NASA DEVELOP National Program
Idaho - Pocatello
Summer 2024

Northern Rockies Ecological Conservation

Leveraging Earth Observations to Monitor and Predict Populations of Federally Threatened Whitebark Pine (*Pinus albicaulis*) across the Intermountain West

DEVELOP Technical Report
August 9th, 2024

Hannah Rogers (Project Lead)
Dustin Corbridge
A H M Mainul Islam
Joshua Daniel Carrell

Advisor(s):
Keith Weber, Idaho State University GIS Training and research Center (Science Advisor)
Joseph Spruce, Analytical Mechanics Associates, NASA Langley Research Center (Science Advisor)

Lead:
Kait Lemon (Idaho – Pocatello)

1. Abstract

Whitebark pine (WBP; *Pinus albicaulis*) is an ecologically important species in North America. As a federally listed threatened species, an understanding of WBP habitat, distribution, and health is important for the natural resource managers of the National Park Service, United States Forest Service, Bureau of Land Management, Fish and Wildlife Service, and non-profit organizations such as the Whitebark Pine Ecosystem Foundation. Previous attempts to develop models of WBP habitat suitability and distribution lack confidence in their validity and integrity for these organizations. The updated models of habitat suitability and distribution developed by this study would provide managers with a capability to be employed in the conservation and future research direction for WBP. Thus, we developed a habitat suitability model of WBP at a high spatial resolution (Landsat 9 Operational Land Image-2, National Land Cover Database, NASA Shuttle Radar Topography Mission; 30 m pixels) using a generalized logistic regression with an area under the curve value of 0.754. We extracted spectral reflectance signatures from overlapped ground sample points and Sentinel-2 Multispectral Instrument. The spectral signature analysis indicates WBP is separable from other tree species. We also utilized a visual validation approach and random forest (RF) modeling to separate WBP from limber pine. Through visual validation the RF classifier successfully identified 8 out of 10 WBP trees gathered through ground truth points. Additionally, we achieved an overall accuracy of 91% in our confusion matrix for the distribution model using a dependent validation approach. The derived products from this study allow project partners to assess current suitable habitat and apparent health status in areas of identified WBP occurrence, providing data to aid future research regarding WBP health.

Key Terms

Whitebark Pine, Habitat Suitability, Rocky Mountains, Spectral Signatures, Landsat 9 Operational Land Imager (OLI-2), Sentinel-2

2. Introduction

2.1 Background

Whitebark Pine (WBP; *Pinus albicaulis*) is a long-lived, (~1,200 years) keystone species found in the upper subalpine Rocky Mountain region (Burns & Honkala, 1990; Perkins et al., 1996; Keane et al., 2017). WBP was recently listed as a federally threatened species in 2023, with a 51% estimated reduction from its original range across the United States (Goeking & Izlar, 2018; U.S. Fish and Wildlife, 2022). WBP seeds provide food for Grizzly Bears (*Ursus arctos horribilis*), Red Squirrels (*Sciurus vulgaris*), and Clark's Nutcracker (*Nucifraga columbiana)* and its branches provide shelter for wildlife (Tomback et al., 2014). WBP's extensive root systems help maintain the integrity of steep mountain terrain, and its decline has increased soil erosion and decreased water quality (Tomback & Achuff, 2010; Keane et al., 2017). Previously protected by its high elevation, WBP is now experiencing decline due to a warming climate (Tomback et al., 2014; Buotte et al., 2016; Keane et al., 2017). Eco-stressors such as mountain pine beetle *(Dendroctonus ponderosae)* outbreaks, blister rust fungal disease caused by *Cronartium ribicola*, and wildfires are responsible for high levels of WBP mortality in many areas across its range. These stressors disrupt nutrient flows and reduce reproductive capacity, ultimately exacerbating threats to already eco-stressed trees, destabilizing WBP populations (Tomback D. & Achuff, P, 2010; Buotte et al., 2016).

Synoptic remote sensing is an essential tool in monitoring and managing WBP populations due to its ability to observe remote high-elevation regions that are difficult to access. Kokaly et al. (2003) used spectral signature analysis with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data to identify vegetation types in Yellowstone National Park, providing spectral signature data for WBP identification. Later, Landenburger & Gessler (2008) employed boosted classification trees, and Landsat 7 Enhanced Thematic Mapper plus (Landsat 7 ETM+) data to map WBP in the Greater Yellowstone Ecosystem, providing a validated remote sensing method for mapping WBP stands. Additionally, the High Five Working Group Technical Team developed a conservation probability model using a likelihood of treatment success logical model to predict the success of conservation treatments and prioritize areas of intervention (Jenkins et al., 2022). Integrating aerial and satellite high-resolution remote sensing data with spatial models has enhanced the accuracy of WBP

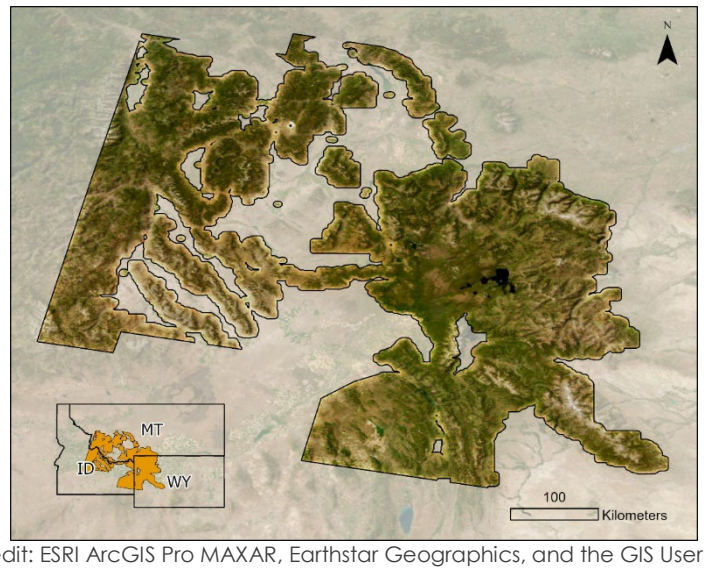
mapping and resulting conservation efforts. For example, Airborne Laser Scanning (ALS) (i.e., Lidar altimetry) has been used to identify forest inventories, resulting in increased accuracy for the classification of forest tree species (Coops et al., 2023). Areas of WBP need to be mapped locally as well as regionally, especially for areas that are only generally mapped (e.g., coniferous forest).

2.2 Project Partners & Objectives

Our NASA DEVELOP project included partnerships with the National Parks Service, Yellowstone Inventory and Monitoring; USDA, US Forest Service (Region 1); Bureau of Land Management, (Salmon Idaho Field Office); US Fish and Wildlife Service, (Wyoming and Montana Field Offices); and the Whitebark Pine Ecosystem Foundation. Our partners are working to conserve WBP ecosystems and are advancing WBP sustainability interests through public engagement, health monitoring programs, restoration projects, and assisted migration efforts of genomic resilient trees (Palmer & Larson, 2014). WBP habitat is between 1,300 and 3,700 meters, and the remoteness of this habitat makes it difficult to easily access WBP stands in the field (U.S. Fish and Wildlife Service, 2022). Our main objective was to develop spatial models that accurately estimate the current extent, accessibility, and probable distribution of WBP within our study area using remote sensing data and Geographic Information System (GIS). We delivered updated models to our partners allowing them to accurately locate areas of WBP occurrence.

