) for tree species classification using hyperspectral imagery (spatial and spectral resolution of 3.35 m and 288 bands, respectively) in the Northwestern Poland. However, they used a higher number of training data (120 sample pixels for each class) as compared to the number of training data we used (23 - 60 sample plots per class). Further, they did not use the internal validation technique available in the RF algorithm that would have provided more data for RF model training, preventing the need for extra samples for model validation and their overall accuracy was not higher than what we obtained in this study. Future studies may also consider the use of multi-objective algorithms (Legaard et al. 2020) for species classification.

3.4.2 Role of site variables in classifying tree species

There are multitudes of studies on tree species composition and forest type classification modeling using remote sensing data; nevertheless, little attention has been given to incorporating topographic variables in the model. We tested several available site variables in our host species classification models and reported their importance. The best model derived using SAR variables performed better with the inclusion of site variables (elevation, aspect, and slope) in our study. Our overall best model (Sentinel-2 and site variables) includes two site variables, namely, elevation, and aspect, which is in agreement with the study conducted by Grabska et al. (2020) on tree species classification in the Polish Carpathians, combining Sentinel-2, and site variables. Their study demonstrates the highest contribution of elevation among all other variables used for

modeling. Similarly, Bhattarai et al. (2021) and Hoscilo and Lewandowska (2019) also emphasize the importance of elevation in tree species mapping when used together with the remote sensing data. Elevation is also an important factor related to landscape-level SBW defoliation patterns during an outbreak (Cen et al. 2021), which further supports its use in our analysis.

3.4.3 Contribution of L-band vs C-band SAR variables in tree species classification

The better performance of L-band PALSAR-2 over C-band Sentinel-1 data for tree species classification in our study aligns well with the results from similar previous studies (Li et al. 2012; Turkar et al. 2012; Wolter and Townsend 2011). There have been several research reports on tree species (Bhattarai et al. 2021) or forest types classification (Erinjery et al. 2018) using Sentinel-1, and Sentinel-2 imagery where the major setback of C-band SAR for tree species discrimination has been attributed to its shorter wavelength (lower canopy penetration). The issue of lower canopy penetration could be overcome using longer wavelength SAR data like L-band PALSAR-2, which is also well demonstrated by our results. There is a breadth of literature currently available on the contribution of C-band SAR data (Sentinel-1) to forest types mapping when coupled with Sentinel-2 optical data (Biswas et al. 2020; Erinjery et al. 2018); however, these studies are focused on the classification of broad forest types with very few SVIs used from Sentinel-2 data. For example, Eringery et al. (2018) used Sentinel-1 and Sentinel-2 datasets for mapping forest types in the tropical rainforest of Western Ghats, India, into five classes and reported an improved accuracy after the inclusion of Sentinel-1 bands. However, NDVI was the only SVI used from Sentinel-2 data. Similarly, the study on forest types mapping (six classes) in southern Myanmar using Sentinel-1, Sentinel-2, and Landsat-8 conducted by Biswas et al. (2020) mainly used spectral bands from optical sensors failing to make optimum use of numerous possible red-edge SVIs

(Table A.1), especially from Sentinel-2 imagery, which would have better vegetation identification capabilities as compared to spectral bands only.

3.4.4 Contribution of SAR vs optical multi-temporal variables in tree species classification

The better performance of individual Sentinel-2 based variables as compared to SAR variables in our study could be attributed to the potential of Sentinel-2 sensor to capture information for a wide range of tree canopy properties like structure, biochemistry, and physiology (Bhattarai et al. 2021; Erinjery et al. 2018). On the other hand, SAR images are generally helpful for detecting the tree canopy structure and are not sensitive to the chemical composition of foliage. The integration of SAR (particularly L-band SAR) with site data produced a promising result (OA: 77.9%) for species classification. However, the model with Sentinel-2 and site variables performed the best (OA: 83.4%). Yet, the SAR and site data based model could be a good alternative in the absence of atmospherically uncontaminated optical images like Sentinel-2. Our findings demonstrate the dominance of multi-temporal Sentinel-2 based red-edge SVIs (NDVIRE, NDVI45, S2REP, and ARI1; see Table A.1 for the abbreviations), as well as NIR (B8a) and SWIR (B11) bands for tree-species identification. Similar to our outcomes, Bolyn et al. (2018) reported the prime role of Sentinel-2 based SWIR (B11), NIR (B8a), and red-edge variables for species composition mapping using multi-temporal Sentinel-2 imagery in the Belgian Ardenne ecoregion. Furthermore, Bhattarai et al. (2021) indicated the importance of Sentinel-2 based red-edge, SWIR, and NIR variables for forest composition mapping.

The efficacy of multi-temporal Sentinel-2 based red-edge, SWIR, and NIR variables for tree-species classification has been widely reported in several other recent studies (Hoscilo and Lewandowska 2019; Grabska et al. 2020). Our study demonstrates the usefulness of imagery from the onset of growing season (spring to summer) and senescence (fall) for discriminating tree

species, which aligns with the findings from similar investigations for optical (Grabska et al. 2019; Hoscilo and Lewandowska 2019; Immitzer et al. 2019) as well as SAR imagery (Bhattarai et al. 2021). The significance of multi-temporal imagery for species classification is grounded on the fact that they have the potential to capture a series of unique phenological events taking place in plant canopy throughout the year (Bhattarai et al. 2021; Grabska et al. 2019; Immitzer et al. 2019). In case of dual polarized SAR bands (VV and VH from Sentinel-1 and HH and HV from PALSAR-2), the cross-polarized bands (VH, and HV) were more important for tree species discrimination in our study (Figure 3.4), which is in line with the results from Rignot et al. (1994), and Bhattarai et al. (2021). However, the importance of Sentinel-1 based SVI (NRPB) was not noticeable in our study.

3.4.5 Evaluating tree species classification model performances

The high species classification accuracy we attained based on Sentinel-2 and site variables (OA: 83.4%) was encouraging. In particular, the class-wise accuracies for the major host species classes, BF (PA: 90%, UA: 98%) and SP (PA: 94%, UA: 97%) were notable. A recent study conducted on SBW host species classification using Sentinel-1, Sentinel-2, and site variables in northern New Brunswick reported the highest OA of 73% for classifying host species into five classes with class-wise accuracies of PA: 73% and UA: 71% for balsam fir and PA: 77% and UA: 71% for spruce classes (Bhattarai et al. 2021). In this work, we classified SBW host species into eight classes and obtained significantly higher accuracies. Similarly, Wolter and Townsend (2011) used a combination of Landsat TM, SPOT, Radarsat (C-band SAR), and PALSAR (L-band SAR) to classify tree species in northern Minnesota into 12 classes (three SBW host species classes) and achieved an OA of 78%. Among the SBW host species classified, the class-wise accuracy for balsam fir (PA: 67%, UA: 80%), black spruce (PA: 83%, UA: 80%), and white spruce (PA: 64%,

UA: 91%) were comparatively lower than what we achieved. Regarding general tree species classification models, Hoscilo and Landowska (2019) used multi-temporal Sentinel-2 data to classify tree species in southern Poland with an OA of 81.7% (lower than ours). Furthermore, Grabska et al. (2020) achieved an OA of 85.6% (comparable to ours); classifying Polish Carpathians' tree species into 11 classes using multi-temporal Sentinel-2 imagery and site variables. However, they used a complex model with 64 variables for the prediction. Similarly, Grabska et al. (2019), and Immitzer et al. (2019) attained a better OA compared to ours (92.4% and 89%, respectively), for tree species classification using multi-temporal Sentinel-2 imagery; yet, those studies focused on classifying relatively homogenous forests. In contrast, we attempted to identify the closely associated and mixed host species classes, for example, BF, Con/BF, BF/BL, SP, Con, and Con/SP. On the other hand, several other related studies (Biswas et al. 2020; Liesenberg and Gloaguen 2013; Yu et al. 2018) achieved better OA than our model; nevertheless, their emphasis was on vegetation type classification rather than individual tree species or host species classification.

3.4.6 SBW susceptibility mapping based on stand impact types and its implications in Maine

In addition to the spatial delineation of host species, their age assessment is crucial from the pest vulnerability point of view, especially for the SBW (Bhattarai et al. 2020; Hennigar et al. 2011); yet, it is a challenging task to estimate the age of forests on a landscape level. Presumably, there have not been any publications to compare with our study regarding the maturity mapping of tree species in light of pest vulnerability using remote sensing data. We used the freely available archives of LCMAP (LCPRI and LCACHG) products to estimate the age of host species and classify them into two classes (young and mature) for the first time. The maturity map created

could be used as a baseline product for the future to predict the annual vulnerability of the host stands incorporating several dynamic variables like winter temperature, wind characteristics, SBW larval counts, etc. Besides, the final product SBW susceptibility based on stand impact types map obtained from our study independently possesses the potential to inform the stakeholders about the host stands which are at a higher risk of damage from defoliation. The SBW susceptibility map evaluated using the latest SBW defoliation ground data (2020) from Maine demonstrates high promise with an overall agreement of 88.3% between our impact rating and SBW defoliation on the ground (Table 3.3, and Figure 3.7). The SBW susceptibility based on stand impact types map covering almost the entire state of Maine is the first-ever product attempted towards the landscape level (~6.5 million ha area) potential susceptibility mapping. Taking note of the forthcoming SBW outbreak in Maine, our map products at finer spatial resolution of 20 m will serve as a valuable supplement for forest managers to conduct management interventions on a larger scale to prevent the forest from possible future damage.

3.5 Conclusion

Our study presents a SBW susceptibility product derived from the integration of host species composition and their maturity on a landscape level (~6.5 million ha) at 20 m spatial resolution using freely available multi-source remote sensing and site data. We obtained the best modeling accuracy using the RF algorithm, demonstrating its robustness and suitability for tree species classification even with limited data; however, the MLP algorithm is worth exploring further, especially when larger training samples are available. As expected, incorporating SAR data with optical data for tree species classification was helpful, particularly the L-band PALSAR-2 data, but the performance of our best model was already high without the contribution of SAR data. In addition, our study demonstrates strong evidence of the importance of site variables for

tree species identification in complex and diverse landscapes, which should be integrated and investigated more in related future endeavors. Furthermore, we suggest an alternative model with a decent accuracy (OA: 77.9%) for tree species classification using SAR and site data in case good-quality optical images are unavailable.

Age determination of forest stands on a landscape level is challenging; nonetheless, our study proposes a promising method to estimate it using a time series of historical land-cover classification data. As of present, our final product (SBW susceptibility based on stand impact types map) is believed to benefit the forest stakeholders in developing effective pest management strategies to minimize the impact of the approaching SBW outbreak. At the same time, our study provides a platform for future studies to improve the proposed SBW susceptibility classification model based on stand impact types capturing the advancements in remote sensing technology.

CHAPTER 4

ESTIMATING NUTRITIVE, NON-NUTRITIVE AND DEFENSE FOLIAR TRAITS IN SPRUCE-FIR STANDS USING REMOTE SENSING AND SITE DATA

"The contents of this chapter have been published in Bhattarai et al. (2023)."

4.1 Introduction

Leaf foliar chemistry plays an important role in host plant quality for herbivores by providing necessary components of their diet that can increase feeding, or by producing defensive compounds that could deter feeding. The composition of compounds and elements in leaves can also be used as indicators of plant stress as well as productivity over time, and the variability in foliar chemistry between different host species has been correlated to different levels of resistance to herbivory (Herms and Mattson 1992; Fuentealba and Bauce 2016). Collecting foliage samples followed by laboratory analyses is a traditional and seemingly easy approach to assess foliar traits; however, it can be extremely time consuming and expensive to make temporally extensive comparisons (Chlus and Townsend, 2022). Furthermore, mapping the concentration of these elements in forestlands at a landscape level can be even more challenging.

One major herbivore threatening the health and productivity of the forests in northeastern USA and neighboring Canadian provinces is eastern spruce budworm (SBW; *Choristoneura fumiferana* Clemens). SBW is a native forest pest in the region and its outbreak has a return period of approximately three to seven decades (Royama, 1984), mainly affecting balsam fir [*Abies balsamea* (L.) Miller] and spruce (*Picea* spp.). SBW has eight stages in its lifecycle including six larval instars (L1 being the youngest and L6 being the oldest), along with pupal and adult stages where L5 and L6 instars cause most of the damage (consumption of ~ 95 % of the host needles). The larva enters hibernation in late summer, emerges in the following spring as L2 and starts

feeding on host buds and new foliage (Miller, 1977). Multiple consecutive years of feeding can cause the removal of all age classes of foliage in host trees. The feeding larva constructs feeding sites in host branches by webbing together several branches and creating a tunnel. SBW has the potential to completely defoliate and kill a host tree in around half a decade after the onset of outbreak if no interventions are applied (Maclean et al., 2019). The outbreaks can last between 5 and 20 years in the affected area and they can be highly catastrophic at the landscape scale (Rahimzadeh-Bajgiran et al., 2018; Bhattarai et al., 2021).

SBW's main host species, balsam fir and spruce, possess a variety of chemical elements and compounds (referred to as foliar traits hereafter), such as iron (Fe), calcium (Ca), cellulose, condensed tannins, terpenes and phenolics, that can reduce herbivory (Mattson et al., 1991; Fuentealba and Bauce, 2016; Fuentealba et al., 2020). In contrast, the host foliage also contains other foliar traits, including copper (Cu), nitrogen (N), potassium (K), phosphorous (P), and sugars, that could increase the palatability of the needles and enhance insects' fitness (Mattson et al., 1991).

