Classifying the 2022 status of Tsuga Canadensis (Eastern Hemlock) along the Kentucky portion of the Pine Mountain Wildlands Corridor.

Grace M. Embree

University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

Part of the Physical Sciences and Mathematics Commons

Recommended Citation

Retrieved from https://ir.library.louisville.edu/etd/3972

This Master's Thesis is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.
CLASSIFYING THE 2022 STATUS OF TSUGA CANADENSIS
(EASTERN HEMLOCK) ALONG THE KENTUCKY PORTION OF THE
PINE MOUNTAIN WILDLANDS CORRIDOR

By

Grace Embree
B.A, University of Louisville 2020

A Thesis or Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Master of Science of Applied Geography

Department of Geographic and Environmental Sciences
University of Louisville
Louisville, Kentucky

August 2022
CLASSIFYING THE 2022 STATUS OF TSUGA CANADENSIS
(EASTERN HEMLOCK) ALONG THE KENTUCKY PORTION OF THE
PINE MOUNTAIN WILDLANDS CORRIDOR

By

Grace Embree
B.A, University of Louisville 2020

A Thesis Approved on

July 27, 2022

By the following Thesis Committee:

____________________________
Dr. Andrea Gaugan, Thesis Chair

____________________________
Dr. Forrest Stevens, Committee Member

____________________________
Dr. Sarah Emery, Committee Member
ACKNOWLEDGMENTS

I would like to thank the University of Louisville for providing me with both the facilities and connections needed to make this research possible. I would also like to thank Kentucky Natural Lands Trust, Office of Kentucky Nature Preserves, and Eastern Kentucky Universities’ Lilley Cornett Research facility for allowing me to conduct my research in your preservations. Additionally, a very sincere thank you to Murray State University for choosing me to award the KentuckyView Graduate Fellowship for funding this project. I would like to thank the Geographic and Environmental Sciences Department, with special thanks to Dr. Gaughan and Dr. Stevens for their continual enthusiasm and guidance throughout this study. Additionally, thank you to Isaiah Kingsberry and Laura Krauser for accompanying me in field research and aiding in data collection.
ABSTRACT
CLASSIFYING THE 2022 STATUS OF TSUGA CANADENSIS (EASTERN HEMLOCK) ALONG THE KENTUCKY PORTION OF THE PINE MOUNTAIN WILDLANDS CORRIDOR
Grace Embree
August 9, 2022

The invasion of the Hemlock Woolly Adelgid (HWA) (Adelges tsugae) has posed a continual threat in the United States to the Eastern Hemlock (Tsuga canadensis) trees since the 1950s. HWA feed on eastern hemlock needles, reducing the amount of healthy photosynthesizing vegetative area. The use of satellite imagery has been instrumental in identifying areas of eastern hemlock presence. Satellite platforms like Landsat and AVIRIS are commonly used for identification, classification, and mapping of eastern hemlock. Sentinel-2 imagery was released in 2015 for free access. It has a finer spatial grain of with the majority of the bands at 10 and 20 m compared to the 30m resolution of Landsat, for example, and has multiple NIR and SWIR bands where previously used satellites have only one of each, making it ideal for the classification of eastern hemlock trees in the eastern United States. The study will use summer and winter Sentinel-2 imagery in an attempt to answer three questions: 1) What is the current extent of eastern hemlock along the portion of the Pine Mountain Wildlands Corridor within Kentucky? 2) Can various stages of hemlock decline be identified within areas of known hemlock presence? Using a Random Forest classification method in the ArcGIS Pro Environment,
hemlock presence was predicted with a 94% accuracy. The variation in spectral signature of eastern hemlock due to decline led to the inability to predict health stages, however, hemlock canopy coverage was predicted with an 83.6% accuracy. Mapping eastern hemlock trees can inform land management of the status of hemlock death, implications on forest health for areas of death, and identify areas in which treatment is needed on their lands.
# TABLE OF CONTENTS

Acknowledgements .............................................................................. iv

Abstract .............................................................................................. iv

List of Tables ....................................................................................... vii

List of Figures ..................................................................................... viii

Introduction ......................................................................................... 1

Study Area ........................................................................................... 11

Data/Preprocessing ............................................................................. 15

Classification Method ......................................................................... 19

Results ................................................................................................. 24

Discussion ............................................................................................ 33

References ........................................................................................... 39

Curriculum Vita ................................................................................... 42
LIST OF TABLES

Field data collection associated to each ground truth point..........................12
Hemlock Canopy Coverage Classes .......................................................19
Model Diagnostics .................................................................21
Hemlock Presence Model OOB Errors ...........................................22
Confusion Matrix for Hemlock Presence-Absence Model .........................24
Hemlock Canopy Coverage Model OOB Errors ...................................25
Training Diagnostics .................................................................26
Validation Diagnostics ...............................................................26
Hemlock Canopy Coverage Confusion Matrix .....................................27
LIST OF FIGURES

Hemlock Woolly Adelgid Infestation by County .......................... 2
Eastern Hemlock Decline Stages ................................................ 5
Sentinel-2 Spectral Bands ............................................................ 7
Rhododendron Understory ........................................................... 9
Study Area on PMWC ................................................................. 11
Seamline Feathering ................................................................. 14
Seasonal Difference ................................................................. 15
Spectral Signatures ................................................................. 18
Hemlock Presence-Absence Map .............................................. 22
Hemlock Presence Model Variable Importance Graph ................. 23
Hemlock Canopy Coverage Map .............................................. 25
Land Coverage Distribution of Hemlock Canopy Coverage Classes 28
Hemlock Canopy Coverage Model Variable Importance Graph ....... 29
INTRODUCTION

Across the globe, many environments are currently suffering the impacts of invasive species. The importation of exotic goods has a history of introducing species, which when introduced into environments without natural predators or resistance may become a problematic species. The invasive insect species, *Adelges tsugae*, hemlock woolly adelgid (HWA) was introduced into the U.S. in 1951 in Virginia. The HWA has caused a significant decline in the abundance of eastern hemlock (*Tsuga Canadensis*), throughout its natural range. The insect sucks nutrients from the needles, leading to progressive dieback which eventually results in death of the tree. HWA infestation primarily impacts the eastern hemlock tree, but has been known to infest the much less prevalent Carolina hemlock (*Tsuga caroliniana*). HWA has since spread throughout a large portion of the eastern US (Figure 1), stretching from Maine to Georgia, with reports of their presence in some ornamental tree nurseries of Michigan as well working through the native range of the eastern hemlock (McCarty and Addesso 2019). The HWA has been estimated to spread 20-30 km per year, and are wiping out significant stands of eastern hemlock trees as their range continues to expand (Morin, Liebhold, and Gottschalk 2009).
HWA infestation affects not only the eastern hemlock, but the forest as a whole. The tree is a keystone species, meaning that its health has a disproportionate effect on its ecosystem. Eastern hemlocks support upwards of 120 species by providing habitat, food, and regulation of water circulation (The National Wildlife Federation). Die off of eastern hemlocks affects the overall water cycle in forests due to the change in respiration rates. Because the species that are replacing hemlock have lower respiration rates, or are deciduous, meaning that respiration will spike in the summer season, overall streamflow and water circulation in hemlock inhabited forests will change drastically. The effects of
hemlock death cascades through the entire food web, resulting in population decline of other reliant species (Farmer 2013; McCarty and Addesso 2019).