2.3 Study Area

The study area (Figure 1) encompasses 114,050 km², and spans parts of Idaho, Montana, and Wyoming. Focusing on the Middle Rockies and Idaho Batholith ecoregions (Omernik, J.M., 1987), landcover is dominated by montane forests and includes 10 management areas of U.S. Forest Service and two national parks. We selected the study period (2021 – 2024) based in part on satellite imagery and data availability. Our project used satellite imagery from Landsat 9 Operational Land Imager-2 (OLI-2; 2023 – Present) and Sentinel-2 Multi-Spectral Instrument (MSI; June 2024). These years are important for examining forests for disturbances like drought, wildfire, and biotic agents that have caused mortality (Buotte et al., 2016; Keane et al., 2016; Tomback D. & Achuff, P, 2010). By focusing on recent years, we also have greater confidence that the satellite imagery and ancillary geospatial data (e.g., occurrence records, disturbance history, etc.) are consistent.



Basemap Credit: ESRI ArcGIS Pro MAXAR, Earthstar Geographics, and the GIS User Community

Figure 1. Map depicting spatial extent of the study area.

3. Methodology

3.1 Data Acquisition

We used Landsat 9 OLI-2 and Sentinel-2 MSI data to identify the distribution of WBP in the study area (Table 1). We obtained Landsat 9 OLI-2 imagery using Earth Explorer hosted by the United States Geological Survey. We obtained Sentinel-2 MSI imagery using Google Earth Engine (Gorelick et al., 2017; GEE).

Table 1
Data Specifications for Satellite Data

Data Product	Data Source	Spatial Resolution	Temporal Resolution	Acquisition Date	Acquisition Method
Landsat 9 Operational Land Imager (OLI)-2	USGS and NASA Earth Observing Systems	30 m	8 days	06/01/2023 - 06/30/2024	Earth Explorer
Sentinel-2 MSI: Multispectral Instrument, Level-2A	European Space Agency	10 m, 20 m, 60 m	5 days	06/01/2024 - 06/30/2024	Google Earth Engine

Ancillary data used as model inputs includes species occurrence records of WBP, represented by point data with latitude and longitude coordinates (Table 2). We collected these occurrence records from iNaturalist & Idaho Department of Fish and Game. We vetted data collected from iNaturalist to incorporate exclusively research-grade data, which indicates that more than one user examined the species visually and agreed the species to be correctly identified, instilling greater confidence in the data. Additionally, we collected elevation data from the NASA Shuttle Radar Topography Mission (SRTM) at a 30 m spatial resolution from which other topographic features of interest (i.e., slope, aspect, etc.) were derived. Spatial resolution of other predictor variables such as, National Landcover Database (NLCD) percent tree canopy, national land cover database, and Landsat 9 OLI-2 derived Normalized Difference Vegetation Index (NDVI) is also 30 m.

Table 2
Ancillary Data used as Model Inputs

Parameter	Data Source	Acquisition Date	Acquisition Method
WBP observation points	iNaturalist	2024	Downloaded from iNaturalist
WBP observation points	Idaho Fish and Game (IDFG)	2024	Idaho Species Diversity Database
Elevation	NASA Shuttle Radar Topography Mission (SRTM)	2024	Downloaded from OpenTopography
Slope	NASA Shuttle Radar Topography Mission (SRTM)	2024	Google Earth Engine Catalog

Aspect	NASA Shuttle Radar Topography Mission (SRTM)	2024	Google Earth Engine Catalog
NLCD Percent Tree Canopy Cover	United States Geological Survey	2021	Google Earth Engine Catalog
National Landcover Database	United States Geological Survey	2021	Google Earth Engine Catalog
Normalized Difference Vegetation Index	NASA Landsat 9 OLI level 2, collection 2, tier 1	2023	Google Earth Engine Catalog
Topographic Diversity	NASA Shuttle Radar Topography Mission (SRTM)	2024	Google Earth Engine Catalog

Furthermore, we acquired tree coordinates for validation and spectral analysis through our own field survey. We selected potential locations of WBP based on elevation and habitat suitability near Bonneville Peak in Idaho. Conducted on 06-01-2024, the field survey involved a 2,600 feet elevation hike to reach the desired location where we found clusters of WBP and limber pine (LP; *Pinus flexilis*). Accurate identification was facilitated by a dichotomous key, examining characteristics such as needle clusters, basil growth patterns, branch dexterity, bark colorization, cone size and deformation; moreover, Chris Earle, Conifer and Wildlife Biologist collaborated on the identification using photographs of the ten WBP and two LP trees located, ensuring the accuracy and validity of our field surveyed trees. We employed ArcGIS Field Maps to record Global Navigation Satellite System (GNSS) coordinates through WGS 1984 Web Mercator (auxiliary sphere) for identified trees (e.g., WBP and LP). To ensure high accuracy, we utilized a Trimble R1 external GNSS receiver, achieving a horizontal positional accuracy of approximately < 1.0 meter while triangulating coordinates with other team members' mobile devices to enhance precision to mitigate potential discrepancies in the coordinates collected for WBP and LP. The collected GNSS points and field notes were subsequently imported into ArcGIS Pro for further analysis, allowing our team to use the coordinates to extract spectral signatures for future analysis (Figure 2).