The level of impact from SBW attack on host trees can be influenced by the combination of both nutritive and non-nutritive foliar elements. Therefore, evaluating the chemical profile of host trees would be valuable in assessing their risk to SBW attacks. In addition, the mechanical properties of needles, particularly their toughness, can be instrumental in the success of herbivores. These morphological traits, which are influenced by the amount of water, fiber, and cellulose, are as important, if not more so, than chemical properties (Fuentealba et al., 2020). Information on foliar traits can also be used for assessing ecological functions including forest productivity, nutrient content and leaf litter decay rates (Curran, 1989).

Different traits in plants are responsible for unique roles, and the need for their estimation depends on the objectives of the study. For instance, while the estimation of N and chlorophyll

abundance can be used to assess the vigor and productivity of a forest (Abdullah et al., 2019), these components alone might not be enough to estimate the quality of a tree for an attacking herbivore. When evaluating the host quality with respect to a pest such as SBW, interactions between the plant and insect warrant the assessment of both nutritive and non-nutritive traits. In the past, several researchers attempted to assess the susceptibility of host species to SBW in a laboratory setting by evaluating the effects of several micro- (Fe and Cu) and macro-nutrients (Ca, K, Mg, N, P) on SBW health and reproduction (Mattson et al., 1991). More recently, studies have focused on assessing the nutrients and defensive chemicals present in SBW host species (Fuentealba and Bauce, 2016; Funtealba et al., 2020). The relationship between foliar chemistry and SBW defoliation proposed by these studies have the potential to be linked with remote sensing sensors and up-scaled to a landscape level to produce comprehensive SBW susceptibility/resilience maps. Currently, there is no literature on using remote sensing data to map either nutrients/non-nutrients or the defensive mechanisms/chemicals present in SBW host species.

There are well-established remote sensing techniques (from field spectroscopy to satellite remote sensing) that relate foliar chemistry to the reflectance spectra, which are worth exploring for foliar trait measurements. Advancements in technology have revolutionized the field of remote sensing, allowing for signal acquisition from the ground in different spatial, spectral, and temporal scales suited for target applications. There have been numerous successful attempts to map canopy traits and leaf pigments using satellite imagery. For instance, high resolution multispectral imagery such as RapidEye (Darvishzadeh et al., 2019) has been used in estimating several plant foliar traits including chlorophyll, Mg, carbon (C), N, P, K, Ca, and water content. Studies on foliar pigment estimation using hyperspectral imagery (Huber et al., 2008; Axelsson et al., 2013; Chlus and Townsend, 2022) suggest that high spectral resolution imagery may be even better at assessing

foliar biochemical traits (Curran, 1989; Chlus and Townsend, 2022). However, the use of medium resolution multispectral satellite data, such as Sentinel-2 imagery, for estimating leaf chemistry is just starting to attract attention within the scientific community with promising results (e.g., Gara et al., 2019; Ma et al., 2019; Gara et al., 2022). This success could be attributed to the presence of red-edge bands, which are sensitive to canopy pigments, and higher temporal resolution of the data. Using Sentinel-2 data to estimate canopy traits like chlorophyll (Darvishzadeh et al., 2019; Gara et al., 2019) and N (Abdullah et al., 2019; Gara et al., 2019; Ma et al., 2019) has been reported but other essential traits such as P, K, Ca, Fe, Cu, Mg, etc., are less reported in the literature.

Generally, it has been a common practice to rely on multivariate statistical methods to model foliar traits using remotely sensed data (Curran, 1989). However, with the emergence of artificial intelligence, several machine-learning algorithms have been employed to model the complex relationship between plant foliar traits and the associated reflection spectra better (Axelsson et al., 2013; Gara et al., 2019; Chlus and Townsend, 2022). In this study, we used Sentinel-2 spectral bands and derived spectral vegetation indices (SVIs) to model foliar traits on a landscape level using modern machine learning algorithms. Our study aimed to estimate and map several important foliar traits of SBW host species (spruce and fir), with an emphasis on their defensive characteristics against herbivory. In general, it is often assumed that the concentration of non-nutritive and defensive traits is inversely related, while the concentration of nutritive traits is positively related to host susceptibility. However, it is important to consider that certain factors can influence these relationships. For instance, when it comes to the concentration of nutrients like N and P, they are essential components of photosynthesis. Their low concentration will force plants to allocate fewer resources to the production of defensive compounds as they prioritize using these nutrients for photosynthetic processes. Consequently, this reduced investment in defensive

compounds can lead to higher susceptibility of host species to herbivory. Additionally, equivalent water thickness (EWT), although considered a defensive trait in our study, is directly associated with the softness of the plant's needles, indicating that a higher EWT could make the host species more susceptible to herbivorous attacks. In contrast, leaf mass per area known as LMA is a measure of needle toughness with higher values preventing herbivory. Specifically, our study sought to address the following objectives:

1. Estimate and map both the nutritive (N, P, K, and Cu) and non-nutritive (Fe, and Ca) elements using Sentinel-2 bands and SVIs, along with site variables in SBW host species.
2. Estimate and map mechanical defensive traits, EWT and LMA of SBW host species using Sentinel-2 bands and SVIs, along with site variables.
3. Compare the performance of modern machine learning algorithms for modeling SBW host species foliar traits.
4. Propose a framework for determining the susceptibility of SBW host species by assessing the relationship between host foliar trait values and SBW defoliation levels.

4.2 Materials and Methods

Figure 4.1 provides an overview of the study workflow. Subsequent sections will explain each step in more detail.

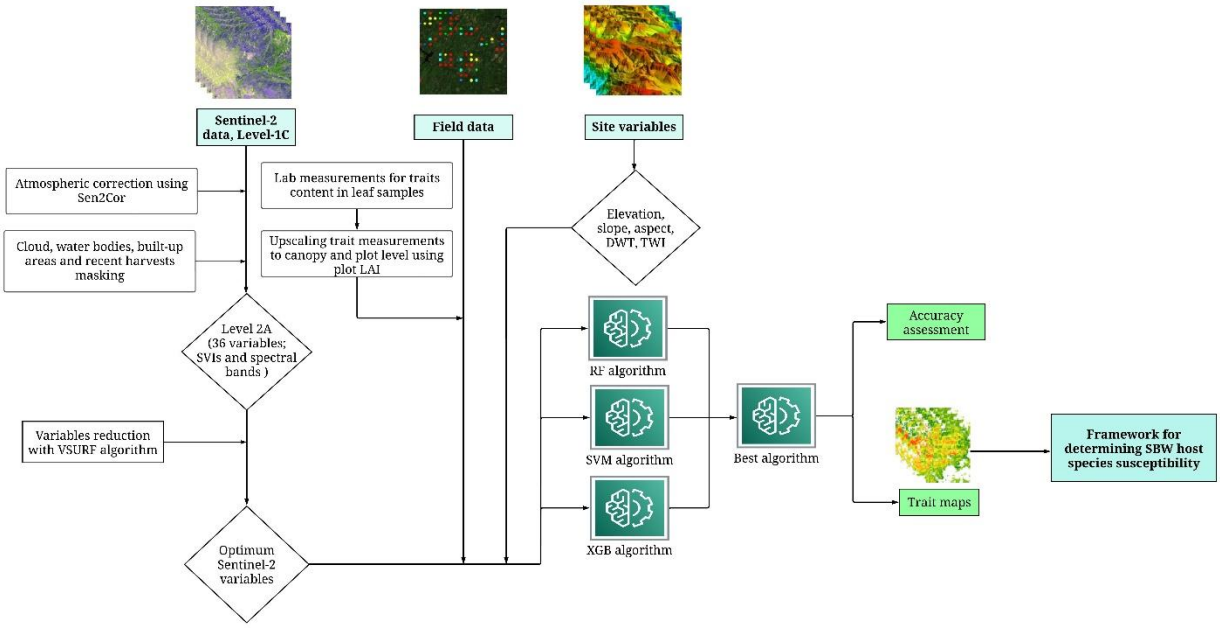

Figure 4. 1. The typical sequence of the steps involved in our study. The entire study was organized into four major segments: 1) remote sensing data collection and preprocessing (including site variables), 2) leaf samples collection, laboratory analysis, and upscaling the leaf traits to plot level, 3) modeling the traits using random forest (RF), extreme gradient boosting (XGB) and support vector machine (SVM) algorithms, and 4) establishing a framework to link trait measurements to spruce budworm (SBW) host species susceptibility to the attack. DWT: depth to water table; LAI: leaf area index; TWI: topographic wetness index.

4.2.1 Study area

The study area is located in Hancock County, Maine, USA. It encompasses a portion of Donnell Pond, which is public property situated only a few miles from the Atlantic coast (Figure 4.2). The study area is completely contained within the Acadian forest system and the southeast coast zone among the nine predominant climatic zones in Maine (Briggs and Lemin, 1992). The prevailing tree species in our study area are spruce species, in particular red spruce (*Picea rubens* Sargent), and balsam fir. Other conifer species are northern white cedar (*Thuja occidentalis* L.), and eastern hemlock [*Tsuga Canadensis* (L.) Carrière].

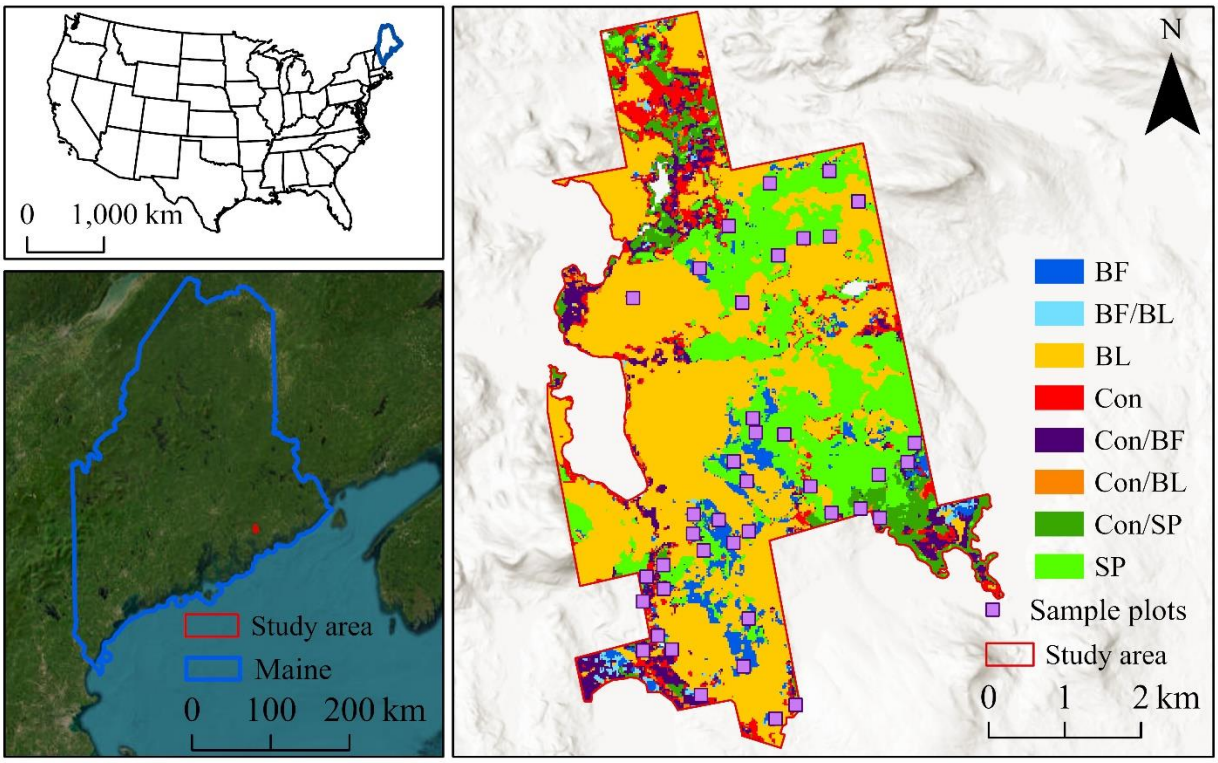

Figure 4. 2. Location of the study area and the distribution of sampling plots. Foliar samples were collected from 40 locations. Species composition map of the study area was adopted from Bhattarai et al. (2022a). BF: balsam fir; SP: spruce; BL: broadleaved species; Con: conifers other than spruce-fir; BF/BL: mixed balsam fir/broadleaved; Con/BF, Con/BL and Con/SP are mixed coniferous/ balsam fir, coniferous/broadleaved and coniferous/spruce, respectively.

4.2.2 Leaf sample collection, lab measurements and upscaling the foliar traits

A field campaign was carried out between July and August 2021. We established forty 20 x 20 m plots in the SBW host-dominated stands where we sampled needles. The central coordinates of each plot were recorded with sub-meter accuracy (after post processing) using a Trimble Geo 7x GPS (Sunnyvale, CA). Within each plot, diameter at breast height (DBH) was measured, and was used as proxy for species abundance in a plot for community weighting of foliar traits (explained in detail in section 2.3) before upscaling them to canopy level. In addition, tree heights were recorded to identify dominant and co-dominant trees in a plot, which were eventually used

as sample trees to collect needles. Furthermore, to measure leaf area index (LAI), we used an LAI-2200 TC plant canopy analyzer (LI-COR, Inc., Lincoln, NE) in each plot in order to upscale the leaf level measurements to canopy level. In case of a completely homogeneous plot, five trees from dominant and co-dominant categories (visible to the satellite sensor) were selected (one in the center and four in corners) to collect leaf samples; however, in a plot with mixed species, samples were collected from every dominant and co-dominant tree species in that plot. We collected at least one sunlit branch from each sampled tree. The upper sections of the canopies were accessed using a slingshot to collect sunlit foliar samples. The collected needles were wrapped with moist paper towels, placed in zip-lock polythene bags, and transferred to a laboratory for additional measurements.