HWA does not have any natural predators in the United States, and eastern hemlock trees do not have a mitigating biological response to infestation (Williams et al. 2016). HWA are spread passively, meaning they disperse through wind, are carried by animals like birds and deer, and by human activity. Rate of dispersal has been estimated at 16-24 km per year and has encompassed the majority of the native range of eastern hemlock (McCarty and Addesso 2019). Many management plans rely on trajectory modeling of HWA spread. Using data from historically HWA infested areas coupled with environmental factors such as climate or topography, predictions have been made about where and how quickly the insect will spread in the future (Trotter et al. 2020). Dispersion modeling has been done at both the national and landscape levels to assess HWA movement. Research facilities like Coweeta in North Carolina have networks of monitoring plots that have been gathering HWA and eastern hemlock status data for decades and are used to help predict the trajectory of HAW in similar landscapes (USDA). Some management strategies have even incorporated citizen science by soliciting locals to report on HWA infestation when venturing into the forest (Kanoti 2022; Pistolese 2018).

In addition to tracking and monitoring HWA infestation, spatial technology has proven very useful in evaluating the ongoing affects the insect causes to eastern hemlocks themselves. The use of satellite imagery and other remote sensing technology, like aerial photography, have been used widely to evaluate forest composition, and more recently have been incorporated into HWA studies (Hanavan, Pontius, and Hallett 2015; Kong et
Multispectral imagery offers the ability to study the effects of HWA on both eastern hemlock and forest status as a whole by capturing reflectance patterns which cannot be observed by the human eye. Sensors that capture reflectance patterns in both visible and non-visible portions of the electromagnetic spectrum (EMS) are important for vegetation mapping, especially wavelengths in the near infrared (NIR) as healthy, green vegetation reflects very highly in this portion of the EMS. Even slight variations in NIR reflectance can be used to differentiate between healthy versus unhealthy vegetation, and if reflectance patterns are known can be used to identify individual species from one another (Carter and Miller 1994; Grabska and Socha 2021; Hanavan, Pontius et al. 2015).

By incorporating remote sensing technology into the tracking of HWA, infestation can be studied on a larger scale, as well as save time and expenses that would be needed to investigate trees on the ground alone. When eastern hemlocks are infested with HWA, there is a biological response that can be detected via satellite imagery in which chlorophyll content within the needles themselves spike (Williams et al. 2016). As the infestation persists, however, dieback causes the density of the needles and branches to decline, and the overall area of photosynthetically active needles declines. Because eastern hemlocks can live up to a decade after initial infestation, these trees go through multiple stages of decline (Jennifer Pontius et al. 2010). These various stages of decline (Figure 1) will produce different spectral signatures due to the ratio of healthy photosynthesizing needles to dead and bare areas, which can be identified and analyzed via vegetation indices such as the Normalized Difference Vegetation Index ((NIR - Red) / (NIR + Red)). NDVI uses the ratio of reflectance in the NIR band, and absorption of the red band of a given sensor which will highlight areas of healthy, photosynthetically active
vegetation (Williams et al. 2019; Carter and Miller 1994). Other types of remotely sensed imagery such as those produced from Light Detecting and Ranging (LiDAR) have been used to identify defoliated trees as well (Boucher et al. 2020). However, the added ability of optical imagery being able to produce health indicators like those mentioned above, often make it favorable for use in eastern hemlock studies.

Figure 2 Eastern Hemlock Decline Stages (Photos by Gina Davis (USDA 2015))

Many satellite platforms have been used to study eastern hemlocks such as Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Pontius et al. 2010) and the Global Ecosystem Dynamics Investigation (GEDI) spaceborne LiDAR (Boucher et al. 2020). Landsat satellites and aerial photography are the most commonly used platforms for classifying eastern hemlock (Hanavan, Pontius, and Hallett 2015; Kong et al. 2008). Due to its freely accessible imagery archive which provides new imagery every 16 days (Cohen and Goward 2004) Landsat imagery has been utilized for detecting the change in hemlock presence over time on a landscape scale (Royle and Lathrop 2002; Cohen and
Goward 2004; Kong et al. 2008; Williams et al. 2016). The 30-meter spatial resolution offers the ability to identify certain individual species on a landscape when coupled with environmental and ground reference data (Royle and Lathrop 2002; Walsh 1980).

The range of eastern hemlock trees spans a large portion of the eastern United states. Throughout its native range, species composition and abundance vary. In many areas eastern hemlock is a codominant species and tends to grow alongside other hard woods such as the hemlock-red spruce forests in Maine or beech-red maple ecological types in New Hampshire (Solomon and Leak). There are some areas, however, that eastern hemlock is the dominant tree, growing in large stands across most or all of the forested area. This is common in both Ohio and North Carolina where hemlock trees grow in large valleys, dominating those moist, shaded areas (Lake Forest College 2022).

In mixed mesophytic forests, neither hemlock or any other tree truly dominate the forests. Eastern hemlock is often the most dominant conifer in this type of ecological system (National Parks Service). Identifying eastern hemlock stands in a mixed mesophytic forest as opposed to areas where it is a more dominant species, provides additional challenges in trying to separate the spectral signature of other green, healthy, photosynthesizing vegetation.

The newly free-access Sentinel-2 platform offers a finer spatial grain of 10-20 meters for ten out of its 13 spectral bands (Figure 3) as well as near infrared (NIR) and multiple red-edge bands which have been shown to aid in separating species on a landscape level. The presence of multiple NIR bands provides a larger range in which healthy vegetation types can be distinguished from one another compared to platforms with only a single NIR band. The Sentinel-2 platform consists of two sensors with 180
degrees of separation between their orbits. The additional sensor in the platform allows imagery to be taken of any one area in the mid-latitudes every 2-3 days (Grabska and Socha 2021).