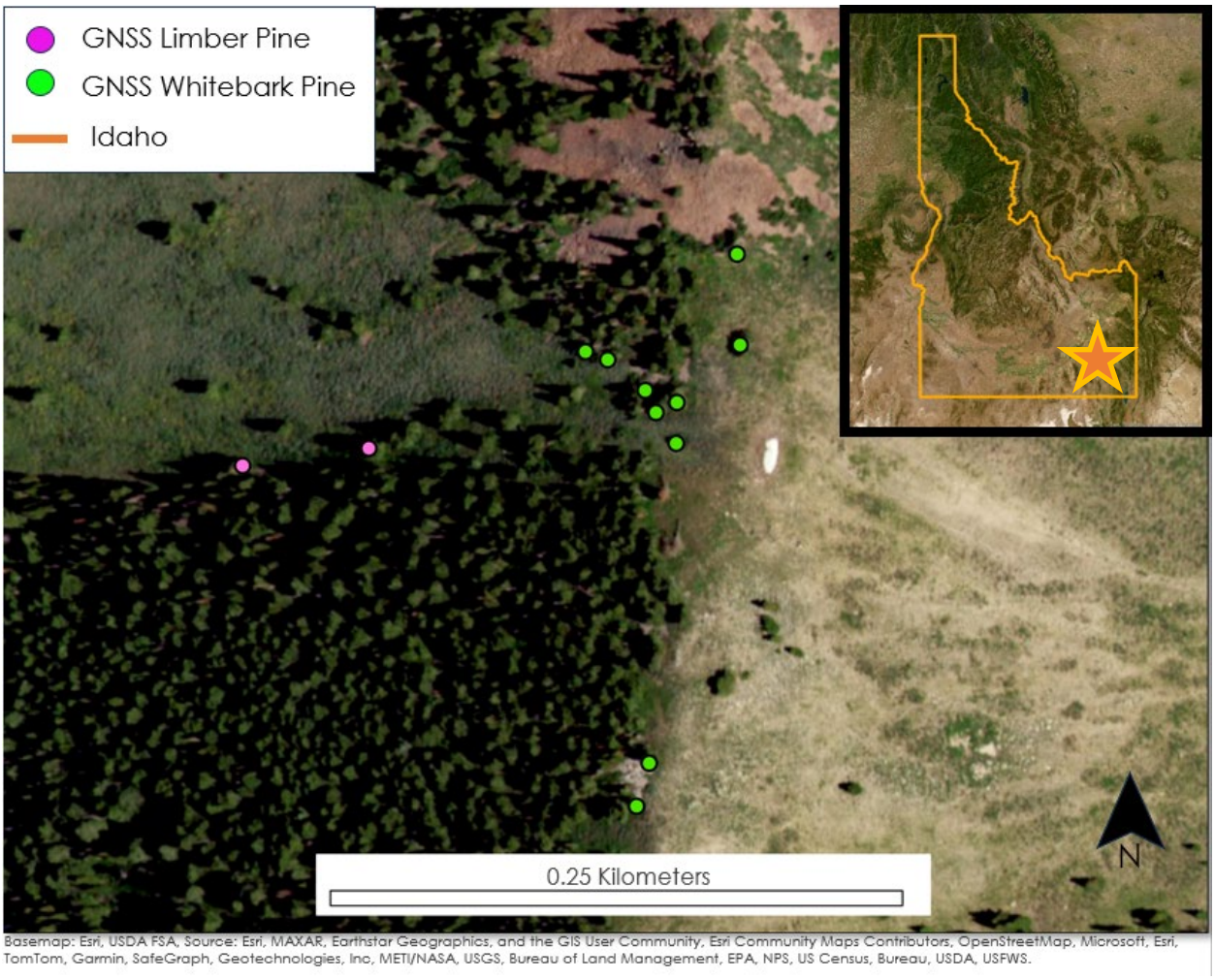


Figure 2. Ground collected points of whitebark pine collected by DEVELOP team (in green) and limber pine (in pink) from Bonneville Peak in Caribou-Targhee National Forest (Starred area)

3.2 Data Processing

3.2.1 Habitat Suitability Modeling

1,083 Whitebark pine occurrence records were collected through citizen science iNaturalist observations and the Idaho Species Diversity Database maintained by Idaho Fish and Game. With occurrences in shapefile format, the two datasets were merged and clipped to the study area using ArcGIS Pro 3.2.2 geoprocessing tools. Additionally, environmental predictor variables for modeling whitebark pine habitat suitability were collected in raster format and clipped to the study area. Collected predictor variables primarily represent topographic and biophysical properties of suitable WBP sites such as elevation, landcover, NDVI from Landsat 9 OLI-2, and percent tree canopy cover (Table 2).

Using the R 4.4.0 programming language, environmental predictors of WBP habitat suitability were uploaded, stacked, and projected to the study area coordinate system using the *rast() and project()* functions from the *terra* R package (Hijmans, 2022). Occurrence data was additionally loaded into the R environment through the *st_read()* function from the *sf* R package and spatially overlaid with the stacked raster dataset of predictors (Pebesma and Bivand, 2023). Habitat suitability models often include the use of presence occurrence and absence data. Without access to recorded species absences, randomized pseudo-absence generation was implemented using the *st_sample()* function with pseudo-absences being created at double the number of

presences within the study area boundaries. Using the *extract()* function from the *terra* package, values from each cell spatially overlapping with an occurrence point were extracted to create an R dataframe used for model training.

3.2.2 Preliminary Accessibility Model

The habitat suitability model output provides representations of where suitable WBP habitat may occur. An important aspect of monitoring WBP health is field data collection, which may be targeted by selecting pixels of high suitable habitat probability. Thus, to improve the field data collection logistics, we developed a model for identifying locations that depict habitat suitability and general distance from roads using ArcGIS Pro ModelBuilder (Figure 3). This model requires the input of the study area, the habitat suitability model, and any line shapefile representing access (e.g., trails, roads). The model creates a hexagonal grid of 1km^2 containing the minimum, maximum, and mean averages of the habitat suitability model. Next, a 5-km buffer is created around road or trail shapefiles and the hexagonal grid of habitat suitability is clipped to it. Through a series of spatial joins and using the *near* function, each grid cell is assigned a distance from the nearest road, the road name, and road quality attributes. The created map enables researchers and field campaigns to rapidly toggle grid cells and search for suitable habitats and information on the nearest roads and trails to access those areas for field data sampling.