We randomly selected approximately three grams from the multi-year needles from the SBW host tree branches (dominated by red spruce) of each plot and then recorded measurements of the fresh weight using an analytical (0.0001 g) digital scale, and their surface area using an AM350 leaf surface scanner (ADC BioScientific Ltd, Hoddesdon, UK). The needles were then dried in an oven at 65°C for approximately 72 hours, and their dry weights were measured. Based on these measurements, we calculated the foliar traits of SBW host species that represent the properties associated with mechanical defenses using Equation 1 and 2.

$$EWT\ (g\ cm^{-2}) = \frac{FW - DW}{SA} \qquad (1)$$

$$LMA\ (g\ cm^{-2}) = \frac{DW}{SA} \qquad (2)$$

where, FW is the fresh weight of needles, DW is the dry weight of needles and SA is the surface area of the fresh needles. In order to obtain an accurate surface area measurement, we corrected the surface area obtained after scanning the cylindrical spruce needles using the AM350 by

multiplying the values by 2.57, as recommended by Waring (1983). The dried needles were then carefully powdered using an electric grinder and sent to a chemistry lab for the estimation of our target nutritive and non-nutritive elements (Ca, Cu, Fe, K, N, and P).

Concentrations of Ca, K, N, and P in the foliar samples were given as percentages of the sample dry weight, which was then converted to area-based estimation (g cm^{-2}) by multiplying the percentage values with their respective dry mass per unit area (g cm^{-2}) (Gara et al., 2022). Cu and Fe estimates were received as mg kg^{-1} of the dry sample weight, which were then converted to area-based measurements (g cm^{-2}) by multiplying the values with their corresponding dry mass per unit area (g cm^{-2}) values.

4.2.3 Upscaling foliar traits to plot and canopy levels

Our sample plots were predominantly composed of red spruce; however, to account for the effects of other species on plot level trait expression, the mean canopy trait of each sample plot was derived by community weighting the trait values of all the species present in the plot by their relative basal area. We subsequently employed an area-based trait upscaling approach to upscale the community-weighted trait values to an individual plot. This approach involved using the leaf area index (LAI) values to upscale foliar traits from the leaf to the plot level, as described in Equation 3 (Homolova et al., 2013).

$$Trait_{Plot} = \left(\sum_{n=1}^{x} T_n \times BA_n \right) \times LAI \qquad (3)$$

where, $Trait_{Plot}$ is the community weighted average trait expression for a plot, x is the total number of species in a plot, T_n is the trait value for the species n, BA_n is the relative basal area of the species n in the particular plot and LAI is the leaf area index of that plot.

4.2.4 Satellite and site data acquisition and preprocessing

We downloaded a single cloud-free Sentinel-2A Level 1C Top-Of-Atmosphere (L1C-TOA) reflectance image (https://earthexplorer.usgs.gov/) for the peak growing season (July 7, 2021) corresponding to the date of field data collection. The L1C-TOA imagery was converted to surface reflectance products (L2A) using the Sent2Cor plugin within the Sentinel Application Platform (SNAP). L2A products were resampled to 20 m and used to derive several SVIs. The SVIs were particularly selected based on their sensitivity to canopy biochemistry. (see Bhattarai et al., 2022a; Bhattarai et al., 2022b).

We used a total of 36 remote sensing variables to predict several foliar traits. These included 27 SVIs (Table A.1) and nine spectral bands (Bhattarai et al., 2022a). The quality scene classification image, obtained after atmospherically correcting L1C-TOA images in SNAP, was used to mask out everything except for forested areas in the image. Next, a SBW host and non-host species mask for the state of Maine (Bhattarai et al. 2022a) was used to exclude all the non-host species from the analysis. Lastly, five site variables including aspect, elevation, depth to water table (DWT), slope, and topographic wetness index (TWI), all at 20 m spatial resolution, were integrated with Sentinel-2 variables to improve model performance.

4.2.5 Variable selection and machine learning algorithms

Three techniques were used for modelling foliar traits of SBW host species. Random forests (RF) is a non-parametric method that uses an ensemble of decision trees to make decisions (Breiman, 2001). The bootstrapping techniques used in RF to generate individual trees is widely utilized to normalize the potential effects of multicollinearity. Support vector machine (SVM) is also a non-parametric modeling algorithm known for its ability to resolve regression and classification problems by identifying an optimum hyperplane (Cortes and Vapnik, 1995).

Moreover, we evaluated the effectiveness of extreme gradient boosting (XGB) algorithm for modeling canopy traits. Similar to RF, XGB (Chen and Guestrin, 2016) is an algorithm which makes decisions using an ensemble of decision trees; however, the trees are not independent. Additional trees are built one after another in order to improve the weaknesses of the previous tree and improve the accuracy of the overall model.

We used Caret R package to execute all models (Kuhn, 2008). Model formulation and evaluation involved the implementation of a five-folded cross-validation technique, which was repeated five times to ensure the robustness of the model. Regarding the model cross validation, the entire training dataset was divided into five folds where four folds were used to train the model and the remaining fold for validation until all the folds were used for training and validation once. The final accuracy reported was obtained after averaging the accuracies of all the models formulated within the cross-validation framework. For each trait of interest, we used the optimum variables to obtain model performances using three different algorithms (RF, XGB and SVM).

Optimum variables for modeling each trait were determined in two steps: i) obtaining optimum remote sensing variables using VSURF algorithm, and ii) integrating site variables with the remote sensing variables obtained from the first step and evaluating different combinations based on the overall accuracy of the created models. VSURF R package (Genuer et al., 2010) was used to reduce the dimensionality of the variables and to identify the optimum variables for modeling the canopy traits. This approach helps to eliminate redundant or less informative variables in three steps: 1) eliminating irrelevant variables, 2) selecting all the variables contributing to prediction and 3) excluding redundant/correlated variables from the list of variables acquired from the second step (Bhattarai et al., 2022a).

The models' performances were evaluated using the normalized RMSE (nRMSE) (Bhattarai et al., 2022b) and their respective standard deviations (sd). RMSE was normalized with the range of the data (nRMSE = RMSE / Data range). Finally, we selected the best-performing algorithm to produce the final trait map for the study area.

4.3 Results

4.3.1 Selected variables and their importance evaluation

The variables used for modeling the eight traits obtained from VSURF algorithm are presented in Figure 4.3. A variable importance plot was not generated for the Ca estimation model as it was built using only one variable (B5). The optimum variables were subsequently used for modeling traits using three different algorithms (RF, XGB, and SVM) (Figure 4.4) and the importance of variables for the best model was ranked for the studied traits.

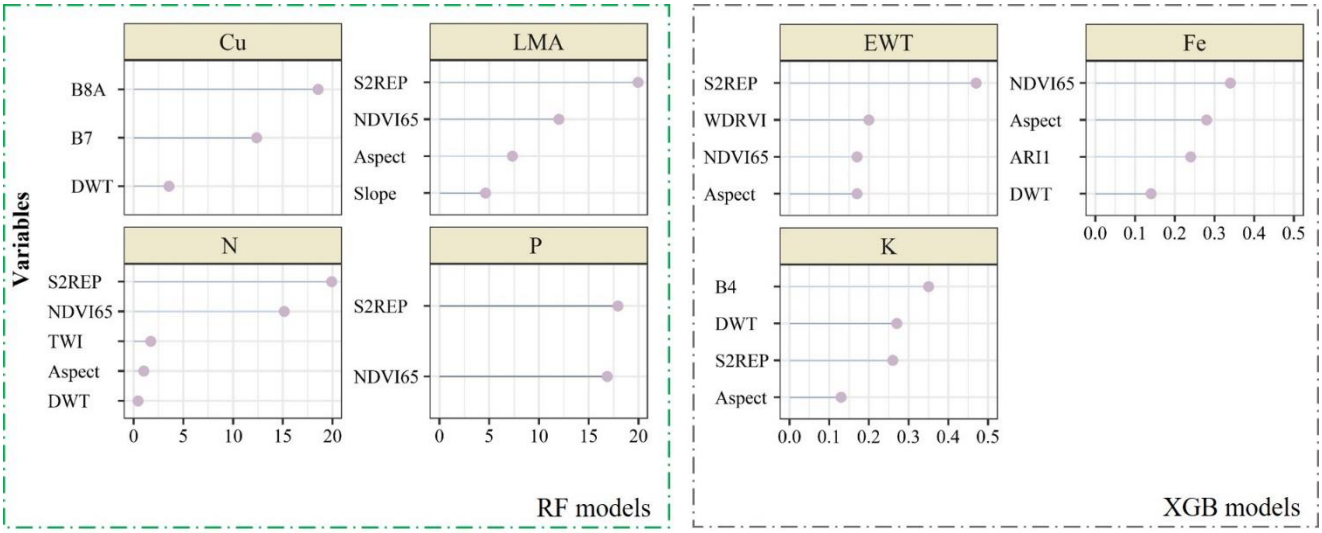

Figure 4. 3. Variables used for modeling different traits along with their overall importance (presented in the x-axis) in the model. Sentinel-2 Red Edge Position (S2REP); Normalized Difference Vegetation Index65 (NDVI65); Wide Dynamic Range Vegetation Index (WDRVI); Anthocyanin Reflectance Index1 (ARI1).

Evaluating the effectiveness of several variables used for modeling, it was evident that the red-edge variables were playing a pivotal role in terms of nutritional/non-nutritional and host species defensive trait estimations. It also demonstrates the potential of Sentinel-2 data for predicting the foliar traits in general. In particular, Normalized Difference Vegetation Index65 (NDVI65) was dominantly expressed for the estimation of five out of eight traits (EWT, Fe, LMA, N, and P) (Figure 4.3). Similarly, Sentinel-2 Red Edge Position (S2REP) was also one of the important variables for modeling five traits (EWT, LMA, N, P, and K). In addition, Anthocyanin Reflectance Index1 (ARI1) was picked by the VSURF algorithm as one of the important variables for estimating Fe, whereas Wide Dynamic Range Vegetation Index (WDRVI) was found useful for predicting EWT. Individual spectral bands also expressed themselves strongly for the prediction of several traits (B7 and B8A for Cu, B4 for K, and B5 for Ca) (Figure 4.3). Moreover, site variables, particularly DWT, slope, aspect, and TWI, contributed significantly to the prediction of foliar traits and were important in six of the eight trait models. P and Ca were the two traits for which models did not recognize site variables as important for modeling.

4.3.2 Assessing the performance of machine learning algorithms for traits estimation

The performances of different machine learning algorithms for all traits are presented in Figure 4.4. The grid search for all the targeted algorithms (RF, XGB, and SVM) was executed in R using the Caret package.

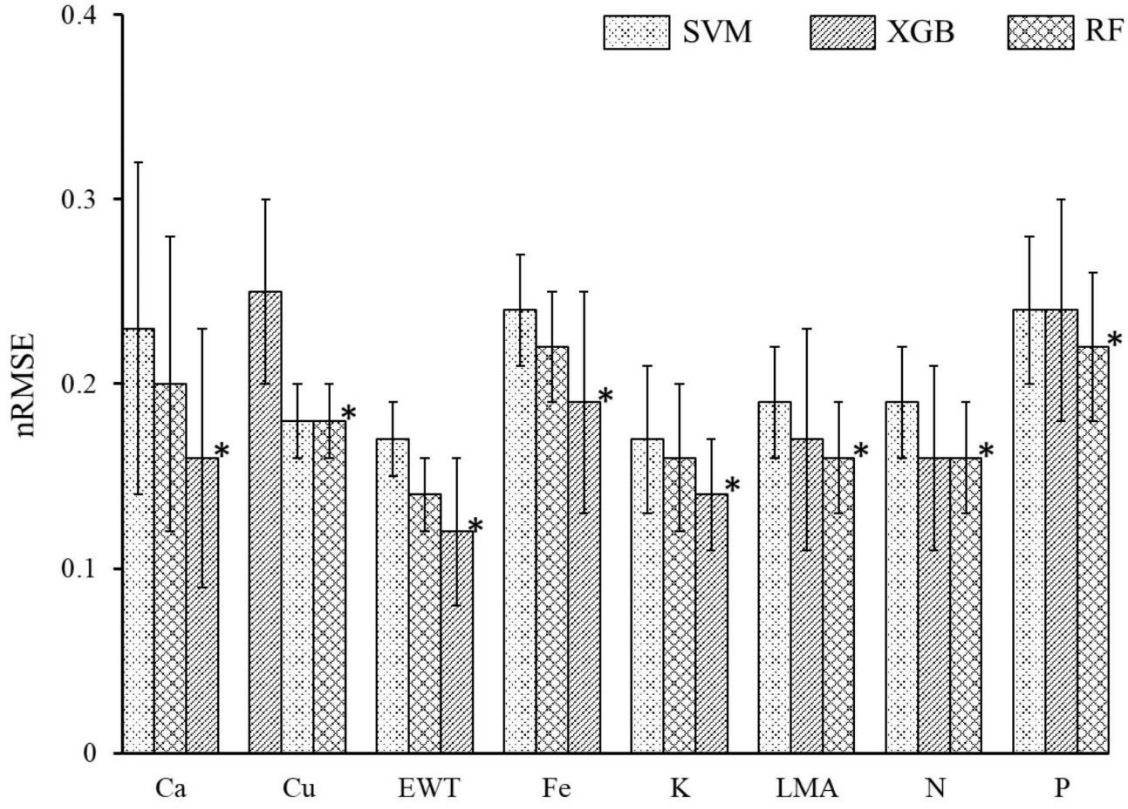

Figure 4. 4. Cross-validated nRMSE values obtained for modeling eight different plot level traits using RF, XGB, and SVM. The error bars represent one standard deviation of nRMSE. An asterisk adjacent to the corresponding bar indicates the final model used for prediction.