<table>
<thead>
<tr>
<th>Sentinel-2 Bands</th>
<th>Central Wavelength (µm)</th>
<th>Resolution (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1 - Coastal aerosol</td>
<td>0.443</td>
<td>60</td>
</tr>
<tr>
<td>Band 2 - Blue</td>
<td>0.490</td>
<td>10</td>
</tr>
<tr>
<td>Band 3 - Green</td>
<td>0.560</td>
<td>10</td>
</tr>
<tr>
<td>Band 4 - Red</td>
<td>0.665</td>
<td>10</td>
</tr>
<tr>
<td>Band 5 - Vegetation Red Edge</td>
<td>0.705</td>
<td>20</td>
</tr>
<tr>
<td>Band 6 - Vegetation Red Edge</td>
<td>0.740</td>
<td>20</td>
</tr>
<tr>
<td>Band 7 - Vegetation Red Edge</td>
<td>0.783</td>
<td>20</td>
</tr>
<tr>
<td>Band 8 - NIR</td>
<td>0.842</td>
<td>10</td>
</tr>
<tr>
<td>Band 8A - Vegetation Red Edge</td>
<td>0.865</td>
<td>20</td>
</tr>
<tr>
<td>Band 9 - Water vapour</td>
<td>0.945</td>
<td>60</td>
</tr>
<tr>
<td>Band 10 - SWIR - Cirrus</td>
<td>1.375</td>
<td>60</td>
</tr>
<tr>
<td>Band 11 - SWIR</td>
<td>1.610</td>
<td>20</td>
</tr>
<tr>
<td>Band 12 - SWIR</td>
<td>2.190</td>
<td>20</td>
</tr>
</tbody>
</table>

*Figure 3 Sentinel-2 Spectral Bands*

The first documentation of HWA infestation in the state of Kentucky came in 2006, starting in Harlan and Letcher Counties (University of Kentucky 2018). Within this area of Kentucky resides a large stretch of protected lands that make up the Pine Mountain Wildlands Corridor. HWA infestations pose a threat not only to the eastern hemlock species, but to this ecosystem as a whole. As a keystone species (Russo et al. 2019), the health of this corridor is significantly affected as eastern hemlocks die off. In order to combat HWA infestations, forest management has turned to pesticides since the trees have no biological defenses of their own and the HWA has no natural predators in
North America (University of Kentucky 2018). The primary chemical used for HWA control is a broad spectrum insecticide known as imidacloprid. The pesticides used are able to stave off death for eastern hemlocks, but must be administered every 5-10 years. Trees are treated on an individual basis by injecting the pesticide at the base of the tree to be taken up by the roots and killing the HWA when it feeds on needles (Havila and D’Amico 2017). Because it is a broad spectrum treatment, when this chemical is injected into the tree or the soil surrounding it, there is no guarantee that it will not kill or harm other organisms that interact or feed on eastern hemlocks (Dilling et al. 2009). Because the ecosystem is so heavily influenced by hemlock death and treatment regimes, assessing the current state of eastern hemlock coverage is necessary for the continued efforts toward protecting the health of the Pine Mountain Wildlands Corridor (PMWC).

Previous studies have assessed the extent of eastern hemlock presence in Harlan County (Kong et al. 2008) but since 2008 no publicly available results have been published. HWA infestation has now spread to the majority of areas in Kentucky where eastern hemlocks grow causing widespread hemlock death throughout the state. As eastern hemlock decline progresses, the composition of the forest surrounding the trees will be altered as well. Ecologists have found a trend in the growth of \textit{Rhododendron maximum}, an evergreen understory shrub, has increased as the eastern hemlock defoliates and dies off (Brantley, Ford, and Vose 2013; Farmer 2013). The increase of light coming through the canopy where needles and branches die allows for increased growth of the shrub. Increases in rhododendron growth has been linked to issues such as changes in water use and availability as well as effects on the growth and productivity of surrounding vegetation. As rhododendron bushes grow, they are close to the ground and
have thick layers of broadleaf, meaning that very little of the light that penetrates the canopy is able to reach the ground below this understory bush. As a result, saplings are unable to receive the sunlight needed to grow into mature trees (Farmer 2013; Hawthorne, Miniat, and Elliott 2017). This change to understory forest structure and suppression of tree saplings that would repopulate the canopy could potentially have a long-lasting effect on the ecosystem as a whole and demands the attention of researchers in attempts to mitigate the fallout caused by hemlock death.

Figure 4 Rhododendron Understory

In this research, I seek to address two questions: 1) What is the current extent of eastern hemlock along the portion of the PMWC within Kentucky in 2021? 2) Can various stages of hemlock decline be identified within areas of known hemlock presence? I expect that along the PMWC within Kentucky using a random forest ensemble-based classifier combined with the highest spatial resolution imagery from Sentinel-2, that this research will help refine spatial mapping of eastern hemlock. Using a combination of vegetation indices with ground-truth data, it is predicted that eastern hemlock trees can be
classified by level of decline (i.e. defoliation and dieback) on a four-class basis from healthy to standing-dead status.
STUDY AREA

This study encompasses the extent of the Pine Mountain Wildlands Corridor (PMWC) that lies within Kentucky, spanning three counties in southern Kentucky; Harlan, Bell, and Letcher Counties. The PMWC comprises many parcels of land which are owned both privately and by Kentucky based organizations. The Kentucky Natural Lands Trust (KNLT) and the Office of Kentucky Nature Preserves (OKNP) collectively oversee a significant portion of the corridor within Kentucky. For this study, research permits were acquired to work within preserves owned by both organizations across the corridor.

The PMWC is considered a mixed mesophytic forest meaning that the region is covered by mixed vegetation rather than one dominant species assemblage. The temperate region receives moderate precipitation and varies in humidity. Vegetation ranges in age, but have characteristically distributed areas of old growth (Martin 1992). Coal harvesting and fracking is common in the region, but the PMWC does not have merchantable coal or active harvesting of such resources. With only nine roads crossing through the corridor, there is little surrounding residential area (KNLT 2022). Total population of each county of the study area is approximately 20 thousand residents (U.S. Census Bureau 2021).