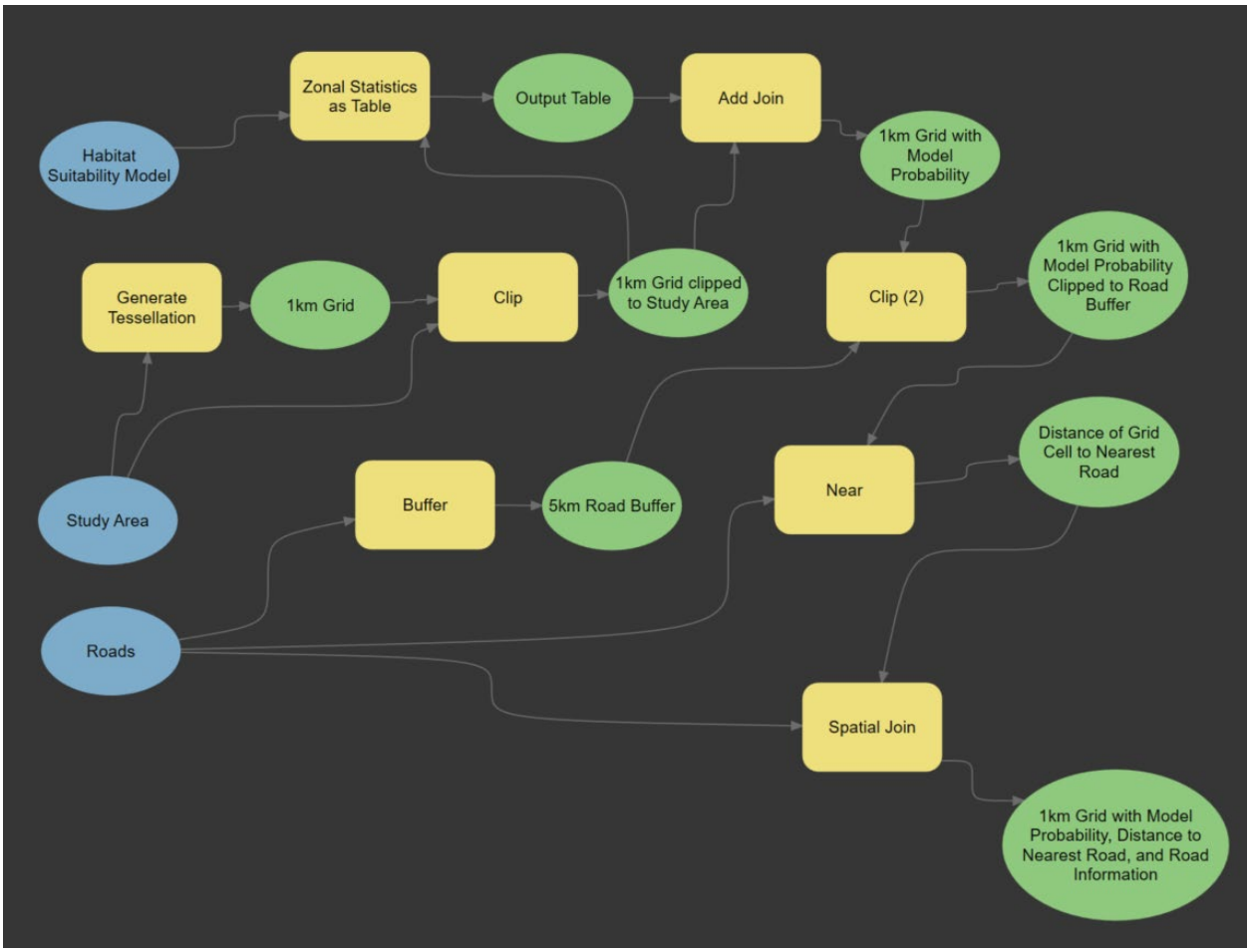


Figure 3. ArcGIS ModelBuilder flowchart for developing field site accessibility map.

3.2.3 Spectral Signatures & Distribution Model

We used Sentinel-2 MSI satellite data from January 1, 2022, to December 31, 2022, to ensure complete coverage and capture seasonal variations for locations of interest in the study area. We initially calculated average spectral signatures of the ground collected WBP (Figure 2; 10 points) and LP (two points) to implement in our GEE-based preliminary WBP distribution model. Later we added Water (one polygon), Urban (three polygon), Grassland (five polygons), and Rocky outcrop (one polygon) to the model as additional training classes. We then extracted spectral signatures for all the variables across Sentinel-2 wavebands. In GEE, the filtering process involves selecting images with less than 10% cloud cover, thereby reducing noise in the imagery. We also masked out cloud shadows, clouds low probability/unclassified, clouds medium probability, clouds high probability, cirrus, and snow/ice using the scene classification layer (SCL). Next, we selected vegetation indices, NDVI and Enhanced Vegetation Index (EVI) to assess vegetation health, assisting in differentiating between various vegetation types. NDVI is a widely used index that measures the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs) (Equation 1; Moreira et al., 2016; Wu et al., 2018; Huete et al., 2002). EVI, on the other hand, optimizes the vegetation signal with improved sensitivity in high biomass regions and improves vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmospheric influences (Equation 2; Moreira et al., 2016; Wu et al., 2018; Kriegler et al., 1969). Using NDVI and EVI implemented together improves sensitivity in both low and high-density vegetation areas, leveraging the strengths of both indices while compensating for their weaknesses, resulting in a more comprehensive and accurate detection of spectral signatures, enhancing our distribution model (Moreira et al., 2016; Wu et al., 2018).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$EVI = G \cdot \frac{NIR - Red}{NIR + C1 \cdot Red - C2 \cdot B + L} \quad (2)$$

Where G is a gain factor, L is the adjustment for canopy background, C values are atmospheric resistance coefficients, and B represents the values from the blue band. Our team utilized normalizedDifference() function to calculate the indices based on the specific bands in the sensor. For the EVI, our team used G = 2.5, C1 = 6, C2 = 7.5, and L = 1.

Later we incorporated topographic data by loading a Digital Elevation Model (DEM) to mask out regions above 1,828.8 m. This approach helped us account for terrain influences on vegetation and WBP distribution preferences while reducing noise and calculation fluctuations in our model. Then we applied a random forest (RF) model to classify the training variables. We chose the RF classifier for its effectiveness in handling large datasets and its ability to manage the complexities of spectral data (Breiman, 2001). The RF classification process uses spectral bands and vegetation indices to categorize each pixel from Sentinel-2 imagery within the area of interest (AOI) into a specific output class or category. The layers provide a visual representation of the distribution of various targeted vegetation types, based on supervised training data on the occurrence of such land cover types.

3.3 Data Analysis

3.3.1 Habitat Suitability Modeling

A logistic regression model (equation 3) was fit using the glm() (Generalized Linear Model) function in R. The model employs a logit link function, which is designed to model a binary response (e.g., suitable vs. non-suitable habitat) using several predictor variables, and is mathematically derived as:

$$logit(P(A)) = \beta 0 + \beta 1 \cdot aspect + \beta 2 \cdot elevation + \beta 3 \cdot landcover + \beta 4 \cdot slope + \beta 5 \cdot NDVI + \beta 6 \cdot CanopyCover \quad (3)$$

where logit(P(Y=1)) is the log odds of the probability of the response variable being 1 (e.g., presence). The model coefficients (β0 −β6) are estimates derived from maximum likelihood estimation, which finds the parameter values that maximize the probability of the observed data given the model.

Model threshold optimization was identified through the *optimal.thresholds()* function, which was used to determine the optimal threshold value for binary classification and employs a MaxKappa methodology which aimed to maximize Cohens Kappa (Cohen, 1968), a statistic that measures agreement for categorical terms (e.g., presence vs. absence). Model accuracy metrics were developed using the *presence.absence.accuracy()* function which calculates commonly identified model accuracy metrics like sensitivity, specificity, the area under the curve and the true skill statistic. Additionally, model cross-validation was performed using the *CV binary()* function, which performs cross-validation for a binary classification model and employed 5 folds. Lastly, model predictions were fit and spatially created using the *predict()* function using the logistic regression model shown above.