In general, SVM algorithm produced inferior results compared to RF and XGB algorithms for modeling all the foliar traits in our study. SVM and RF algorithms produced identical nRMSE value (0.18) for the Cu model (Figure 4.4). However, the RF algorithm was preferred over the SVM due to its lower sd. The RF algorithm outperformed the XGB algorithm for four trait models, Cu (nRMSE: 0.18), LMA (nRMSE: 0.16), N (nRMSE: 0.16), and P (nRMSE: 0.22) (Figure 4.4). On the other hand, XGB algorithm performed better for the remaining models, Ca (nRMSE: 0.16), EWT (nRMSE: 0.12), Fe (nRMSE: 0.19), and K (nRMSE: 0.14) (Figure 4.4). Of all the traits, EWT was predicted with the highest accuracy (nRMSE of 0.12 using XGB), while P was predicted

with the lowest accuracy (nRMSE of 0.22 using RF). Additional performance metrics, including R^2 and RMSE, for all the formulated models are presented in Table 4.1.

Table 4. 1. XGB, RF and SVM algorithms in estimating different canopy traits (see Figure 4.1 and Figure 4.3 for abbreviations). RMSE values for all the traits are in g m^{-2} except for Fe and Cu (mg m-2). Bold values are the metrics for the best performing models.

Traits (algorithm chosen)	**Algorithm**	**Variables used**	**R^2 (sd)**	**RMSE (sd)**	**nRMSE (sd)**
Ca (XGB)	**XGB**	B5	**0.47 (0.33)**	**0.66 (0.29)**	**0.16 (0.07)**
	RF		0.32(0.24)	0.85 (0.34)	0.20 (0.08)
	SVM		0.19 (0.13)	0.95 (0.36)	0.23 (0.09)
Cu (RF)	XGB	B8A, B7, DWT	NA (NA)	0.2 (0.038)	0.25 (0.05)
	RF		**0.42 (0.13)**	**0.15 (0.016)**	**0.18 (0.02)**
	SVM		0.42 (0.16)	0.15 (0.018)	0.18 (0.02)
EWT (XGB)	**XGB**	S2REP, WDRVI, NDVI65, aspect	**0.72 (0.12)**	**50.9 (15.1)**	**0.12 (0.04)**
	RF		0.58 (0.18)	59.34 (9.63)	0.14 (0.02)
	SVM		0.44 (0.20)	71.09 (11.34)	0.17 (0.02)
Fe (XGB)	**XGB**	NDVI65, ARI1, DWT, aspect	**0.48 (0.27)**	**2.6 (0.78)**	**0.19 (0.06)**
	RF		0.40 (0.17)	2.7 (0.38)	0.22 (0.03)
	SVM		0.17 (0.15)	3.3 (0.39)	0.24 (0.03)
LMA (RF)	XGB	NDVI65, S2REP, aspect, slope	0.59 (0.29)	57.83 (18.80)	0.17 (0.06)
	RF		**0.51 (0.19)**	**55.50 (10.58)**	**0.14 (0.02)**
	SVM		0.37 (0.19)	66.18 (10.22)	0.19 (0.03)
N (RF)	XGB	NDVI65, S2REP, DWT, aspect, TWI	0.69 (0.19)	0.52 (0.14)	0.16 (0.05)
	RF		**0.64 (0.01)**	**0.52 (0.08)**	**0.16 (0.03)**
	SVM		0.55 (0.14)	0.59 (0.08)	0.19 (0.03)
P (RF)	XGB	NDVI65, S2REP	0.50 (0.22)	0.09 (0.02)	0.24 (0.06)
	RF		**0.44 (0.14)**	**0.08 (0.01)**	**0.22 (0.04)**
	SVM		0.43 (0.14)	0.09 (0.02)	0.24 (0.04)
K (XGB)	**XGB**	B4, S2REP, DWT, aspect	**0.63 (0.28)**	**0.25 (0.05)**	**0.14 (0.02)**
	RF		0.50 (0.18)	0.30 (0.07)	0.16 (0.04)
	SVM		0.43 (0.23)	0.31 (0.06)	0.17 (0.04)

Prediction maps were made for all canopy traits of SBW host species using their respective best models (Figure 4.5). The predicted foliar traits ranged from 2 – 5 g m^{-2} for N, 0.2 – 0.5 g m^{-2} for P, 0.8 – 3 g m^{-2} for K, 0.8 – 4.8 g m^{-2} for Ca, 0.4 – 1 mg m^{-2} for Cu, 7 – 19 mg m^{-2} for Fe, 219

– 629 g m^{-2} for EWT and 171 – 560 g m^{-2} for LMA in the study area. It should be noted that these ranges were obtained for the SBW host stands from a coastal environment.

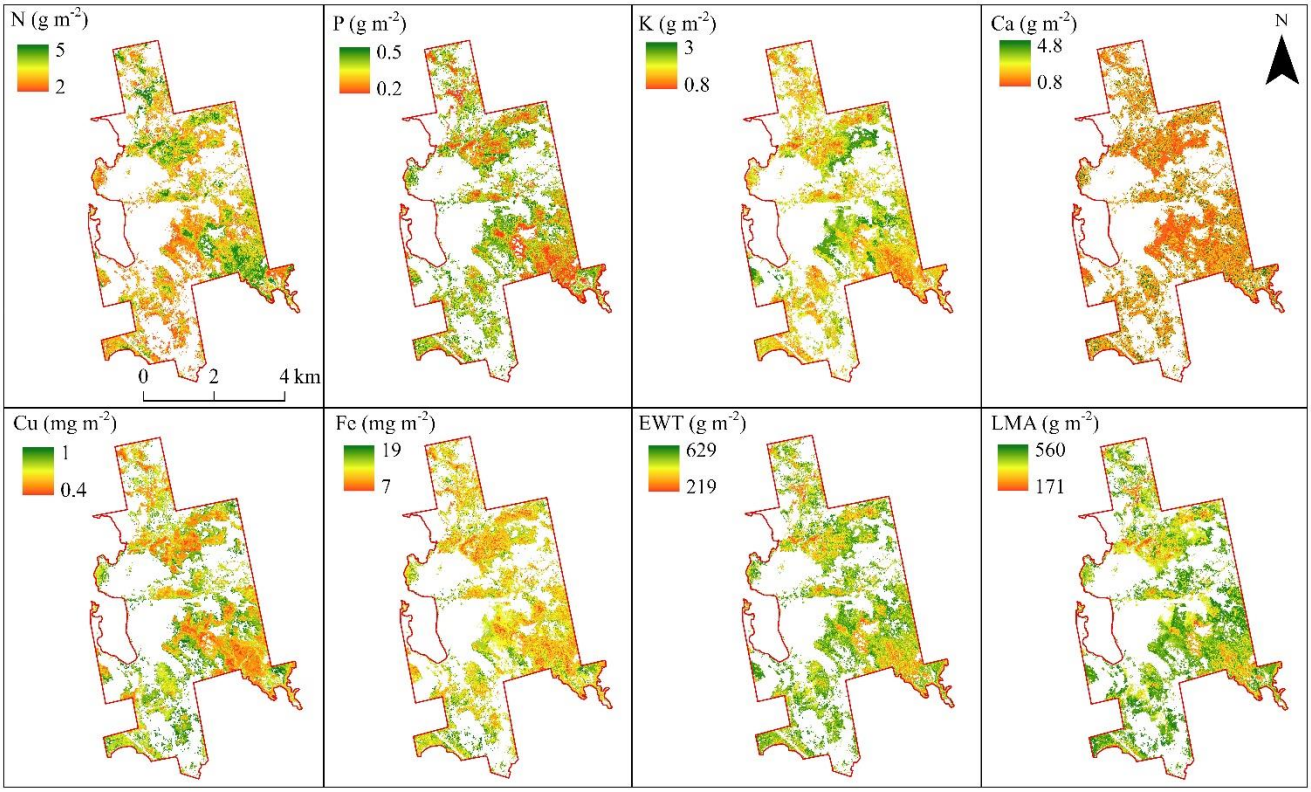

Figure 4. 5. SBW nutritional, non-nutritional and host species defense traits prediction maps at 20 m spatial resolution. The final models for Cu, LMA, N, and P were developed using random forest (RF) algorithm while the extreme gradient boosting (XGB) algorithm was employed to create the final model for Ca, EWT, Fe, and K.

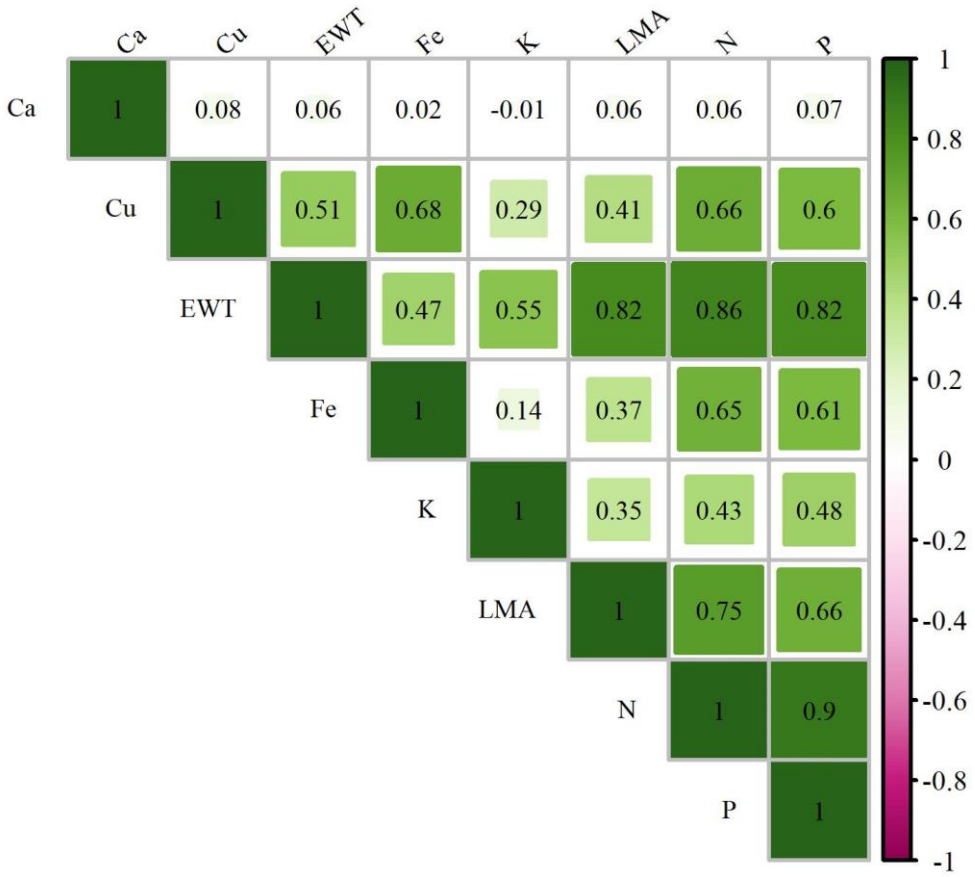

Figure 4. 6. Correlation between the predicted plot level (20 m spatial resolution) traits for the entire study area as demonstrated in Figure 4.5. It is important to mention that this relationship is not derived from the field data, but from the predicted maps depicted in Figure 4.5.

The degree of correlation between all the predicted trait maps obtained from our study area was evaluated (Figure 4.6). It is apparent that most of the traits are positively correlated with each other. However, Ca is the only trait not indicating significant positive relation with any of the traits under investigation. Similar to Ca, K also exhibit a weaker relationship with the rest of the predicted traits. In contrast, N, P, EWT, and LMA show a general moderate to high level of positive correlation with each other (N-P: 0.9; N-EWT: 0.86; N-LMA: 0.75; P-EWT: 0.82; P-LMA: 0.66; EWT-LMA:0.82) (Figure 4.6).

4.3.3 Implications of trait measurements in understanding host-pest interactions

Spatially explicit information on foliar traits can provide valuable insights into plant growth, productivity and health dynamics. By utilizing remote sensing data, the scope of application can be extended to a landscape level. However, the absence of threshold values for the concentration of canopy traits that significantly impact ecological processes, such as forest productivity or susceptibility to herbivory, hinders the effective use of produced canopy trait maps in real-world scenarios. In this study, we estimated eight different canopy traits from three categories in terms of their usefulness towards SBW success: 1) nutritive traits (N, P, K, and Cu), 2) non-nutritive traits (Fe and Ca), and 3) defensive morphological traits (LMA and EWT). Although, due to the constraints related to multi-temporal field data collection and experimental setup, the thresholds for host canopy traits associated with host species' susceptibility to SBW attack were not determined. We present a framework (Figure 4.7) that could be used to classify host species into different susceptibility classes (least susceptible, moderately susceptible, and highly susceptible) based on the estimated trait values. We used LMA (defensive trait), K (nutritive trait), and N (nutritive trait but favors defense) as an example, where LMA and N values are negatively correlated with the susceptibility of SBW host species while K is positively correlated. All the experimental plots proposed in our framework should be control (untreated) plots since any application of SBW suppression treatment between two measurements might influence the natural activities of the pest.