Reference data was collected in the summer (leaf-on season) of 2021 within the boundaries of the preserves owned by the partner organizations of this study (Figure 2). Accessible areas were limited to those not on private land and nature preserves with public trails due to the rough terrain of the mountain. Reference points were taken within
a 60-meter boundary of the public trails. Prior to field work, shape files of each preserve were shared by owners and used to create a boundary in which reference points would be taken. The ArcGIS Pro tool Create Random Points was used to place randomized points around the preserves at least 250 meters apart. This helps to eliminate bias in reference points and aid in collecting a larger range of reference data.

**Kentucky Portion of the Pine Mountain Wildlands Corridor**

![Study area on Pine Mountain Wildlands Corridor, True Color Summer 2021 Imagery](image)

*Figure 5. Study area on Pine Mountain Wildlands Corridor, True Color Summer 2021 Imagery*

Ground reference points were collected in June and July of 2021. Using the Esri Field Maps app, an offline, editable map was made which contained the boundary layers
for the study area and randomized points placed within those boundaries and along hiking trails to act as a guide for collecting reference points. A handheld Bad Elf GNSS Surveyor (Bad Elf 2014) unit was used for more accurate data by connecting the unit to the mobile device in which the Field Maps app was being used. At each reference site information on hemlock presence and stage of decline, as well as the ratio of hemlock and rhododendron presence within an approximate 30-60m area surrounding the point which would represent a 3x3 pixel in the imagery. Once the randomized point (or closest possible location to it) was reached, information was compiled into a table within the Field maps app which was associated with each individual point. All points were assessed by at least two field researchers and corroborated before logging them into the table.

Stages of decline were based on the stages identified in Figure 2 from the USDA’s guide for assessing eastern hemlock health. These stages included healthy (1), light decline (2), moderate decline (3), and severe decline (4) (USDA 2015). The level of decline was assessed for a surrounding 30 meter area, which would equate to an approximate 9x9 pixel area in the Sentinel-2 imagery. Site information was assessed for that area using a method of analysis similar to that of quadrant sampling, with the 3 x 3 pixel area being the quadrant. The decline stage of all hemlock trees within the 30 x 30 m area was averaged and corroborated between all researchers in the field. The same estimation strategy was done for, rhododendron and ground cover. Hemlock canopy coverage was also collected by estimating the percentage of the 30 x 30 meter area was covered by hemlock foliage. Five classes in total were identified on a 25 percent increment from 0% to 100% coverage, with the 0% class having no hemlock present. In total, between two research trips across June and July, 102 reference points were taken.
<table>
<thead>
<tr>
<th>Hemlock Information</th>
<th>Rhododendron information</th>
<th>Environmental information</th>
</tr>
</thead>
<tbody>
<tr>
<td>· Hemlock Dominant (+,-)</td>
<td>· % Rhododendron coverage</td>
<td>· Elevation</td>
</tr>
<tr>
<td>· % Hemlock coverage</td>
<td>· Dominant understory</td>
<td>· Canopy coverage (light penetration)</td>
</tr>
<tr>
<td>· Infestation (0,1)</td>
<td></td>
<td>· Dominant ground cover and % coverage</td>
</tr>
<tr>
<td>· Decline stage (1-5)</td>
<td></td>
<td>· Issues that may affect spectral signature</td>
</tr>
</tbody>
</table>

*Table 1 Data Collected at Each Reference Point*
DATA COLLECTION AND PREPROCESSING

Sentinel-2 imagery has been made freely available via United States Geological Survey’s (USGS) Earth Explorer data portal (U.S. Geological Survey 2021) and the Copernicus Open Access hub. Winter imagery was downloaded from the Copernicus portal for February 20th, 2022 (leaf-off season). Level 2A imagery was selected for the extent of the study area as the processing level includes atmospheric and topographic correction. Each image was then resampled to be 10 meter pixel size of the RGB and NIR bands. This process was performed in ENVI 5.6 (Exelis Visual Information Solutions, Boulder, Colorado), using the Layer Stacking tool. Level 2-A Sentinel-2 imagery does not include band 10, used to discuss cirrus cloud contamination (1.375 nm), so the remaining 13 bands were stacked together using nearest neighbor resampling with the blue band (band 2) as the reference image. Three image tiles were stitched together in the ENVI processing environment using the Seamless Mosaic tool. In order to reduce the appearance of seam lines where the images meet or overlap, the Seam Feather selection with histogram matching for overlapping areas was used within the tool. Histogram matching uses the selected reference image of the images being mosaiced together to adjust the grayscale of each band to the reference’s corresponding histogram. This correction method was only used for the overlapping areas of imagery, meaning that any portions of the image that do not overlap with another are left unadjusted. Seamline feathering determines how many pixels away from the seam line will be blended between images. Seamline feathering uses pixels on both sides of the actual seamline, meaning that both the reference and the adjusted images will be blended on each side of where the
images meet up to more cohesively join them and produce a less obvious seam in the 
final mosaiced (Exelis Visual Information Solutions, Boulder, Colorado).

Figure 6 Seamline Feathering (Adapted from Harris Geospatial Solutions, 2020)

The available summer imagery for 2021 which encompassed the full study area 
were all found to have significant cloud cover in areas where reference data was 
collected. Because cloud cover obscures the surface reflectance data below, summer 
imagery was acquired in a different manner. A composite of summer imagery at multiple 
dates through the season was obtained using Google Earth Engine (GEE), a cloud-based 
computing platform increasingly used for remote sensing analyses. Sentinel-2 Level 2-A 
imagery was identified between June 1st and July 31st of 2021. This imagery contained 
less than or equal to 20% cloud cover. Due to differences in coverage across the dates, 
the median of each pixel value was used to create the resulting image. This excludes any 
extreme pixel values, but will still accurately represent the value of each pixel for the full 
leaf-on season. Four individual image tiles were produced from this process which were a 
stack of all bands excluding band 10 the Cirrus band, at 10 meter resolution. Mosaicing
in the ENVI environment proved problematic as the resulting image using the Seamless Mosaic tool repetitively created an output which had a single pixel line of No-Data values stretching horizontally across the study area where the images meet. This issue was theorized to be caused by the orientation point within each individual image being a central pixel rather than the first pixel in each column and row, causing a misorientation of the images when trying to stitch them together. For this reason, the summer composite imagery was mosaiced in ArcGIS Pro using the Mosaic to New Raster tool. The image which covered the majority of the study area and the training points was used as the reference image of the four. Pixel values were color corrected to correspond to that of the reference image, and overlap areas were made to be the average of those pixels. The resulting mosaiced image was then clipped to the extent of the study area to be used in classification (Figure 7).