3.3.2 Spectral Signatures & Distribution Model

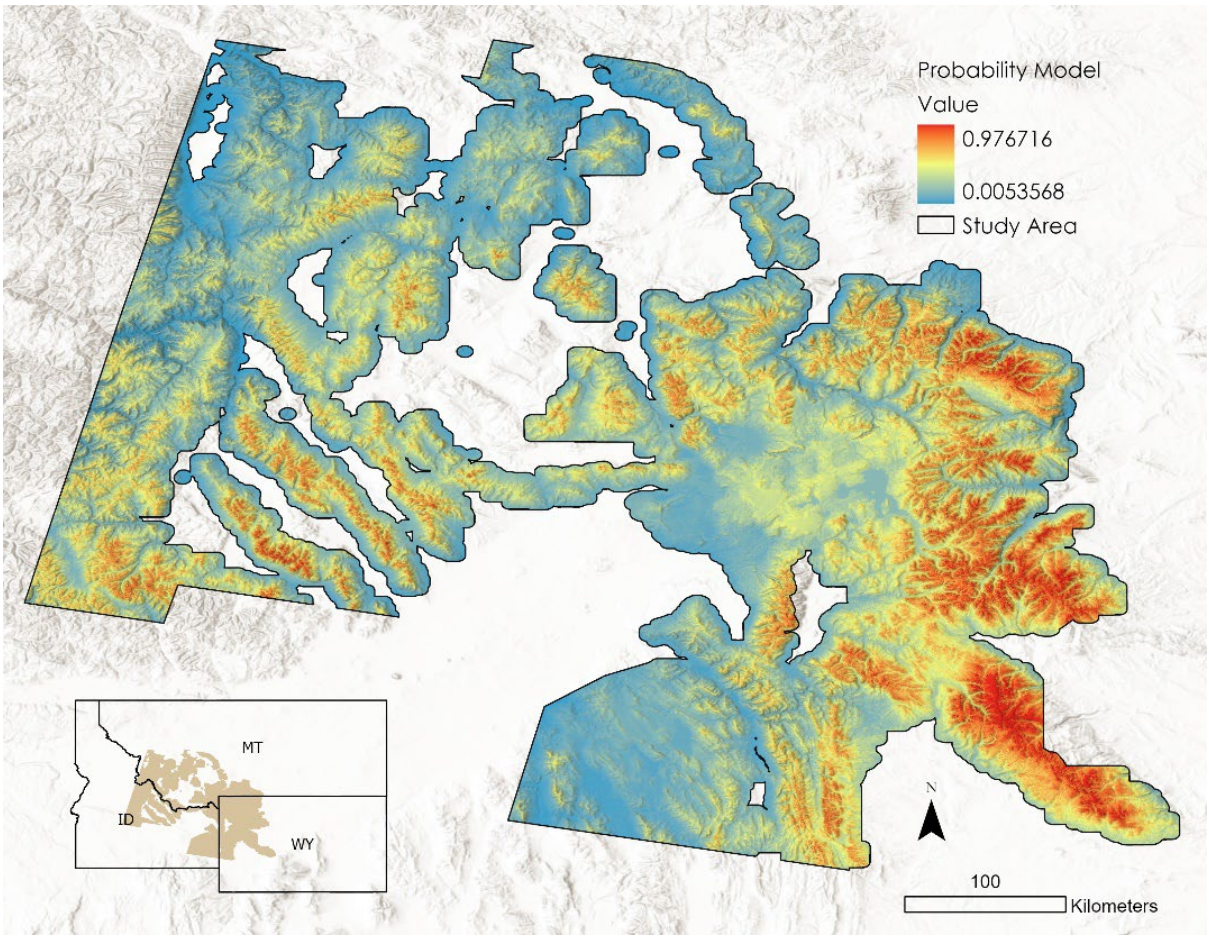
We used a visual validation approach along with random forest modeling to explore the distribution and accuracy of classified WBP compared to other classes. We used NDVI and EVI from applicable Sentinel-2 spectral bands to classify the variables. Then spectral signatures were combined with the indices' scores to develop the training bands. Finally, we used a RandomForest() (Google Earth Engine, 2022) classifier to classify the variables. We also calculated a confusion matrix to understand the accuracy of the model. The confusion matrix compares our classifications and ground truth data, providing insights into the model's performance. The confusion matrix calculates overall accuracy, user's accuracy, and producer's accuracy which helps assess the performance of the model.

4. Results & Discussion

4.1 Habitat Suitability Modeling & Preliminary Accessibility Model

A generalized linear model was used to develop a single probability raster data depicting suitable habitat of WBP across the study area (Figure 4). Suitability was assigned as a percentage to each pixel, ranging from 0.01% to 97.67%, and varied considerably across the study area. Higher suitability was highly favored among higher elevations where rockier topography and steeper slopes existed. Visually, higher suitability was also biased towards the southern and western sections of the study area, most likely due to limitations in occurrence records located in the northern extents of the study area. In addition to the probability raster, the model produced specific metrics commonly used to define the performance of habitat suitability models. Model sensitivity and specificity were reported as 0.704 and 0.707, respectively. Sensitivity, also known as the true positive rate, measured how well the model identified positive cases (i.e., suitable habitat). Specificity, the true negative rate, described how well the model identified negative cases, or non-suitability. Another principal metric in determining the modeling predictive performance was the area under the curve (AUC). The model AUC was reported as 0.756 out of 1, performing higher than a random classifier (e.g., AUC of 0.5). Lastly, the true skill statistic (equation 4) was reported as 0.411, suggesting that the model performed better than random chance:

$$\text{Sensitivity} + \text{Specificity} - 1 = 0.719 + 0.689 - 1 = 0.411 \quad (4)$$



Basemap Credit: ESRI ArcGIS Pro MAXAR, Earthstar Geographics, and the GIS User Community

Figure 4. Habitat suitability model of WBP within the study area. The probability gradient is visualized in a color gradient of blue to red representing low to high suitability, respectively.

Using ArcGIS bivariate symbology, we created a map to visualize the relationship between habitat suitability and proximity to access routes like trails and roads (Figure 5). This symbology approach allowed us to simultaneously display two variables: the habitat suitability scores and the distance from the nearest road or trail. By applying a color gradient, the map effectively highlights areas with high habitat suitability and varying levels of accessibility. This dual representation aids in quickly identifying optimal field sampling locations, where high suitability habitats are within manageable distances from access routes. The bivariate map enhances our understanding of the spatial distribution and accessibility of suitable WBP habitat, providing a valuable tool for planning and conducting field data collection.

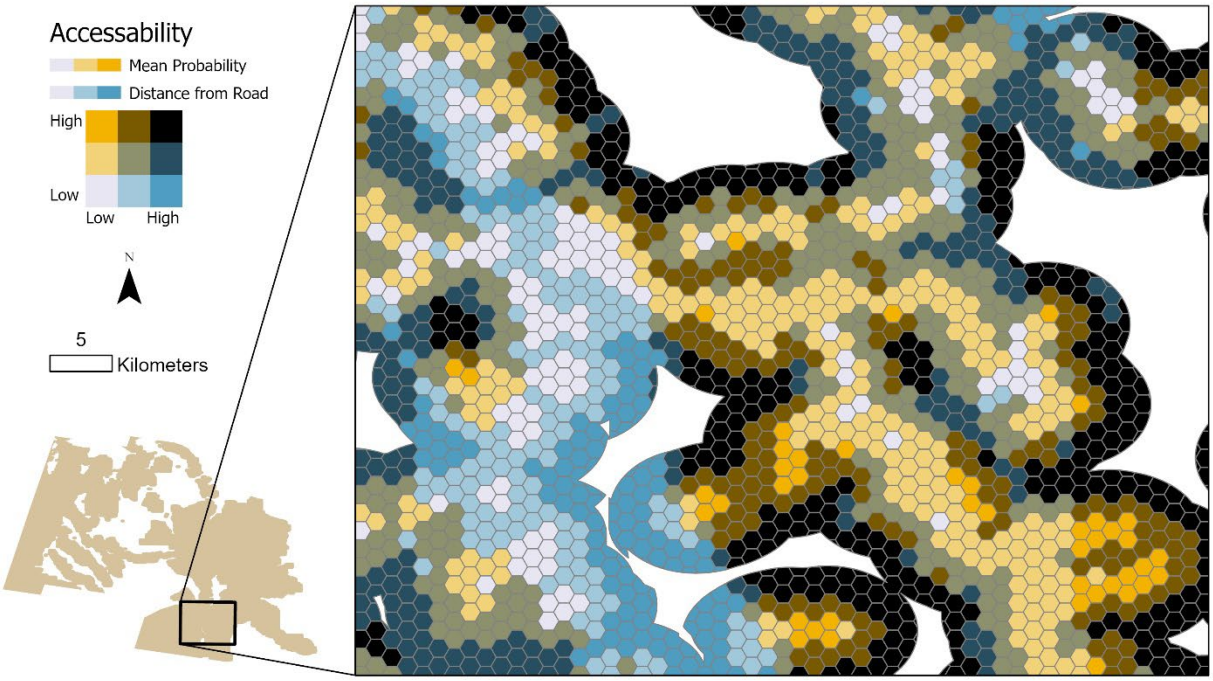


Figure 5. Accessibility map with a bivariate symbology depicting mean habitat suitability probability and distance from nearest road per 1km hexagon.