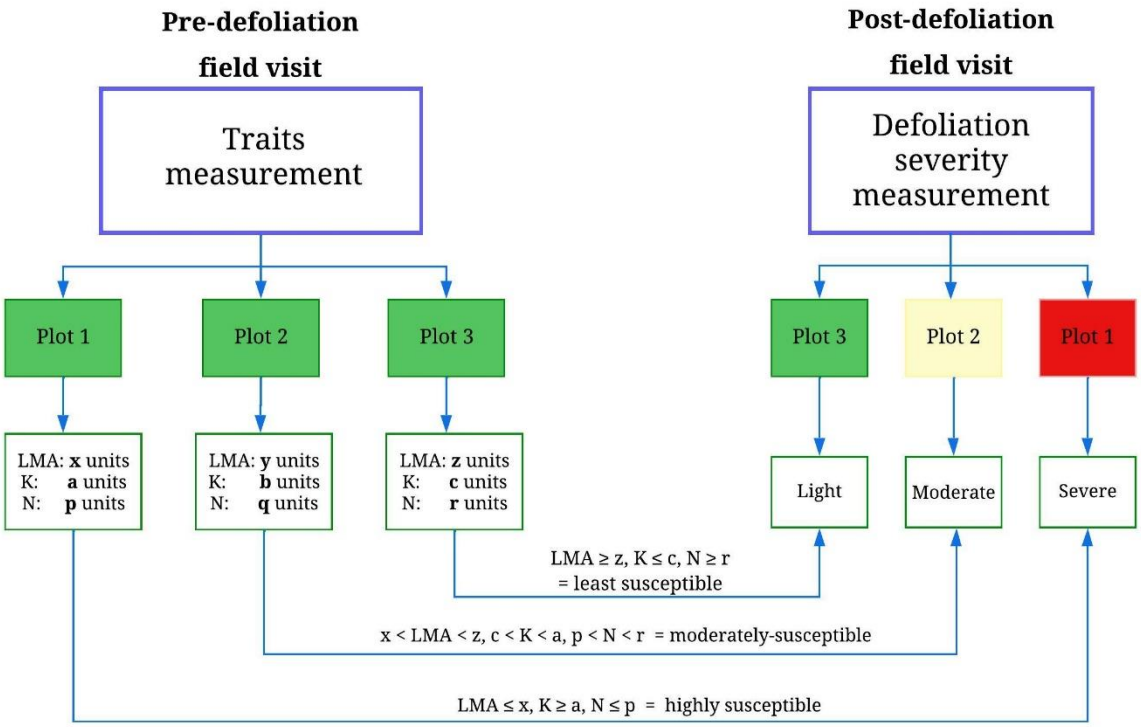

Figure 4. 7. Framework for linking plot-level trait values (such as plot LMA, K, and N) to the susceptibility of SBW host species against SBW attack. The tool involves conducting visits to the plots before and after infestation. During the initial visit, traits are measured, whereas the degree of defoliation is measured during the subsequent visit.

4.4. Discussion

This study investigated a new method of estimating plant canopy traits by using remote sensing and site data. Recently, there has been growing interest among researchers in estimating canopy foliar traits using this technique due to its ability to be an easy and cost-effective approach to evaluate plant health and productivity. However, to date, no studies have attempted to relate estimated foliar traits to a plants' ability to reduce herbivory. In fact, most studies focus only on estimations of very common plant pigments, such as chlorophyll and N, which are less valuable for assessing forest well-being and productivity from an ecological perspective. Our research

bridges two popular disciplines: foliar trait estimation and insect-plant interactions, especially in regard to traits associated with potential defensive mechanisms of plants. We also used modern machine learning algorithms to make predictions of foliar traits and to compare their accuracies for trait estimation.

4.4.1. Machine learning algorithms and their performances for traits mapping

Overall, our results indicate that XGB and RF algorithms outperformed the SVM algorithm for estimating targeted canopy traits using remote sensing and site variables (Figure 4.4; Table 4.1). However, it is worth noting that the SVM algorithm still achieved decent performance [nRMSE range: 0.17 (EWT) – 0.24 (Fe and P)]. We observed that XGB and RF were both suitable for estimating four trait models (XGB: Ca, EWT, Fe, and K; RF: Cu, LMA, N, and P). Nevertheless, XGB generally performed better than RF in terms of the R^2 values (Table 4.1). These findings are consistent with several other studies that have reported the performance of RF, XGB, and SVM algorithms for trait estimation. The study conducted by Miao et al. (2022) on estimating leaf nutrients of Mangrove during summer using Sentinel-2 variables noted a comparable accuracy between RF and XGB for N [RF: nRMSE (0.09) vs. XGB: nRMSE (0.10)] and a better performance of RF over XGB for P [RF: nRMSE (0.15) vs. XGB: nRMSE (0.16)]. Similarly, Prado-Osco et al. (2019) reported a comparable RMSE of 2.09 for RF and 2.04 for XGB in an attempt to estimate canopy N content of citrus trees using UAV imagery, whereas SVM exhibited inferior performance with an RMSE of 5.15.

In contrast, Zhang et al. (2022) reported better performance of XGB over RF for the estimation of foliar N [RF: nRMSE (0.47) vs. XGB: nRMSE (0.44)], P [RF: nRMSE (0.40) vs. XGB: nRMSE (0.29)] and specific leaf area [inverse of LMA; RF: nRMSE (0.41) vs. XGB: nRMSE (0.26)]. Nonetheless, it is worth noting that their study differs from ours in terms of the

study area location and vegetation types involved, as they focused on alpine meadows while our study estimated canopy traits of temperate forest trees. Regarding data modeling, Georganos et al. (2018) reported superior performance of XGB, followed by RF and SVM, with a small number of training samples (< 60) in their study on land use-land cover modeling, which is in agreement with our findings.

To our best understanding, no studies have reported on the estimation of Ca, Cu, EWT, Fe, and K using the XGB algorithm, which could be compared with RF and SVM. However, there are several instances where some of the aforementioned foliar traits have been measured using either a single algorithm or a different algorithm than the ones used in our study. When comparing the best models from the literature with the ones we created, our results are consistently comparable or even superior in terms of the estimation of N (Wang et al. (2018b): nRMSE (0.18); Gara et al. (2019): nRMSE (0.19); Zhang et al. (2022): nRMSE (0.22)), P (Axelsson et al. (2013): nRMSE (0.57); Zhang et al. (2022): nRMSE (0.20)), K (Axelsson et al. (2013): nRMSE (0.29); Chlus and Townsend (2022): nRMSE (0.17)), Ca (Axelsson et al. (2013): nRMSE (0.57); Chlus and Townsend (2022): nRMSE (0.26)), EWT (Liu et al. (2016): nRMSE (0.20)), and LMA (Gara et al. (2019): nRMSE (0.19); Chlus and Townsend (2022): nRMSE (0.17)). There were no records available in the literature for modeling canopy Fe and Cu using satellite imagery and machine learning algorithms to compare with our results.

4.4.2. Remote sensing and site variables for modeling canopy traits

Our study demonstrates the primary role of remote sensing variables in estimating a suite of diverse canopy traits; yet site variables were influential in all the of the models except for Ca and P. Focusing on the remote sensing variables, red-edge variables had a significant contribution in estimating the entire set of canopy traits. Our results on the selection of red-edge variables for

N estimation are consistent with several other studies (e.g., Gara et al., 2019; Prado-Osco et al., 2019; Miao et al., 2022). In addition, several researchers have demonstrated the sensitivity of NIR and SWIR wavelengths to N (e.g., Curran, 1989; Kokaly and Clark, 1999), with the absorption of the wavelengths attributed to, for example, N-Hydrogen (N-H) stretch (1510 nm), Oxygen-Hydrogen (O-H) stretch/O-H deformation (1940 nm), N-H bend (second overtone)/Carbon-N (C-N) stretch (2180 nm) etc.

Among the limited literature on the estimation of canopy P using remote sensing data, Milton et al. (1991) suggested the significance of red-edge wavelengths for estimating foliage P, which was consistent with our results; however, the study from Mutanga and Kumar, (2007) contradicts this finding. The authors claim that the wavebands from the SWIR region are better to estimate the P content in plant canopies. The inconsistency in which spectral bands should be used to predict canopy P content, plus the lower accuracy for the estimation (nRMSE: 0.22 in our study), may be attributed to the generally low P concentration in leaves. In addition, the absence of prominent P-bonded compounds sensitive to the range of electromagnetic spectrum that researchers generally work with (400 nm – 2500 nm) may be playing a role (Porder et al., 2005; Axelsson et al., 2013; Homolova et al., 2013).

Similar to the findings reported by Gara et al. (2022), the results from our study demonstrate the key role of red-edge variables and a visible band (B4) in estimating canopy K. Nevertheless, Asner et al. (2011) recommended a full spectrum of electromagnetic radiation from visible (400 nm) to SWIR (2500 nm) for the better estimation of canopy K. In general, K, similar to P, does not have prominent absorption compounds, which could affect the absorption throughout the visible to SWIR range, forcing us to approximate it directly (Axelsson et al., 2013). However, we found that it was possible to estimate it with decent accuracy (nRMSE: 0.14) by leveraging its

correlation with other compounds, particularly hydrocarbon compounds like cellulose, starch, sugar, etc., present in the leaves (Chlus and Townsend, 2022). Moreover, the same trajectory is followed by canopy Ca and Fe (Asner et al., 2011; Axelsson et al., 2013). It is crucial to underscore that the concentration of these elements (except for Cu) vary seasonally in the plant canopy, which also might impact the strength of their imprint in the range of electromagnetic spectrum we are working with (Chlus and Townsend, 2022).

As opposed to our expectation, the models for traits expressing canopy water (EWT), and dry matter/structure (LMA) content did not find SWIR-based variables important as proposed by Curan (1989). It is not uncommon to observe the inclusion of variables sensitive to LAI (red-edge indices in our case) for the estimation of EWT (Liu et al., 2016) and LMA (Gara et al., 2019) as witnessed in our study. Furthermore, the SWIR region of the electromagnetic spectrum is very sensitive to the dry matter content in leaves. Nevertheless, dry matter content can be masked by the water content when dealing with green foliage (Riano et al., 2005; Feret et al., 2019). Our results are similar to a study conducted by Liu et al. (2016) that found NDVI performed the best among all other SWIR variables when evaluating various hyperspectral indices for estimating EWT. Furthermore, the authors revealed the existence of a very high correlation (correlation coefficient of 0.94) between LAI and EWT in their study that could be attributed to the influential role of red-edge variables for water content estimation.

Similar to the remote sensing variables, as we expected, site variables, in particular, DWT, TWI, and aspect played a significant role in estimating spruce and fir foliar traits. Taking into account the literature available on traits estimation using remote sensing data, it is apparent that researchers have rarely incorporated site or environmental variables into their analyses. However, a recent research conducted by Gara et al. (2022) estimated the foliar traits of conifer species and

reported the important role of a site variable (DWT). Moreover, in support of our findings, Loozen et al. (2020) have emphasized the significance of site variables in estimating foliar nitrogen levels in European forests.

4.4.3 Correlation between traits and the limitations of proposed susceptibility framework

Based on our study, it is evident that there exists a significant correlation among the nutritive (N, P, K, and Cu), non-nutritive (Fe, and Ca), and defense (EWT, and LMA) foliar traits (Figure 4.6). Specifically, there were certain pairs exhibiting strong correlations: N and P (correlation coefficient: 0.9), EWT, and LMA (correlation coefficient: 0.82), EWT and N (correlation coefficient: 0.86), and EWT and P (correlation coefficient: 0.82). Given these strong correlations, it is possible to strategically select specific variables and exclude others for the sake of convenience in analysis. In addition, eliminating correlated traits can significantly reduce the cost of laboratory analysis.

The presented framework for linking canopy traits to host-species susceptibility is both simple and promising. However, it is important to acknowledge its limitations. One notable limitation is the current inability of the framework to account for the interaction effects between the nutritive, non-nutritive, and defensive traits in both the plant and insect system. This interaction is complex and needs further exploration in order to fully understand its implications. Moreover, the existing framework represents a relatively simplistic model, and there is room for improvement by incorporating additional trait variables. Expanding the range of traits considered, we could refine the model and enhance its robustness. Additionally, it is crucial to recognize that the coevolutionary relationship between herbivorous insects and host plant species has the potential to modify the nature and/or importance of the traits as a metric of susceptibility. This is especially true based on if the insect is native vs. non-native, or a generalist vs. a specialist. It is therefore

important to verify and update the framework regularly to account for any modifications resulting from these plant-insect interactions.

4.5 Conclusions

Our study modeled eight different nutritive, non-nutritive and defensive foliar traits in spruce-fir stands and compared model performances using remote sensing, site variables, and three different machine-learning algorithms. Among the three algorithms, XGB and RF exhibited better performance in overall traits estimation compared to SVM. In particular, we recommend the XGB algorithm over RF due to its superiority evaluating multiple accuracy metrics such as R^2 and RMSE. However, the accuracy of SVM was decent and could be tested in future for similar studies.

Remote sensing data, especially the red-edge variables, proved essential for estimating canopy traits of spruce-fir stands, which played a crucial role in estimating foliar traits from green foliage. However, additional investigations are required to validate the adequacy of red-edge bands in estimating a broader range of foliar traits. Additionally, site variables were found to be an integral part of the trait models and we strongly recommend future studies to incorporate site variables in their analyses to improve the accuracy of trait estimations.