**Figure 7 Seasonal Difference of True color RGB Sentinel-2 Imagery between Winter and Summer**

Digital Elevation Models (DEMs) for the study area were downloaded from the USGS TNM data portal. Original pixel sizes were non-square 8 by 10-meter pixels.
Using the quick mosaic feature in ENVI 5.6, all DEMs were compiled to cover the study area and resampled to 10m square pixels using the Bilinear resampling method to avoid creating linear noise when used in the topographic correction method. In addition, the Slope raster layer was also created in ArcGIS Pro with the Slope tool which used the DEM mosaic obtained from USGS as the input data. The slope layer was also used to create a topographic wetness index (TWI) in the ArcGIS Pro environment. This raster layer is an indicator for areas that accumulate water and will be used in the classification as a parameter for identifying eastern hemlock and rhododendron habitat. Creating the TWI followed the method outlined in the online tutorial of the Formation SIG team which adapted the formula from Bevin & Kirby (1979).

\[ TWI = \ln \left( \frac{SCA}{\tan \phi} \right) \]

Where SCA is Specific Catchment Area, or the contributing upslope area where water will flow from, and \( \phi \) is the slope angle. Creating the TWI followed the method outlined in the online tutorial of the Formation SIG team which adapted the formula from Bevin & Kirby (1979). Changes in slope are used to estimate water flow, meaning that well drained areas where slope is steep will be associated with low TWI values and areas with low slope will be associated with high TWI values. As an indicator for areas that accumulate water this data layer will be used in the classification as a parameter for identifying eastern hemlock and rhododendron habitat which are known to be associated with high water accumulation.
CLASSIFICATION METHOD

Identifying Eastern Hemlock Presence

Classification workflow was adapted from the methods for identifying eastern hemlock by Kong et al. (2008). Because eastern hemlock and rhododendron are both evergreen species, a Normalized Difference Vegetation Index (NDVI) derived from the winter imagery (Feb 2022) was used to identify the evergreen vegetation in the area. Using the NDVI tool in ArcGIS Pro, the ratio of the difference in reflection in the near-infrared (band 8) and absorption in the red (band 4) were calculated. The result is an image scaled from -1 to +1 with higher values typically associated with areas of dense, healthy, green, and photosynthesizing vegetation. To identify the areas where evergreen vegetation persists within the study area, the difference between summer and winter NDVI was identified. Using the Change Detection Wizard tool in ArcGIS Pro, a raster layer was calculated by subtracting winter NDVI values from summer NDVI values in order to identify areas of significant negative change in vegetation coverage. Using the histogram provided, the distribution of negative change was identified as beginning at approximately -0.623. This threshold was used to create a binary raster in which values equal to or less than this value were considered significant negative change, and all values higher than -0.623 were considered areas of little to no change. The areas of little to no change did, however, include land cover classes such as water, roads, or built up structures as the values for these areas would not change significantly. Once used in the classification however, these areas would not have similar spectral signatures of known
hemlock points and would therefore not likely be contributing to misclassification of these land covers as hemlock.

Using the spectral signatures of three separate points where: 0% hemlock canopy coverage, 26-50% canopy coverage, and 76-100% canopy coverage was observed in the field. The image below shows the spectral signature from a representative reference point as its reflectance corresponds to each band in the Summer 2021 Sentinel-2 composite image. The difference between reflectance values for each of these classes is highest for bands 6, 7, 8, and 8A NIR (total range: 0.705-0.865 nm), and the SWIR bands 9, 11, and 12 (total range: 0.945-2.190 nm). Band 10 cirrus is not present in this imagery as it is removed from the stack during atmospheric corrections in the SNAP environment but is represented by the tenth band here, while the ninth band is representing the Sentinel-2 band 8A. These bands should prove to be the most important in distinguishing between eastern hemlock, as higher hemlock canopy coverage percentages appear to have lower reflectance than those lower hemlock canopy coverage points in these bands.

![Spectral Profile](image)
Random Forest classifiers have been shown to accurately classify individual species (Knauer et al. 2019; Nguyen et al. 2019). The ArcGIS Pro Random Forest classification and regression tool is an adapted version of Breiman’s random forest model (Breiman 2001). The ensemble based approach has been shown to predict better than other classification methods (Cutler et al. 2007). Random Forest classification also has the ability to produce model characteristics and rank variables based on their importance in splitting trees which can both be used to improve the model. Additionally, using a Random Forest classification for a multi-class prediction has been shown to perform better in approximating boundaries between observations across classes (Cutler et al. 2007). In the model building process for this research, the training points taken in the field will determine the classes based on a binary of presence/absence where presence is denoted as 1 and absence 0. The raster imagery used for the classification will include all bands of the summer composite and winter imagery loaded to the tool individually, a DEM, TWI, summer and winter NDVI, NDVI change detection, and a slope layer. Using 90% of the training points, the model will then use the decision trees, group them, and then vote based on majority whether a pixel is classified with or without hemlock presence.

Identifying Eastern Hemlock Canopy Coverage

After eastern hemlock areas were identified, the classified image of hemlock presence was used to mask out areas that were predicted to have no hemlock presence. Using the percent of hemlock coverage value of each training point and the subset of hemlock areas, a hemlock canopy coverage classification was created to identify areas of
higher hemlock presence using the same random forest classifier method. This produced a five-class output (the values of each class can be found in Table 2).

<table>
<thead>
<tr>
<th>Hemlock Canopy Coverage Classes</th>
<th>Values of % Hemlock Canopy Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>0%</td>
</tr>
<tr>
<td>Class 2</td>
<td>1-25%</td>
</tr>
<tr>
<td>Class 3</td>
<td>26-50%</td>
</tr>
<tr>
<td>Class 4</td>
<td>51-75%</td>
</tr>
<tr>
<td>Class 5</td>
<td>76-100%</td>
</tr>
</tbody>
</table>

*Table 2 Hemlock Canopy Coverage Classes*

Model Assessment

After incorporating the necessary data layers into the Random Forest model, a preliminary run of the classification was performed with default settings used in ArcGIS Pro which are shown in Table 3, with the exclusion of the number of trees which is defaulted to 100. Along with a classification raster, the tool provides diagnostics that are used to assess the robustness of the model. The model diagnostics product provides an overall model and individual class score for the proportion of the Out of Bag (OOB) errors identified in the model (Arc GIS Pro 2020). The closer this score to 0, the less error is present in predicting values. The diagnostics produced (Table 3) provide information about the parameters used in the model. Along with the proportion of OOB errors, the model diagnostics inform the user on how the model is performing based on the number of trees, leaf size, and the percent of the training data available for each tree. Comparing the proportion OOB error can help the user assess if these parameters need to be adjusted, and how well the model performs when they are adjusted. Accuracy and
sensitivity are also calculated for the training and validation data for each class. Sensitivity is the percentage of known training points correctly classified while the accuracy takes into account both the percentage of known points correctly classified and how often points are misclassified for any one class. Both sensitivity and accuracy range from 0-1.