4.2 Spectral Signatures

We extracted the spectral signatures of WBP and LP, which is the spectral reflectance across the Sentinel-2 wavebands (Figure 6). We used an image collection from the most recent month (June 2024) for spectral analysis. The curve was fitted using the smoothing functions in GEE. Spectral variations and differences are important to develop spectral indices and distinguish between tree species. We found the average spectral reflectance of WBP is higher than LP in visible and shortwave infrared regions while lower in near infrared regions. The reflectance of WBP and LP span extends from 0.05 to 0.35 (Figure 6, Axis Y). Hence, the differences in these wavebands can be used to distinguish WBP and LP.

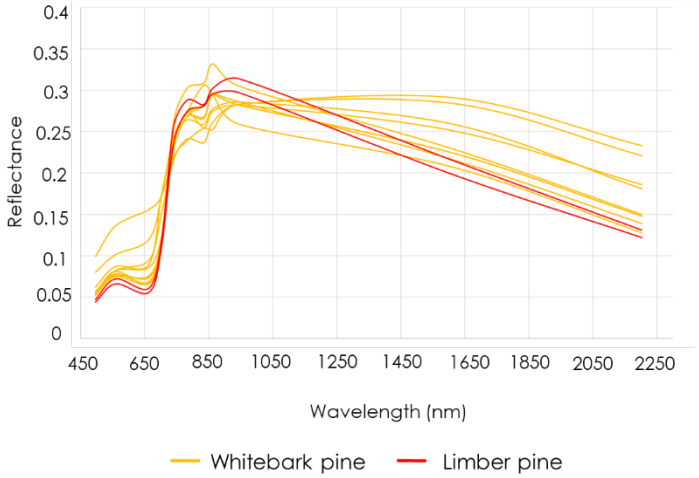


Figure 6. Spectral signatures of WBP (in orange) and LP (in red).

4.3 Preliminary Distribution Model

We tested our model in a selected area within our ground truth collection site. Our Confusion matrix derived through dependent validation, identified the overall accuracy of the model at 91% (Table C1) with WBP user's accuracy of 80% and producer's accuracy of 100%. Considering elevation and other environmental variables, our habitat suitability distribution model identified the ground-truth region as a high probability area. Through Visual Validation of our gathered WBP ground-truth points, we successfully identified 8 out of 10 WBP samples (Figure 7). Additionally, visual validation from our model identified pixels where IDFG also identified WBP points (Figure A1). Furthermore, with the unknown accuracy of the GNSS coordinates of WBP from IDFG, a percentage accuracy cannot be attached to the visual accuracy of the occurrence. Consequently, the model identified clusters of WBP in unexplored or unreached areas of WBP occurrence, suggesting potential new areas to apply ground-truth data collection points for the future (Figure B1).

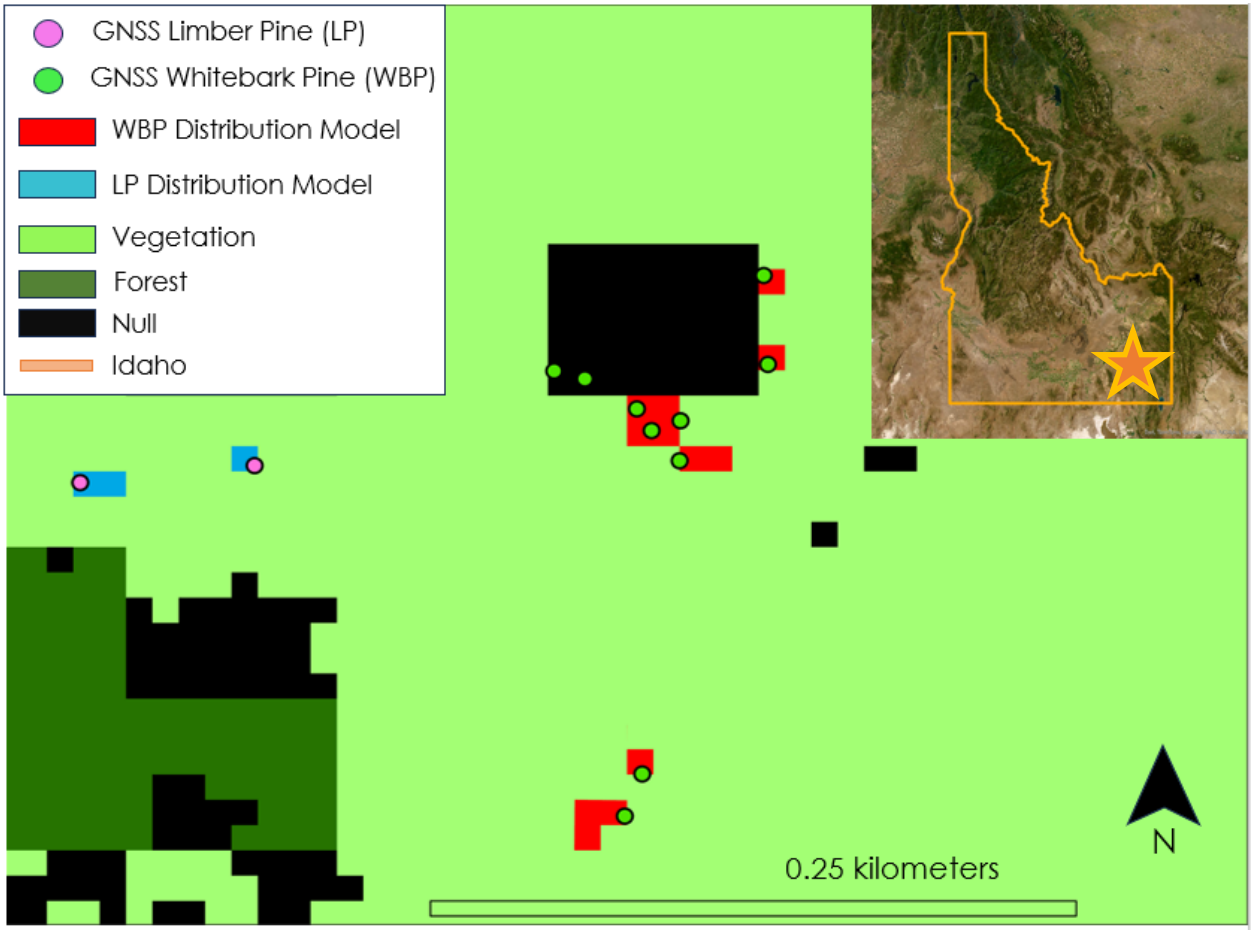


Figure 7. Random forest derived validation of classified WBP (in red) and LP (in blue) with the ground truth points (WBP- green and LP- pink) from Bonneville Peak in Caribou-Targhee National Forest (starred area). This classification uses Sentinel-2 Imagery at a 10 m pixel resolution.