Ultimately, we propose a novel framework for linking the estimated trait values with susceptibility of the host species against their pests that could be utilized by researchers to create host susceptibility maps using canopy trait values. We believe that this approach can be a complementary method to existing trait estimation tools. Additionally, trait maps could help forest managers make meaningful inferences at a landscape level regarding the health and productivity of their forestlands.

CHAPTER 5

INTEGRATION, CONCLUSIONS, AND FUTURE DIRECTIONS

5.1 Integration of spruce budworm (SBW) susceptibility assessment techniques

This study investigated several techniques to assess forest susceptibility to SBW as well as exploring product needs for SBW outbreak planning and management. The focus was on improving preparedness prior to actual forest defoliation and damage enabled by remote sensing technology. The common and widely adopted technique of evaluating pest susceptibility of a forest using host-composition layer has been reinforced in our study using the stand maturity information derived using Land Change Monitoring, Assessment, and Projection (LCMAP) products. Moreover, we also used several machine learning algorithms to model SBW host abundance in terms of leaf area index (LAI) and basal area per hectare (BAPH) in the areas where SBW host species are present. This provides us with in-depth information about the host species richness in the area mapped as host species by our host composition model. The host abundance in terms of leaf area could be useful for the forest community to derive information about the foliage density available for SBW to defoliate while basal area could be used as a proxy for wood materials or value under potential risk. Moving forward, our study further strengthens the susceptibility mapping technique incorporating the insect and host physiology and their interaction. Given that, the host species use several defense strategies against the pests, not all the attempts from pests to attack on hosts turn out to be successful and not all the individual host trees are equally susceptible. In this regard, the quantification of defensive, nutritive and non-nutritive traits present in SBW host species can aid in the comprehensive analysis of host species resilience against the pest.

While each of the three aforementioned techniques for assessing the susceptibility of SBW host species is individually robust enough, their collective use seemingly produces better

outcomes. The individual techniques could be sequentially used, with three distinct methods corresponding to three stages within a process. As illustrated in Figure 5.1, the initial step involves discriminating between host species and non-host species, followed by assessing the maturity of the identified host species. Moving to the second stage, the quantification of abundance through indicators such as LAI and BAPH takes place in the regions classified as host forests from the preceding step. Ultimately, in the final stage, host abundance is coupled with the evaluation of nutritive, non-nutritive and defense traits present in the host species, particularly foliage. It is expected that this integrated approach is comprehensive and is likely to yield a robust, reliable, and cost-effective alternative for conducting landscape-level assessments of forest susceptibility in general.

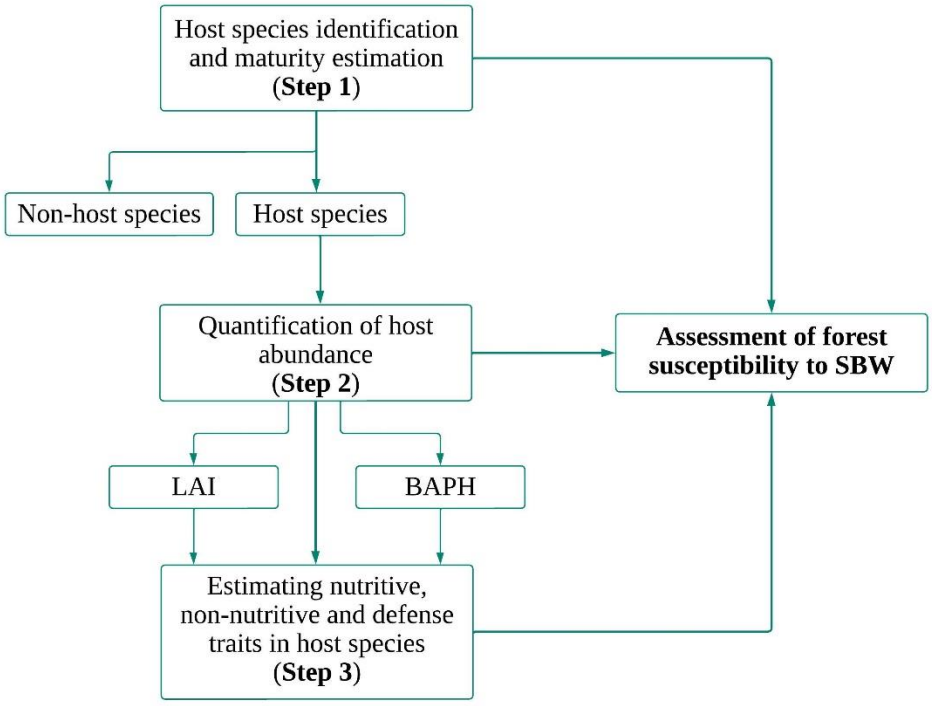

Figure 5. 1. Integrated approach for the estimation of forest susceptibility to SBW defoliation. LAI: leaf area index; BAPH: basal area per hectare.

5.2 Limitations and future directions

Monitoring of insect/pest induced disturbances, particularly SBW in our case is crucial from economical, ecological and societal points of view. Our study sought to devise various individual techniques to assess the susceptibility of landscapes to SBW which could be eventually combined to formulate a robust framework. The study largely focused on one of the major factors causing an outbreak, forest susceptibility where we modeled host species composition, their abundances and foliar traits. However, there are still areas where we can integrate scenarios such as phenology of insects and environmental factors like precipitation, spring temperatures, wind direction, etc. to better predict SBW outbreak at a landscape scale However, such studies require good quality historical SBW defoliation data that were not available for this research in particular with the spatial resolution used in this study (20 m). Nevertheless, our work provides a very helpful insight into monitoring and assessing the status of forested landscapes before the outbreak, which will aid landowners to formulate management plans against the potential future outbreaks.

Through this study, we introduced a novel framework to link canopy foliar traits to susceptibility of forests to pests (SBW in our case) which will put value on the predicted foliar traits for actual application in the real world. In addition, we suggest the use of freely available archives of LCMAP data for estimating the age of forest stands based on their promise from our study.

Throughout this study, several machine learning algorithms such as random forest (RF) support vector machine (SVM), artificial neural networks (ANNs), and extreme gradient boosting (XGB) were used to model diverse forest susceptibility indicators, namely, forest composition, abundance (LAI, and BAPH) and foliar traits (nutritive, non-nutritive, and defense). RF was often found to outperform all other algorithms. However, XGB marked its superiority along with RF for

some foliar traits estimation such as calcium (Ca), equivalent water thickness (EWT), iron (Fe), and potassium (K). In this entire study, Sentinel-2 data, in particular, the red-edge spectral vegetation indices derived from it proved to be the essential remote sensing data for estimating all three host susceptibility indicators. In addition, synthetic aperture radar (SAR), to be specific, L-band SAR data added some benefits during the development of forest composition models (SBW host species included) when integrated with optical data. Furthermore, site variables were pivotal in all the modelling approaches.

Together with numerous advantages, remote sensing techniques come with several limitations for forest health monitoring. Remote sensing techniques rely heavily on the quality of available satellite imagery, which can be affected by environmental factors such as cloud cover and atmospheric conditions, and the resolution (temporal, spectral, and spatial) of sensor itself. With the recent continuous advancements in remote sensing technologies, these challenges could be overcome by fusing multiple sensor types and data as well as taking advantage of artificial intelligence (AI) and big data leading to more efficient and cost-effective forest health monitoring.

REFERENCES

Abdullah, H., Skidmore, A.K., Darvishzadeh, R. and Heurich, M. (2019). Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (Ips typographus, L.) compared with Landsat-8. Remote sensing in ecology and conservation, 5(1), 87-106.

Ahmadi, K., Kalantar, B., Saeidi, V., Harandi, E.K., Janizadeh, S. and Ueda, N. (2020). Comparison of machine learning methods for mapping the stand characteristics of temperate forests Using Multi-Spectral Sentinel-2 Data. Remote Sensing, 12, 3019.

Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C., Carranza, L., Martinez, P., Houcheime, M., Sinca, F. and Weiss, P. (2011). Spectroscopy of canopy chemicals in humid tropical forests. Remote Sensing of Environment, 115, 3587-3598.

Astola, H., Häme, T., Sirro, L., Molinier, M. and Kilpi, J. (2019). Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sensing of Environment, 223, 257-273.

Attarchi, S., and Gloaguen, R. (2014). Classifying complex mountainous forests with L-Band SAR and Landsat data integration: a comparison among different machine learning methods in the Hyrcanian forest. Remote Sensing, 6(5), 3624-3647.

Axelsson, C., Skidmore, A.K., Schlerf, M., Fauzi, A. and Verhoef, W. (2013). Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. International Journal of Remote Sensing, 34(5), 1724-1743.

Bhattarai, Rajeev, "Spruce Budworm Defoliation Detection and Host Species Mapping Using Sentinel Satellite Imagery" (2020). Electronic Theses and Dissertations. 3306.

Bhattarai, R., Rahimzadeh-Bajgiran, P. and Mech, A. (2023). Estimating nutritive, non-nutritive and defense foliar traits in spruce-fir stands using remote sensing and site data. Forest Ecology and Management, 549, 121461.

Bhattarai, R., Rahimzadeh-Bajgiran, P. and Weiskittel, A. (2022a). Multi-Source Mapping of Forest Susceptibility to Spruce Budworm Defoliation Based on Stand Age and Composition across a Complex Landscape in Maine, USA. Canadian Journal of Remote Sensing, 48(6), 873-893.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Homayouni, S., Gara, T.W. and Hanavan, R.P. (2022b). Estimating species-specific leaf area index and basal area using optical and SAR remote sensing data in Acadian mixed spruce-fir forests, USA. International Journal of Applied Earth Observation and Geoinformation, 108, 102727.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., and MacLean, D.A. (2020). Sentinel-2 based prediction of spruce budworm defoliation using red-edge spectral vegetation indices. Remote Sensing Letters, 11(8), 777-786.

Bhattarai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Meneghini, A. and MacLean, D.A. (2021). Spruce budworm tree host species distribution and abundance mapping using multi-temporal Sentinel-1 and Sentinel-2 satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 172, 28-40.

Biswas, S., Huang, Q., Anand, A., Mon, M.S., Arnold, F.E., and Leimgruber, P. (2020). A multi sensor approach to forest type mapping for advancing monitoring of sustainable development goals (SDG) in Myanmar. Remote Sensing, 12(19), 3220.

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., Bonnet, and S. (2018). Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. Biotechnologie, Agronomie, Société et Environnement, 22(3), 16.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Briggs, R.D. and Lemin, R.C. Jr. (1992). Delineation of climatic regions in Maine. Canadian Journal of Forest Research, 22(6), 801–811.

Broge, N.H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sensing of Environment, 76, 156-172.

Butler, B.J. (2018). Forests of Maine, 2017. Resource Update FS-160. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station. 3p.

Chen, C., Rahimzadeh-Bajgiran, P. and Weiskittel, A. (2021). Assessing spatial and temporal dynamics of a spruce budworm outbreak across the complex forested landscape of Maine, USA. Annals of Forest Science, 78(2), 1-14.

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. Canadian Journal of Remote Sensing, 22, 229-242.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Chlus, A. and Townsend, P.A. (2022). Characterizing seasonal variation in foliar biochemistry with airborne imaging spectroscopy. Remote Sensing of Environment, 275, 113023.

Clevers, J., De Jong, S., Epema, G., Addink, E., Van Der Meer, F. and Skidmore, A. (2000). Meris and the Red-edge index. Paper presented at the Second EARSeL workshop on Imaging spectroscopy, EARSeL, Enschede, pp. 1-16.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273-297.

Cui, Z. and Kerekes, J.P. (2018). Potential of red edge spectral bands in future Landsat satellites on agroecosystem canopy green leaf area index retrieval. Remote Sensing, 10, 1458.

Curran, P.J. (1989). Remote sensing of foliar chemistry. Remote Sensing of Environment, 30(3), 271-278.

Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O'Connor, B. and Paganini, M. (2019a). Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. International Journal of Applied Earth Observation and Geoinformation, 79, 58-70.

Darvishzadeh, R., Wang, T., Skidmore, A., Vrieling, A., O'Connor, B., Gara, T.W., Ens, B.J. and Paganini, M. (2019b). Analysis of Sentinel-2 and RapidEye for retrieval of leaf area index in a saltmarsh using a Radiative Transfer Model. Remote Sensing, 11, 671.

Dash, J., and Curran, P. (2007). Evaluation of the MERIS terrestrial chlorophyll index (MTCI). Advances in Space Research, Vol. 39: pp. 100-104.

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., De Colstoun, E.B., and McMurtrey Iii, J.E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sensing of Environment, 74, 229-239.

Delegido, J., Verrelst, J., Alonso, L. and Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors, 11, 7063-7081.

Deo, R.K., Russell, M.B., Domke, G.M., Andersen, H.E., Cohen, W.B., and Woodall, C.W. (2017). Evaluating site-specific and generic spatial models of aboveground forest biomass based on Landsat time-series and LiDAR strip samples in the Eastern USA. Remote Sensing, 9 (6), 598.

DeRose, J.R. and Seymour, R.S. (2010). Patterns of leaf area index during stand development in even-aged balsam fir – red spruce stands. Canadian Journal of Forest Research, 40, 629-637.

Donovan, S.D., MacLean, D.A., Zhang, Y., Lavigne, M.B. and Kershaw, J.A. (2021). Evaluating annual spruce budworm defoliation using change detection of vegetation indices calculated from satellite hyperspectral imagery. Remote Sensing of Environment, 253, 112204.

Dostálová, A., Lang, M., Ivanovs, J., Waser, L.T., and Wagner, W. (2021). European Wide Forest Classification Based on Sentinel-1 Data. Remote Sensing, 13(3), 337.

Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D. and Dalu, T. (2019). Numerical assessments of leaf area index in tropical savanna rangelands, South Africa using Landsat 8 OLI derived metrics and in-situ measurements. Remote Sensing, 11, 829.

Erinjery, J.J., Singh, M., and Kent, R. (2018). Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. Remote Sensing of Environment, 216, 345-354.

Fang, H., Baret, F., Plummer, S. and Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. Reviews of Geophysics, 57, 739-799.

Fassnacht, F.E., Latifi, H., Ghosh, A., Joshi, P.K. and Koch, B., (2014). Assessing the potential of hyperspectral imagery to map bark beetle-induced tree mortality. Remote Sensing of Environment, 140, 533-548.

Fassnacht, K.S., Gower, S.T., Norman, J.M. and McMurtric, R.E. (1994). A comparison of optical and direct methods for estimating foliage surface area index in forests. Agricultural and Forest Meteorology, 71, 183-207.

Féret, J.B., Le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J.C., Ponzoni, F.J., Solanki, T. and De Boissieu, F. (2019). Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. Remote Sensing of Environment, 231, 110959.

Filgueiras, R., Mantovani, E.C., Althoff, D., Fernandes Filho, E.I., and Cunha, F.F.D. (2019). Crop NDVI monitoring based on sentinel 1. Remote Sensing, 11(12), 1441.

Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. Paper presented at the Multidisciplinary Digital Publishing Institute Proceedings, 18(1), 11.

Foody, G.M. (2004). Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. International Journal of Remote Sensing, 25, 3091-3104.

Fuentealba, A. and Bauce, É. (2016). Interspecific variation in resistance of two host tree species to spruce budworm. Acta Oecologica, 70, 10-20.

Fuentealba, A., Sagne, S., Legendre, G., Pureswaran, D., Bauce, É. and Despland, E. (2020). Leaf toughness as a mechanism of defence against spruce budworm. Arthropod-Plant Interactions, 14, 481-489.

Furniss, J., Rahimzadeh-Bajgiran, P., Gara, T.W., Daigle, J., and Costanza, K.K. (2022). Mapping ash species across a mixed forest using hyperspectral imagery. Remote Sensing Letters, 13(5), 441-451.

Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T. and Heurich, M. (2019). Accurate modelling of canopy traits from seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits. ISPRS Journal of Photogrammetry and Remote Sensing, 157, 108-123.

Gara, T.W., Rahimzadeh-Bajgiran, P. and Weiskittel, A. (2022). Determination of foliar traits in an ecologically distinct conifer species in Maine using Sentinel-2 imagery and site

variables: Assessing the effect of leaf trait expression and upscaling approach on prediction accuracy. ISPRS Journal of Photogrammetry and Remote Sensing, 193, 150-163.

Genuer, R., Poggi, J.M., and Tuleau-Malot, C. (2015). VSURF: an R package for variable selection using random forests. The R journal, 7(2), 19–33.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M. and Wolff, E. (2018). Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. IEEE geoscience and remote sensing letters, 15(4), 607-611.

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment, 58, 289-298.

Gitelson, A. A., Merzlyak, M. N., and Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochemistry and Photobiology, 74, 38-45.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., and Merzlyak, M. N. (2002). Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. Photochemistry Photobiology, 75, 272-281.

Gitelson, A., and Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. Journal of Photochemistry Photobiology B: Biology, 22, 247-252.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., and Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophysical research letters, 30, 1248

Grabska, E., Frantz, D., and Ostapowicz, K. (2020). Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sensing of Environment, 251, 112103.

Grabska, E., Hostert, P., Pflugmacher, D., and Ostapowicz, K. (2019). Forest stand species mapping using the Sentinel-2 time series. Remote Sensing, 11(10), 1197.

Guyot, G., and Baret, F. (1988). Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux. Paper presented in 4th International Colloquium on Spectral Signatures of Objects in Remote Sensing, Assois, France, 287: pp.279-286.

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote sensing of environment, 81, 416-426.

Hanavan, R.P., Kamoske, A.G., Schaaf, A.N., Eager, T., Fisk, H., Ellenwood, J., Warren, K., Asaro, C., Vanderbilt, B., Hutten, K. and Leatherman, L. (2022). Supplementing the forest health national aerial survey program with remote sensing during the COVID-19

pandemic: Lessons learned from a collaborative approach. Journal of Forestry, 120(2), 125-132.

Hardisky, M., Klemas, V., and Smart, M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora*. Photogrametric Engineering and Remote Sensing, 49, 77-83.

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., and Carron, J. (2019). Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. Sensors, 19(6), 1284.

Hennigar, C. R., Weiskittel, A., Allen, H.L. and MacLean, D.A. (2017). Development and evaluation of a biomass increment based index for site productivity. Canadian Journal of Forest Research, 47(3),400-410.

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

Hennigar, C.R., Wilson, J.S., MacLean, D.A., and Wagner, R.G. (2011). Applying a spruce budworm decision support system to Maine: Projecting spruce-fir volume impacts under alternative management and outbreak scenarios. Journal of Forestry, 109(6), 332-342.

Herms, D. A. and Mattson, W. J. (1992). The dilemma of plants: to grow or defend. The Quarterly Review of Biology, 67(3), 283-335.

Holmgren, J., and Persson, Å. (2004). Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment, 90 (4), 415-423.

Homolová, L., Malenovský, Z., Clevers, J.G., García-Santos, G. and Schaepman, M.E. (2013). Review of optical-based remote sensing for plant trait mapping. Ecological Complexity, 15, 1-16.

Hościło, A., and Lewandowska, A. (2019). Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. Remote Sensing, 11(8), 929.

Huber, S., Kneubühler, M., Psomas, A., Itten, K. and Zimmermann, N.E. (2008). Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. Forest Ecology and Management, 256(3), 491-501.

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25, 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83, 195-213.

Immitzer, M., Böck, S., Einzmann, K., Vuolo, F., Pinnel, N., Wallner, and A. and Atzberger, C. (2018). Fractional cover mapping of spruce and pine at 1 ha resolution combining very high and medium spatial resolution satellite imagery. Remote Sensing of Environment, 204, 690-703.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., and Atzberger, C. (2019). Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data. Remote Sensing, 11(22), 2599.

Jain, A.K., Mao, J. and Mohiuddin, K.M. (1996). Artificial neural networks: A tutorial. Computer, 29, 31-44.

Jones, H.G., and Vaughan, R.A. (2010). Remote Sensing of Vegetation: Principles, Techniques, and Applications. Oxford University Press.

Kaufman, Y.J., and Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE transactions on Geoscience and Remote Sensing, 30, 261-270.

Key, C., Benson, N., Ohlen, D., Howard, S., and Zhu, Z. (2002). The normalized burn ratio and relationships to burn severity: ecology, remote sensing and implementation. Paper presented at the 9th biennial remote sensing applications conference, San Diego, CA.

Kokaly, R.F. and Clark, R.N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sensing of Environment, 67(3), 267-287.

Korhonen, L., Packalen, P. and Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat-8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sensing of Environment, 195, 259-274.

Kuehne, C., Russell, M.B., Weiskittel, A.R., Kershaw Jr, and J.A. (2020). Comparing strategies for representing individual-tree secondary growth in mixed-species stands in the Acadian Forest region. Forest Ecology and Management, 459, 117823.

Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of Statistical Software, 28, 1-26.

Lardeux, C., Frison, P.L., Tison, C., Souyris, J.C., Stoll, B., Fruneau, B. and Rudant, J.P. (2010). Classification of tropical vegetation using multifrequency partial SAR polarimetry. IEEE Geoscience and Remote Sensing Letters, 8(1), 133-137.

Legaard, K., Simons-Legaard, E. and Weiskittel, A. (2020). Multi-objective support vector regression reduces systematic error in moderate resolution maps of tree species abundance. Remote Sensing, 12(11), 1739.

Li, G., Lu, D., Moran, E., Dutra, L., and Batistella, M. (2012). A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. ISPRS Journal of Photogrammetry and Remote Sensing, 70, 26-38.

Liesenberg, V., and Gloaguen, R. (2013). Evaluating SAR polarization modes at L-band for forest classification purposes in Eastern Amazon, Brazil. International Journal of Applied Earth Observation and Geoinformation, 21, 122-135.

Liu, L., Zhang, S. and Zhang, B. (2016). Evaluation of hyperspectral indices for retrieval of canopy equivalent water thickness and gravimetric water content. International Journal of Remote Sensing, 37(14), 3384-3399.

Loozen, Y., Rebel, K.T., de Jong, S.M., Lu, M., Ollinger, S.V., Wassen, M.J. and Karssenberg, D. (2020). Mapping canopy nitrogen in European forests using remote sensing and environmental variables with the random forests method. Remote Sensing of Environment, 247, 111933.

Ma, X., Mahecha, M.D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S., Kattge, J., Richter, R., Musavi, T., Baeten, L. and Barnoaiea, I. (2019). Inferring plant functional diversity from space: the potential of Sentinel-2. Remote Sensing of Environment, 233, 111368.

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

MacLean, D.A. and MacKinnon, W.E. (1996). Accuracy of aerial sketch-mapping estimates of spruce budworm defoliation in New Brunswick. Canadian Journal of Forest Research, 26(12), 2099-2108.

Majasalmi, T. and Rautiainen, M. (2016). The potential of Sentinel-2 data for estimating biophysical variables in a boreal forest: a simulation study. Remote Sensing Letters, 7, 427-436.

Maltamo, M., Kinnunen, H., Kangas, A., and Korhonen, L. (2020). Predicting stand age in managed forests using National Forest Inventory field data and airborne laser scanning. Forest Ecosystems, 7(44), 1-11.

Mananze, S., Pôças, I. and Cunha, M., 2018. Retrieval of maize leaf area index using hyperspectral and multispectral data. Remote Sensing, 10, 1942.

Manninen, T., Stenberg, P., Rautiainen, M., Voipio, P. and Smolander, H. (2005). Leaf area index estimation of boreal forest using ENVISAT ASAR. IEEE Transactions on Geoscience and Remote Sensing, 43, 2627-2635.

Mattson, W.J., Haack, R.A., Lawrence, R.K. and Slocum, S.S. (1991). Considering the nutritional ecology of the spruce budworm in its management. Forest Ecology and Management, 39, 183-210.

Mensah, A.A., Petersson, H., Saarela, S., Goude, M. and Holmström, E. (2020). Using heterogeneity indices to adjust basal area–leaf area index relationship in managed coniferous stands. Forest Ecology and Management, 458, 117699.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., and Rakitin, V.Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum, 106, 135–141.

Meyer, L.H., Heurich, M., Beudert, B., Premier, J. and Pflugmacher, D. (2019). Comparison of Landsat-8 and Sentinel-2 data for estimation of leaf area index in temperate forests. Remote Sensing, 11, 1160.

Miao, J., Zhen, J., Wang, J., Zhao, D., Jiang, X., Shen, Z., Gao, C. and Wu, G. (2022). Mapping Seasonal Leaf Nutrients of Mangrove with Sentinel-2 Images and XGBoost Method. Remote Sensing, 14(15), 3679.

Milton, N.M., Eiswerth, B.A. and Ager, C.M. (1991). Effect of phosphorus deficiency on spectral reflectance and morphology of soybean plants. Remote Sensing of Environment, 36(2), 121-127.

Ministère des Forêts de la Faune et des Parcs. (2020). Aires infestées par la tordeuse des bourgeons de l'épinette au Québec en 2020. Québec, Gouvernement du Québec, Direction de la protection des forêts, 31.

Mutanga, O. and Kumar, L. (2007). Estimating and mapping grass phosphorus concentration in an African savanna using hyperspectral image data. International Journal of Remote Sensing, 28(21), 4897-4911.

Neinavaz, E., Darvishzadeh, R., Skidmore, A.K. and Abdullah, H. (2019). Integration of landsat-8 thermal and visible-short wave infrared data for improving prediction accuracy of forest leaf area index. Remote sensing, 11, 390.

Persson, M., Lindberg, E., and Reese, H. (2018). Tree species classification with multi-temporal Sentinel-2 data. Remote Sensing, 10, 1794.

Porder, S., Asner, G.P. and Vitousek, P.M. (2005). Ground-based and remotely sensed nutrient availability across a tropical landscape. Proceedings of the National Academy of Sciences, 102(31), 10909-10912.

Prado Osco, L., Marques Ramos, A.P., Roberto Pereira, D., Akemi Saito Moriya, É., Nobuhiro Imai, N., Takashi Matsubara, E., Estrabis, N., de Souza, M., Marcato Junior, J., Gonçalves, W.N. and Li, J. (2019). Predicting canopy N content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. Remote Sensing, 11(24), 2925.

Puhlick, J.J., Weiskittel, A.R., Kenefic, L.S., Woodall, C.W., and Fernandez, I.J. (2020). Strategies for enhancing long-term carbon sequestration in mixed-species, naturally regenerated Northern temperate forests. Carbon Management, 11(4), 381-397.

Qiao, K., Zhu, W., Xie, Z. and Li, P. (2019). Estimating the seasonal dynamics of the leaf area index using piecewise LAI-VI relationships based on phenophases. Remote Sensing, 11, 689.

Raczko, E., and Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50 (1), 144-154.

Rahimzadeh-Bajgiran, P., Hennigar, C., Weiskittel, A. and Lamb, S. (2020). Forest potential productivity mapping by linking remote-sensing-derived metrics to site variables. Remote Sensing, 12, 2056.