F1-Scores and MCCs are other means of assessing the accuracy of a model. F1-Score is a weighted average of precision (the ratio of correct classifications to total classifications) and recall (the ratio of correct classifications to the total number of objects in that class). The F1-Score ranges from 0-1, with 1 being the best. (ArcGIS pro 2020. The Matthew’s Correlation Coefficient is a measure of performance of the model based on the ratios of true positives and negatives to false positives and negatives in regard to the total number of classifications. The score ranges from -1 to 1 with a 1 being perfect agreement across the model and -1 being no agreement (Jurman, Riccadonna, and Furlanello 2012).

<table>
<thead>
<tr>
<th>Model</th>
<th>Hemlock Presence</th>
<th>Hemlock Canopy coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Trees</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Leaf size</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tree Depth Range</td>
<td>87-159</td>
<td>176-280</td>
</tr>
<tr>
<td>Mean Tree Depth</td>
<td>120</td>
<td>223</td>
</tr>
<tr>
<td>% of Training Available per Tree</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Number of Randomly Sampled Variables</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>% of Training Data Excluded for Validation</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

*Table 3 Model Diagnostics*
RESULTS

Hemlock Presence

Figure 9 shows the results of the random forest classifier used to identify areas in which eastern hemlock can be found. Areas classified as presence (1) made up approximately 49% of the study area compared to areas of absence (0) covering 51%. The accuracy of the model can be assessed through the proportion of OOB errors, training and validation diagnostics, and the overall accuracy of the model as well as the user's and producer’s accuracies calculated from the Confusion Matrix. The classifier for the eastern hemlock presence-absence model had an OOB error proportion of 0.523 when setting the number of trees to 300 (Table 3). Additionally, the OOB for the absence and presence classes individually were 0.091 and 0.774, respectively. The scores for the training data reported sensitivity of the absence class to be 100% and 99% for the presence class. Training data accuracy for each class was 1.00. For the validation data, sensitivity was 93% and 95% for absence and presence, respectively. Validation accuracy was found to be 94% for both classes. Validation F1-scores for the absence was 0.92 and 0.95 for presence. The MCC score for each class was 0.87.
When performing a random forest classification, the tool also produces a Variable Importance Table, which ranks the top 20 explanatory rasters in order from most important to least. Importance is calculated using the Gini coefficient, which is a representation of how many splits in a decision tree are caused by an individual variable (in this case, an explanatory raster). For identifying eastern hemlock presence, band 9
SWIR (0.945 nm) from both summer and winter imagery were of highest importance, Gini coefficient of 5.15 and 5.1 respectively. These bands were followed by band 1 coastal aerosol (0.443 nm) from the summer imagery, band 12 SWIR (1.610 nm) from the winter imagery and the DEM. NIR (0.705-0.865 nm) and NDVI bands had less importance, 3.5, with the TWI, Slope, and distance to roads layers having lower Gini coefficients than those appearing on the list.

![Hemlock Presence Model Variable Importance Graph](image)

**Figure 10 Hemlock Presence Model Variable Importance Graph**

The confusion matrix provides a better look at how misclassification is distributed across all classes. Both presence and absence classes had low omission and commission errors, all of which were 91% or higher. The presence class had higher user’s and producer’s accuracies than the absence class. The overall accuracy was calculated to be 94%, with a Kappa statistic of 87.1%
<table>
<thead>
<tr>
<th></th>
<th>Absence</th>
<th>Presence</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence</td>
<td>113</td>
<td>9</td>
<td>122</td>
</tr>
<tr>
<td>Presence</td>
<td>11</td>
<td>199</td>
<td>210</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>208</td>
<td>332</td>
</tr>
</tbody>
</table>

Overall Accuracy = 122/332 = 94.0%  
Kappa = 87.1%

<table>
<thead>
<tr>
<th>Producer’s Accuracy (omission error)</th>
<th>User’s Accuracy (commission error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence = 113/124 = 91.1%</td>
<td>Absence = 113/122 = 92.6%</td>
</tr>
<tr>
<td>Presence = 199/208 = 95.7%</td>
<td>Presence = 199/210 = 94.8%</td>
</tr>
</tbody>
</table>

Table 5 Confusion Matrix for Presence-Absence Model

Hemlock Canopy coverage

Figure 11 depicts the distribution of hemlock canopy coverage across the areas identified to have hemlock presence. Four classes were created with varying densities of hemlock ranging from 1-100% and divided into 25% intervals. The gray areas depict those that were not identified as having hemlock presence, or 0% canopy coverage. The overall proportion of OOB error for the hemlock canopy coverage model is 9.063. Overall accuracy was calculated to be 83.6%.
Figure 11 Hemlock Canopy Coverage

<table>
<thead>
<tr>
<th>Number of Trees</th>
<th>150</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall proportion of OOB error</td>
<td>10.006</td>
<td>9.063</td>
</tr>
<tr>
<td>76-100% OOB error</td>
<td>0.351</td>
<td>0.176</td>
</tr>
<tr>
<td>51-75% OOB error</td>
<td>9.777</td>
<td>8.984</td>
</tr>
<tr>
<td>26-50% OOB error</td>
<td>6.126</td>
<td>5.640</td>
</tr>
<tr>
<td>1-25% OOB error</td>
<td>11.38</td>
<td>10.603</td>
</tr>
<tr>
<td>0% OOB error</td>
<td>12.220</td>
<td>10.833</td>
</tr>
</tbody>
</table>

Table 6 Hemlock Canopy coverage Model Out of Bag Errors

Tables 7 and 8 show the training and validation accuracy and sensitivity scores for the canopy coverage model. The sensitivity, or percentage of correctly classified
training or validation points, is at 90% or higher for training data and in the 80% or higher for the validation data. Accuracy for all classes, excluding the 76-100% canopy coverage class, have higher accuracy scores compared to their sensitivity scores which will be addressed later in the discussion section. The 76-100% canopy coverage class has the highest accuracy, with the 0% canopy coverage class having the lowest accuracy for both training and validation. Validation F1-Scores varied from 0.80 -0.87, with the highest score for the 0% class and the lowest for the 26-50% class. MCC scores for validation points were lower for all classes except the 76-100% class, ranging from 0.80-0.85. The lowest score was again found to be for the 26-50%, but the highest MCC score was found for the 76-100% class.