4.4 Errors & Uncertainties

Our team did not have enough accurate WBP sites to validate the distribution model through independent validation. We trained and tested our machine learning model to separate WBP from other classes only by using 10 WBP and 2 LP ground collected sites. The number of samples is not adequate to develop a robust and reproducible machine learning model. The more samples the better a machine learning model can "learn"

patterns within the data. Hence, less samples from a very specific site can lead to a biased or overfit model. For example, we found 8 out of 10 WBP points were correctly classified (80%). From a statistical point of view, that's a good level of agreement. If the model is not tested with many additional field samples, it will be difficult to say if the model correctly estimates the accuracy or not. Also, collecting samples only from very similar areas (e.g. many samples from ~1600 m and less samples from ~3000 m) can bias the distribution of WBP. Such collection can also lead to periodicity problems of sample collection which can introduce bias in the estimation or distribution of the species. A general guideline is to have a specified confidence interval of 50 samples per species to perform accuracy assessment for any classification study (Congalton, 1991; Brogaard, S., and Ólafsdóttir, R. 1997). Exploring the differences in the pattern of spectral signatures for co-occurring similar tree species is another way to assess potential spectral separability between tree species (Rossi & Gholizadeh, 2023). In that case, a substantial number of additional samples may be needed to derive representative spectral signatures of targeted species and subsequently, reliable land cover classification models.

The number and accuracy of occurrence records can also influence habitat suitability results. During model training, occurrence point data extracts information from overlapping variables and when these points are not accurately placed in space, they may extract information that is not associated with the habitat suitability of that given species. Additionally, while iNaturalist provides "research grade" citizen science data, there is potential for a species to be incorrectly identified visually by multiple individuals. Additional models, such as random forest or gradient boosting machines working with the generalized linear model, may provide more insights into WBP habitat suitability. Further incorporation of downscaled climate, soil, and topographic predictor variables and the addition of new occurrence records may improve the habitat suitability model's predictive power.

4.5 Feasibility & Partner Implementation

Our WBP habitat suitability model, accessibility model, and preliminary distribution model are all applicable for our partners to use in the field. Leveraging our models will effectively and efficiently allow conservation efforts to identify the most accessible habitats where WBP is likely to exist. By overlaying the habitat suitability and distribution models, partners can efficiently locate potential hotspots or clusters of WBP while using the accessibility map to find the optimal locations to pursue research in the field. Enhancing these models with ground truth data collection will refine the preliminary distribution map, improving accuracy in WBP identification clusters or individual trees with high statistical accuracy. Collecting WBP ground truth data and incorporating a quality rating of trails will ensure accurate, up-to-date information on the accessibility of WBP or any other conservation endeavors moving forward. As our partners utilize our deliverables and update newly discovered WBP locations, a second term DEVELOP project can use the resulting data to build more robust and accurate models and maps moving forward; moreover, incorporating additional ecological stressors and identifiers will enhance conservation strategies in the future. The next term can also focus on WBP occurrence map based on spectral data which can be used to assess the transition between stress induced disturbance and post-disturbance recovery.

5. Conclusions

Our study sought to develop a habitat suitability, distribution, and accessibility model for WBP using remote sensing techniques and machine learning algorithms by integrating spectral data from Landsat-9 OLI-2 and Sentinel-2 MSI. Our habitat suitability model utilized a generalized logistic model predicting areas of suitable habitat across our study area. The logistic regression model incorporated topographic and biophysical variables and processed these variables using R programming language, creating a probability raster map indicating habitat suitability for WBP. This estimated WBP habitat probability map ranged from .01% to 97.67%, highlighting higher suitability in higher elevations with rock topography, especially in the southern and western sections of our study area. Further, development of downscaled climate, soil, and topographic predictor variables along with the addition of new occurrence records will improve the habitat suitability model's predictive power. Additional models such as gradient boosting machine working with the generalized linear model may provide a means to improve WBP habitat suitability predictions.

We developed our distribution model using spectral signatures extracted from our ground truth data collection points of WBP, allocated from ArcGIS Field Maps, using Sentinel-2 imagery. Identifying the unique spectral reflectance patterns of WBP gave a distinguishing difference between a comparable conifer, the LP; furthermore, vegetation canopy greenness indices, such as NDVI and EVI, were utilized to assess vegetation health within predicted WBP occurrence sites. Additionally, we used the RF classifier with spectral signatures and other land cover types such as water, urban areas, grasslands, and rocky surfaces. We applied the RF classifier to the Sentinel-2 imagery filtered for minimal cloud cover; furthermore, integrating a DEM to account for terrain influences created less noise in the model. The RF classifier successfully identified 8 out of 10 WBP trees gathered through ground truth points. Our Confusion matrix derived through dependent validation, identified the overall accuracy of the model at 91% with WBP user's accuracy of 80% and producer's accuracy of 100%. Visual validation further supported the model's capability to correctly identify WBP clusters aligned with IDFG's already collected data points, showing a preliminary distribution model success. Consequently, our accessibility model enhances the practical applicability of our findings, ensuring that conservation efforts can focus on areas that are both ecologically significant and accessible while integrating our habitat suitability and distribution models.

Finally, while our study provides a substantial advancement in WBP habitat mapping, ongoing efforts to collect more comprehensive ground truth data and refine spectral signatures are crucial for future work, which will create more resilient and accurate models moving forward.

6. Acknowledgements

- **Node Lead:**
  - Kait Lemon – Idaho - Pocatello
- **Advisors:**
  - Keith Weber – Idaho State University | GIS Director & NASA DEVELOP Science Advisor
  - Joe Spruce – Analytical Mechanics & Associates, NASA Langley Research Center
- **Partners:**
  - Julee Shamhart – Executive Director |Whitebark Pine Ecosystem Foundation
  - Bob Keane – Board Assoc. Chair & Research Ecologist | Whitebark pine Ecosystem Foundation
  - Melissa Jenkins – Board Secretary & Forest Silviculturist | Whitebark Pine Ecosystem Foundation
  - Joe Fortier – Remote Sensing Coordinator | USDA, USFS Region 1
  - Erin Shanahan – National Park Service, Yellowstone Inventory and Monitoring
  - Kristin Legg – National Park Service, Yellowstone Inventory and Monitoring
  - Hannah Alverson – Bureau of Land Management, Salmon Idaho Field Office
  - Jim Lindstrom – Cartographer | US Fish and Wildlife Service, Wyoming Ecological Services Field Offie
  - Destin Harrell – Biologist | US Fish and Wildlife Service, Wyoming Ecological Services Field Office
  - Rachel Arrick – Ecologist | US Fish and Wildlife Service, Wyoming Ecological Services Field Office
  - Laura Strong – US Fish and Wildlife Service, Montana Ecological Services Field Office

- **Collaborators:**
  - Dr. Chris Earle – Wildlife Biologist | The Gymnosperm Database
  - Diana Tomback – University of Colorado | Department of Integrative Biology
  - Jim Strickland – Botany Director |Idaho Fish and Wildlife

- Liam Megraw – Environmental Scientist and GIS Specialist | Virginia Department of Conservation and recreation

This material contains modified Copernicus Sentinel data (2023, 2024), processed by ESA.
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.
This material is based upon work supported by NASA through contract 80LARC23FA024.