Rahimzadeh-Bajgiran, P., Weiskittel, A. R., Kneeshaw, D. and MacLean, D. A. (2018). Detection of annual spruce budworm defoliation and severity classification using Landsat imagery. Forests. 9(6), 357

Riaño, D., Vaughan, P., Chuvieco, E., Zarco-Tejada, P.J. and Ustin, S.L. (2005). Estimation of fuel moisture content by inversion of radiative transfer models to simulate equivalent water thickness and dry matter content: analysis at leaf and canopy level. IEEE Transactions on Geoscience and Remote Sensing, 43(4), 819-826

Riedler, B., Pernkopf, L., Strasser, T., Lang, S., and Smith, G. (2015). A composite indicator for assessing habitat quality of riparian forests derived from Earth observation data. International Journal of Applied Earth Observation and Geoinformation, 37, 114-123.

Rignot, E.J., Williams, C.L., Way, J. and Viereck, L.A. (1994). Mapping of forest types in Alaskan boreal forests using SAR imagery. IEEE Transactions on Geoscience and Remote Sensing, 32(5), 1051-1059.

Riihimäki, J., Kaitaniemi, P., Koricheva, J. and Vehviläinen, H. (2005). Testing the enemies hypothesis in forest stands: the important role of tree species composition. Oecologia, 142, 90-97.

Rose, A. H. and Lindquist, O. H. (1994). Insects of eastern spruces, fir and hemlock. Natural Resource Canada, Canadian Forest Service, Science and Sustainable Development Directorate, Ottawa, Ontario.

Rouse Jr, J. W., Haas, R., Schell, J. and Deering, D.W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA Special Publication, NASA SP-351: pp. 309-317.

Rowe, J. S. (1972). Forest regions of Canada. Publication No. 1300, Canadian Forestry Service, Department of the Environment, Ottawa, Ontario.

Running, S.W. (1992). A bottom-up evolution of terrestrial ecosystem modeling theory, and ideas toward global vegetation modeling. In Modeling the Earth System (D. Ojima, Ed.), UCAR/Office for Interdisciplinary Earth Studies, Boulder, CO, 263–280.

Sevgen, E., Kocaman, S., Nefeslioglu, H.A. and Gokceoglu, C. (2019). A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression, ANN and random forest. Sensors, 19(18), 3940.

Shamsoddini, A. (2012). Radar backscatter and optical textural indices fusion for pine plantation structure mapping. ISPRS Annals of Photogrammetry, Remote Sensing and the Spatial Information Sciences, 7, 309-314.

Soleimannejad, L., Ullah, S., Abedi, R., Dees, M. and Koch, B. (2019). Evaluating the potential of Sentinel-2, Landsat-8, and IRS satellite images in tree species classification of Hyrcanian forest of Iran using random forest. Journal of Sustainable Forestry, 38(7) 615-628.

Sprintsin, M., Karnieli, A., Berliner, P., Rotenberg, E., Yakir, D. and Cohen, S. (2007). The effect of spatial resolution on the accuracy of leaf area index estimation for a forest planted in the desert transition zone. Remote Sensing of Environment, 109, 416-428.

Stankevich, S.A., Kozlova, A.A., Piestova, I.O. and Lubskyi, M.S. (2017). Leaf area index estimation of forest using Sentinel-1 C-band SAR data. In 2017 IEEE Microwaves, Radar and Remote Sensing Symposium (MRRS), Kiev, Ukraine, 29-31 August 2017, 253-256.

Teets, A., Fraver, S., Weiskittel, A.R. and Hollinger, D.Y. (2018). Quantifying climate–growth relationships at the stand level in a mature mixed-species conifer forest. Global Change Biology, 24(8), 3587-3602.

Terhonen, E., Blumenstein, K., Kovalchuk, A. and Asiegbu, F.O. (2019). Forest tree microbiomes and associated fungal endophytes: Functional roles and impact on forest health. Forests, 10(1), 42.

Thompson, S.D., Nelson, T.A., White, J.C. and Wulder, M.A. (2015). Mapping dominant tree species over large forested areas using Landsat best-available-pixel image composites. Canadian Journal of Remote Sensing, 41(3), 203-218.

Townsend, P.A. (2002). Estimating forest structure in wetlands using multitemporal SAR. Remote Sensing of environment, 79, 288-304.

Turkar, V., Deo, R., Rao, Y.S., Mohan, S. and Das, A. (2012). Classification accuracy of multi-frequency and multi-polarization SAR images for various land covers. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(3), 936-941.

United States Department of Agriculture, Natural Resources Conservation Service. (2016). Web soil survey. Downloaded from: https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx

Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G. and Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods–a comparison. ISPRS Journal of Photogrammetry and Remote Sensing, 108, 260-272.

Wagner, R. G., Bryant, J., Burgason, B., Doty, M., Roth, B. E., Strauch, P., Struble, D. and Denico, D. (2015). Coming spruce budworm outbreak: Initial risk assessment and preparation and response recommendations for Maine's forestry community. Cooperative Forestry Research Unit, University of Maine, Orono. 77p.

Wang, L., Chang, Q., Yang, J., Zhang, X. and Li, F. (2018a). Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments. PloS One, 13, e0207624.

Wang, Z., Skidmore, A.K., Darvishzadeh, R. and Wang, T. (2018b). Mapping forest canopy N content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. Agricultural and forest meteorology, 253, 247-260.

Wang, Z., Sun, Y., Zhang, T., Ren, H. and Qin, Q. (2018c). Optimization of Spectral Indices for the Estimation of Leaf Area Index Based on Sentinel-2 Multispectral Imagery. In 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22-27 July 2018, 5441-5444.

Waring, R.H. (1983). Estimating forest growth and efficiency in relation to canopy leaf area. Advances in Ecological Research, 13, 327-354.

Wei, C., Huang, J., Mansaray, L.R., Li, Z., Liu, W. and Han, J. (2017). Estimation and mapping of winter oilseed rape LAI from high spatial resolution satellite data based on a hybrid method. Remote Sensing, 9, 488.

Weiskittel, A.R., Kershaw Jr, J.A., Hofmeyer, P.V. and Seymour, R.S. (2009). Species differences in total and vertical distribution of branch-and tree-level leaf area for the five primary conifer species in Maine, USA. Forest Ecology and Management, 258, 1695-1703.

Wessel, M., Brandmeier, M. and Tiede, D. (2018). Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. Remote Sensing, 10(9), 1419.

White, B., Ogilvie, J., Campbell, D.M., Hiltz, D., Gauthier, B., Chisholm, H.K.H., Wen, H.K., Murphy, P.N. and Arp, P.A. (2013). Using the cartographic depth-to-water index to locate small streams and associated wet areas across landscapes. Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 37, 333-347.

Wolter, P.T. and Townsend, P.A. (2011). Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. Remote Sensing of Environment, 115, 671-691.

Wolter, P. T., Townsend, P. A., Sturtevant, B. R. and Kingdon, C. C. (2008). Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario. Remote Sensing of Environment, 112, 3971-3982.

Yu, Y., Li, M. and Fu, Y. (2018). Forest type identification by random forest classification combined with SPOT and multitemporal SAR data. Journal of Forestry Research, 29, 1407-1414.

Yuan, J., Lv, X. and Li, R. (2018). A speckle filtering method based on hypothesis testing for time-series SAR images. Remote Sensing, 10, 1383

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A. and Kycko, M. (2021). Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the Krkonoše/Karkonosze transboundary biosphere reserve. Remote Sensing, 13(13), 2581.

Zhang, C., Ju, W., Chen, J.M., Li, D., Wang, X., Fan, W., Li, M. and Zan, M. (2014). Mapping forest stand age in China using remotely sensed forest height and observation data. Journal of Geophysical Research: Biogeosciences, 119(6), 1163-1179.

Zhang, Y.W., Wang, T., Guo, Y., Skidmore, A., Zhang, Z., Tang, R., Song, S. and Tang, Z. (2022). Estimating community-level plant functional traits in a species-rich alpine meadow using UAV image spectroscopy. Remote Sensing, 14(14), 3399.

APPENDIX

Table A. 1. Sentinel-1 and Sentinel-2 based spectral vegetation indices (SVIs) used as predictor variables for the SBW host species classification in our study. SVIs derived from Sentinel-2 imagery were selected based on their sensitivity towards canopy structure, physiology/stress, and biochemistry.

SVIs	Equation	Reference
Structure		
Atmospherically Resistant Vegetation Index (ARVI)	$\frac{(B8A-2B4+B2)}{(B8A+2\,B4+B2)}$	Kaufman and Tanre (1992)
Enhanced Vegetation Index7 (EVI7)	$\frac{2.5(B7-B4)}{(1+B7+6B4-7.5B2)}$	Majasalmi and Rautiainen (2016)
Enhanced Vegetation Index8 (EVI8)	$\frac{2.5(B8A-B4)}{(1+B8A+6B4-7.5B2)}$	Huete et al. (2002)
Modified Simple Ratio (MSR)	$\frac{((B7/B4)-1)}{sqrt((B7/B4)+1)}$	Chen (1996)
Normalized Difference Vegetation Index (NDVI)	$\frac{(B8A-B4)}{(B8A+B4)}$	Rouse et al. (1974)
Soil Adjusted Vegetation Index (SAVI)	$\frac{1.5(B8A-B4)}{(B8A+B4+0.5)}$	Huete (1988)
Wide Dynamic Range Vegetation Index (WDRVI)	$\frac{((0.01B7)-B4)}{((0.01B7+B4)+0.98)}$	Majasalmi and Rautiainen (2016)
Physiology/stress		
Anthocyanin Reflectance Index1 (ARI1)	$\left(\frac{1}{B3}\right)-\left(\frac{1}{B5}\right)$	Gitelson et al. (2001)
Anthocyanin Reflectance Index2 (ARI2)	$\left(\frac{B8A}{B3}\right)-\left(\frac{B8A}{B5}\right)$	Gitelson et al. (2001)
Carotenoid Reflectance Index1 (CRI1)	$\left(\frac{1}{B2}\right)-\left(\frac{1}{B3}\right)$	Gitelson et al. (2002)
Carotenoid Reflectance Index2 (CRI2)	$\left(\frac{1}{B2}\right)-\left(\frac{1}{B5}\right)$	Gitelson et al. (2002)
Normalized Difference Infrared Index11 (NDII11)	$\frac{(B8A-B11)}{(B8A+B11)}$	Hardisky et al. (1983)
Normalized Difference Infrared Index 12 (NDII12)	$\frac{(B8A-B12)}{(B8A+B12)}$	Key et al. (2002)
Plant Senescence Reflectance Index (PSRI)	$\frac{(B4-B3)}{(B8A)}$	Merzlyak et al. (1999)
Biochemistry		
Chlorophyll Green Index (GCI)	$\left(\frac{B8A}{B3}\right)-1$	Gitelson et al. (2003)
Chlorophyll Red Edge (Clre)	$\left(\frac{B7}{B5}\right)-1$	Gitelson et al. (2003)
Green Atmospherically Resistant Index (GARI)	$\frac{(B8A-B3-(B2-B4))}{(B8A+B3-(B2-B4))}$	Gitelson et al. (1996)
Green NDVI (GNDVI)	$\frac{(B8A-B3)}{(B8A+B3)}$	Gitelson et al. (1996)
Inverted Red Edge Chlorophyll Index (IRECI)	$\left(\frac{B6}{B5}\right)(B7-B4)$	Clevers et al. (2000)

Table A.1. continued…

MERIS Terrestrial Chlorophyll Index (MTCI)	$\frac{(B6-B5)}{(B5-B4)}$	Dash and Curran (2007)
Modified Chlorophyll Absorption in Reflectance Index (MCARI)	$1-\left(\frac{(0.2(B5-B3))}{B5-B4}\right)$	Daughtry et al. (2000)
Normalized Difference Vegetation Index45 (NDVI45)	$\frac{(B5-B4)}{(B5+B4)}$	Delegido et al. (2011)
Normalized Difference Vegetation Index65 (NDVI65)	$\frac{(B6-B5)}{(B6+b5)}$	Gitelson and Merzlyak (1994)
Red-Edge Normalized Difference Vegetation Index (NDVIRE)	$\frac{(B8A-B6)}{(B8A+B6)}$	Gitelson and Merzlyak (1994)
Sentinel-2 Red Edge Position (S2REP)	$705+35\left(\frac{\left(\left(\frac{(B7+B4)}{2}\right)-B5\right)}{(B6-B5)}\right)$	Guyot and Baret (1988)
Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	$3(B5-B4)-0.2(B5-B3)\left(\frac{B5}{B4}\right)$	Haboudane et al. (2002)
Triangular Vegetation Index (TVI)	$0.5(120(B6-B3)-200(B4-B3))$	Broge and Leblanc (2001)
Sentinel-1		
Normalized Ratio Procedure between Bands (NRPB)	$\frac{(\sigma VH-\sigma VV)}{(\sigma VH+\sigma VV)}$	Filgueiras et al. (2019)

BIOGRAPHY OF THE AUTHOR

Rajeev Bhattarai was born in Bandipur, Tanahun, Nepal on 10th August 1994. He went to Notre Dame School (operated by School Sisters of Notre Dame) for his primary, secondary, and high school (graduated from the high school in 2012). Rajeev was admitted to Institute of Forestry, Pokhara Campus, Nepal for his undergraduate degree (2013- 2017).

Rajeev began his Master of Science degree at the University of Maine, the School of Forest Resources in fall, 2018 and graduated in summer 2020 with a M.S. degree in Forest Resources. His research generally focuses on the monitoring of forest health using remote sensing techniques.

Rajeev is a candidate for Doctor of Philosophy degree in Forest Resources from the University of Maine for December 2023.