<table>
<thead>
<tr>
<th>Class</th>
<th>F1-Score</th>
<th>MCC</th>
<th>Sensitivity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>76-100%</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>51-75%</td>
<td>0.92</td>
<td>0.90</td>
<td>0.92</td>
<td>0.97</td>
</tr>
<tr>
<td>26-50%</td>
<td>0.92</td>
<td>0.90</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>1-25</td>
<td>0.91</td>
<td>0.88</td>
<td>0.90</td>
<td>0.96</td>
</tr>
<tr>
<td>0</td>
<td>0.93</td>
<td>0.88</td>
<td>0.90</td>
<td>0.95</td>
</tr>
</tbody>
</table>

*Table 7 Training Diagnostics*

<table>
<thead>
<tr>
<th>Class</th>
<th>F1-Score</th>
<th>MCC</th>
<th>Sensitivity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>76-100%</td>
<td>0.85</td>
<td>0.85</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>51-75%</td>
<td>0.85</td>
<td>0.81</td>
<td>0.85</td>
<td>0.94</td>
</tr>
<tr>
<td>26-50%</td>
<td>0.80</td>
<td>0.76</td>
<td>0.81</td>
<td>0.94</td>
</tr>
<tr>
<td>1-25</td>
<td>0.82</td>
<td>0.77</td>
<td>0.80</td>
<td>0.92</td>
</tr>
<tr>
<td>0</td>
<td>0.87</td>
<td>0.80</td>
<td>0.86</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Table 8 Validation Diagnostics*
All classes had relatively few omission and commission errors. The 76-100% class had no commission error, but a 26.1% omission error. Excluding this class and the 21-50% class, all producers and users accuracies were at or above 80%. The overall accuracy was calculated to be 83.6% with a Kappa statistic of 78.2%.

<table>
<thead>
<tr>
<th></th>
<th>76-100%</th>
<th>51-75%</th>
<th>26-50%</th>
<th>1-25%</th>
<th>0%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>76-100%</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>51-75%</td>
<td>2</td>
<td>56</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>69</td>
</tr>
<tr>
<td>26-50%</td>
<td>3</td>
<td>3</td>
<td>42</td>
<td>1</td>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>1-25%</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>61</td>
<td>8</td>
<td>76</td>
</tr>
<tr>
<td>0%</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>105</td>
<td>122</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>23</td>
<td>66</td>
<td>53</td>
<td>72</td>
<td>122</td>
<td>336</td>
</tr>
</tbody>
</table>

**Overall Accuracy** = \( \frac{1336}{1336} = 83.6\% \)  
**Kappa** = 78.2%

<table>
<thead>
<tr>
<th>Producer’s Accuracy (omission error)</th>
<th>User’s Accuracy (commission error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>76-100% = 17/23 = 73.9%</td>
<td>76-100% = 17/17 = 100%</td>
</tr>
<tr>
<td>56-75% = 56/66 = 84.8%</td>
<td>56-75% = 56/69 = 81.2%</td>
</tr>
<tr>
<td>21-50% = 42/53 = 79.2%</td>
<td>21-50% = 42/52 = 80.8%</td>
</tr>
<tr>
<td>1-25% = 61/72 = 84.7%</td>
<td>1-25% = 61/76 = 86.3%</td>
</tr>
<tr>
<td>0% = 105/122 = 86.1%</td>
<td>0% = 105/122 = 86.1%</td>
</tr>
</tbody>
</table>

Table 9 Hemlock Canopy coverage Confusion Matrix

The amount of land covered by each of the four canopy coverage classes varied significantly. The lowest total area was the highest canopy coverage class, 76-100% with less than 5% of the study area falling in this class. Following it was the 26-50% class, at 11%, then the 51-75% canopy coverage class at 16.2%. Excluding the areas of 0%, or no hemlock presence, the 1-25% canopy coverage class comprised the most land at 21.9%.
For identifying eastern hemlock canopy coverage, the DEM variable was found to have the highest importance with a Gini Coefficient score of 11.21. Following the DEM was band 9 SWIR (0.945 nm) from summer, 9.48 score, and then the winter band 11 SWIR (1.610) with an 8.69. Subsequent bands had Gini Coefficient scores of approximately 7, until dropping below 6 for the winter band 8A NIR (0.865). The remaining variables had Gini Coefficients of below 3. Both the Change detection and TWI appeared on the list, contrasting their absence in the hemlock presence model, pushing the slope, distance to roads layer, and RGB bands to importance values below those on the list.
Figure 13 Hemlock Canopy Coverage Model Variable Importance Graph
DISCUSSION

Using a combination of Sentinel-2 imagery and environmental variables, the Forest-based Classification was able to identify eastern hemlock presence on the PMWC as well as the surrounding lands with a 94.0% overall accuracy, and individual class accuracies all above 90%. Because the study area is classified as a mixed mesophytic forest, eastern hemlock does not dominate the extent of the areas it resides in but is dispersed throughout in varying degrees of canopy coverage. Using the extent of the presence class, an acceptable canopy coverage classification was created with an overall accuracy of 83.6%. The ability to accurately classify forested areas as eastern hemlock using remotely sensed imagery allows land management teams combating HWA infestation to redirect their efforts away from locating the trees, and toward treatment regimes. Similarly, identifying the areas in which eastern hemlock exists at a higher canopy coverage can help in prioritizing how HWA treatment is distributed across large swaths of land.

The total area of hemlock presence covered approximately 49% of the study area. Of that area, only about 5% was covered by the highest canopy coverage stands of hemlock and the majority of hemlock presence areas were classified as being 1-25% canopy coverage. This sparse coverage may be due to the nature of a mixed mesophytic forest, where there are not typically large areas of a single dominant species. However, the continued infestation of HWA since 2006 has significantly reduced the presence of eastern hemlock in all three counties. Many of the dense areas that once were abundant
with eastern hemlocks, such as the Hemlock Gardens in the Pine Mountain State Nature preserve, have suffered significant hemlock death and decline.

Although adding and modifying input data had significant impacts on the classification output, modifying the parameters of the model itself also increased the accuracy of the classifications. Increasing the number of trees was able to significantly increase the accuracy, and lowered the proportion of OOB error. Tables 4 and 6 show the decline in overall and individual class OOB error when increasing the number of trees from 150 to 300. Setting the number of trees to 300 was based on the fact that there was no significant change in proportion of OOB error between 250 and 300, meaning that increasing the number of trees and further would not result in a noticeable increase in accuracy but running the model would require increased processing time.

Accuracy assessment for the hemlock presence-absence classification indicates that the model was able to produce an accurate and acceptable product with overall accuracies and class accuracies all at 90% or higher. The hemlock canopy coverage model had lower overall accuracy of 83.6%. Class accuracies from the diagnostics output, however, were all found to be 91% or higher. High accuracy but lower sensitivity, or probability of correctly classifying a feature, may point to the imbalance of the data set. Discussed in the limitations section, the number of higher canopy coverage reference points taken in the field were lower than the number of mid- to low-canopy coverage areas of hemlock presence. For this reason, the MCC score may be a better indication of individual class accuracy for this model. MCC scores for each class were found to be at least 77% or higher, with scores as high as 85%, meaning that if limitations of this data set were taken into consideration the model is still acceptable.
The results of the variable importance tables of each classification were found to have a relatively small range in Gini coefficient. It would be expected that one or a few covariates would have a higher importance, such as the NDVIs or NIR bands in identifying eastern hemlock. However, each explanatory raster in the variable importance table only varied between 2.1 at the lowest and 5.1 maximum importance. The highest variables of importance for classifying hemlock presence were both summer and winter imagery band 9 SWIR (0.945 nm), while the highest variable of importance for the hemlock canopy coverage classification was the DEM layer. The 3 point Gini coefficient range indicates that one variable was not significantly more important than the others, and that the classification of eastern hemlock requires many environmental variables to achieve the accuracy represented in the models.

Limitations

Although hemlock presence and canopy coverage were accurately identified, the limitations of this data should be taken into consideration. Due to the difficulty of acquiring imagery with low cloud coverage, the summer imagery used in the Random Forest model was a composite of imagery across the full 2021 summer season. This allowed for clear imagery that accurately represented the summer spectral signature, but does not account for any change that may have occurred in eastern hemlock decline over the two-month period of 2021. Coupled with the fact that not all areas of the PMWC are accessible due to land ownership or dangerous terrain, the unbalanced dataset required the addition of hemlock-dense training points. Identifying points in the 51-75%, and 76-100% classes helped to bolster the dataset, but did not include information on dieback, infestation, rhododendron, or light penetration that were collected in the field. In the
hemlock canopy coverage classification, this issue was lighted by using the Compensate for Sparse Categories option in the Forest-based Classification tool. Selecting this option ensures that each category represented in the training dataset will be included in each tree.

Additionally, those points that were taken in the field combined with the additional points identified from aerial imagery, may not fully encompass the variation in spectral signature of various hemlock densities across all spectral bands used. The study area encompassed three counties in southern Kentucky, and a significant portion of the 125 mile corridor. The variation of vegetation, environmental factors like those used (slope, elevation, water flow) across the entirety of the PMWC may not be encompassed by the limited reference data that we were able to acquire. The model could be improved in the future by adding additional ground referencing points to balance the dataset and provide a wider range of signatures for the model to train from. Given the added difficulty of rhododendron spectral signatures being obscured by canopy vegetation and eastern hemlock, the data was not robust enough to create an accurate rhododendron classification. Additional training points could prove useful in linking high presence of rhododendron bushes with areas of known hemlock die off.

Future Work

The results of both classifications indicate the usefulness of this study’s method in identifying eastern hemlock. The Sentinel-2 data offers the ability to accurately identify hemlock presence when coupled with in-situ data and environmental variables at a spatial specificity unable to be resolved with Landsat or coarser imagery in previous studies. Because the data is free to the public and is archived every year, it can be a viable
resource for future eastern hemlock and HWA monitoring. For those at the organizations this study partnered with, the resulting classifications can help in targeting treatment regimens to areas of hemlock-dense forest where HWA has spread. Incorporating the use of canopy coverage maps could prove useful in informing dispersion patterns of HWA. Understanding the connectivity of HWA habitable zones allows management teams to take re-infestation into consideration when enacting treatment efforts. If canopy coverage maps were created at some interval across time, patterns of decline and death could be identified. Coupled with the knowledge of where treatment has occurred, effectiveness of the insecticide across the mountain could be inferred.

The results of this study would help to identify both areas of greatest threat, as well as what to expect in terms of changing understory vegetation in areas where hemlock decline is high. A comparison of previous hemlock maps to the results of this classification should show a decline in hemlock presence across Harlan, Letcher, and Bell counties in Kentucky. This product coupled with any data that land managers may have concerning pesticide treatment may aid in the ability to show areas within their jurisdiction that have been sustained due to treatment and those that have not shown continued hemlock growth. By continuing to target healthier areas, hemlock decline could be staved off, possibly preventing the concurrent change to the understory from increased light reaching the lower levels of the forest.

Because the rhododendron shrub is understory vegetation, observed spectral signatures in satellite imagery would be a mixture of the signature of rhododendron and that of the canopy. However, if the canopy were to recede, as is the case with eastern hemlocks dying from HWA infestation, the spectral signature of rhododendron would
outweigh that of the canopy vegetation. Identifying the correlation between eastern hemlock decline stage and rhododendron presence is contingent on the ability of the being able to distinguish between the two types of vegetation. Future work could focus on doing so in order to compare the changes in coverage of rhododendron where eastern hemlock has died off and opened up gaps in the canopy. Identifying this change could help managers in predicting how they should respond to rhododendron growth as hemlock decline continues.

The mission of the PMWC is currently geared toward acquiring lands that help build the corridor. The majority of the land within the bounds of what is considered corridor is owned by either KNLT, the OKNP, or the state. Areas that are still privately owned and operated, however, are less likely to be treated or observed for HWA infestation, which can hinder efforts of those organizations which do treat the trees. Identifying lands which have hemlock presence, and at what canopy coverage that hemlock is present, could help organizations like KNLT to identify private lands of highest priority for acquiring or establishing a partnership with to reduce the future death of eastern hemlocks.
REFERENCES


Sen2Cor. 2021.


CURRICULUM VITA

NAME: Grace Embree

Address: Department of Geographic and Environmental Sciences
200 W Shipp Ave
University of Louisville
Louisville, Ky 40208

DOB: Charleston, West Virginia – May 31, 1997

Education: B.A., Biology
University of Louisville
2018-2020