All NASA LP DAAC data products will be removed from EarthExplorer and M2M on August 30, 2024. Users are encouraged to visit the LP DAAC website to familiarize themselves with alternative search and download options. A list of available tools for each dataset are included under the "Access Data" button on each dataset's DOI landing page, these landing pages can be found using the Search Data Catalog. Please see the news announcement to learn more: https://lpdaac.usgs.gov/news/removal-of-nasa-lp-daac-products-from-usgs-earthexplorer-and-machine-api-on-august-30-2024/ please email lpdaac@usgs.gov with any questions.

7. Glossary

Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) - An optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands. The main objective of AVIRIS is to measure, identify, and monitor constituents of the Earth's surface and atmosphere based on molecular absorption and particle signature

Airborne Laser Scanning (ALS) - A remote sensing technique that uses a laser scanner attached to an aircraft to create 3 dimensional models of earth's surface. Also commonly known as LiDAR (Light Detection and Ranging)

Band - A wavelength range in the spectrum of reflected or radiated electromagnetic (EM) energy to which a remote sensor is sensitive. Sensors collect data from a band and store the data in a file or a portion of a file devoted to that range. These files ae also referred to as bands or spectral bands

Earth observations - Satellites and sensors that collect information about the Earth's physical, chemical, and biological systems over space and time

Ecoregion - A major ecosystem defined by distinct geography and receiving uniform solar radiation and moisture

EVI - Enhanced Vegetation Index. It optimizes the vegetation signal with improved sensitivity in high biomass regions and improves vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmospheric influences

GEE - Google Earth Engine is a catalog of satellite imagery

MSI - Multispectral Instrument. We used Sentinel-2 MSI imagery for spectral signature analysis

NDVI - Normalized Difference Vegetation Index. It is a widely used index that measures the difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs)

OLI - Operational Land Imager. We used Landsat 9 OLI-2 imagery for developing study area and NDVI for habitat suitability model

RF - Random Forest is a machine learning model which provides predictions based on decision trees

Spectral Reflectance - Spectral reflectance is the ratio between the energy reflected by the surface and energy incident on the surface. It is measured as a function of the wavelengths

8. References

Brogaard, S., & Ólafsdóttir, R. (1997). *Ground-truths or ground-lies? Environmental sampling for remote sensing application exemplified by vegetation cover data* (Lund Electronic Reports in Physical Geography No. 1). Department of Physical Geography, Lund University. https://typeset.io/pdf/ground-truths-or-ground-lies-environmental-sampling-for-wx4qhpkc7s.pdf

Buotte, P.C., Hicke, J.A., Preisler, H.K., Abatzoglou, J.T., Raffa, K.F. and Logan, J.A. (2016). Climate influences on whitebark pine mortality from mountain pine beetle in the Greater Yellowstone Ecosystem. Ecological Applications, *26*(8), 2507-2524. https://doi.org/10.1002/eap.1396

Breiman, L. Random Forests. *Machine Learning 45*, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Burns, R. M., & Honkala, B. H. (Technical Coordinators). (1990). *Silvics of North America: Volume 1. Conifers* (Agriculture Handbook 654). United States Department of Agriculture, Forest Service. https://www.srs.fs.usda.gov/pubs/misc/ag_654_vol1.pdf

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin. 70*(4), 213–220. doi:10.1037/h0026256. PMID 19673146.

Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35-46. doi:https://doi.org/10.1016/0034-4257(91)90048-B)

Coops, N. C., Tompalski, P., Goodbody, T. R. H., Achim, A., & Mulverhill, C. (2023). Framework for near real-time forest inventory using multi source remote sensing data. *Forestry: An International Journal of Forest Research, 96*(1), 1-19. https://doi.org/10.1093/forestry/cpac015

European Space Agency. (2015). Sentinel 2 Multispectral Imagery (MSI) / Level-2 Surface Reflectance [Dataset]. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/processing-levels/level-2

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.E. (2007). The shuttle radar topography mission: *Reviews of Geophysics*, v. 45, no. 2, RG2004. https://doi.org/10.1029/2005RG000183.

Hijmans, R. (2022). *raster: Geographic data analysis and modeling* (Version 3.5-15) [R package]. Retrieved from https://CRAN.R-project.org/package=raster

Goeking, S. A., & Izlar, D. K. (2018). *Pinus albicaulis* Engelm. (Whitebark Pine) in mixed-species stands throughout its US range: Broad-scale indicators of extent and recent decline. *Forests, 9*(3), 131. https://doi.org/10.3390/f9030131

Google Earth Engine. (2022). ee.Classifier.smileRandomForest [Source code]. Retrived July 30, 2024 from https://developers.google.com/earth-engine/apidocs/ee-classifier-smileRandom

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment.* https://doi.org/10.1016/j.rse.2017.06.031

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & L.G. Ferreira. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of the Environment 83*(1-2), 195-213. https://doi.org/10.1016/S0034-4257(02)00096-2

Jenkins, M. J., Hebertson, E. G., & Page, W. G. (2022). Conservation Probability Models for Whitebark Pine. *Journal of Forestry*, *120*(4), 381-389. https://www7.nau.edu/mpcer/direnet/publications/publications_j/files/Jenkins_MJ_Hebertson_E_Page_W_Bark_beetles_fuels_fires.pdf

Keane, R.E., Holsinger, L.M., Mahalovich, M.F., & Tomback, D.F. (2017). *Restoring whitebark pine ecosystems in the face of climate change*. Gen. Tech. Rep. RMRS-GTR-361. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 123 p. https://www.climatehubs.usda.gov/sites/default/files/rmrs_gtr361.pdf

Keane, R.E., Holsinger, L.M., Mahalovich, M.F., & Tomback, D.F. (2016), Evaluating future success of whitebark pine ecosystem restoration under climate change using simulation modeling. *Restoration Ecology*, 25: 220-233. https://doi.org/10.1111/rec.12419

Kokaly, R. F., Despain, D. G., Clark, R. N., & Livo, K. E. (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment*, *84*(3), 437-456. https://doi.org/10.1016/S0034-4257(02)00197-6

Kriegler, F., Malila, W., Nalepka, R., & Richardson, W. (1969). Preprocessing transformations and their effect on multispectral recognition. *Proceedings of the 6th International Symposium on Remote Sensing of Environment*. Ann Arbor, MI: University of Michigan, 97-131.

Landenburger, L., & Gessler, P. (2008). Whitebark Pine Distribution Mapping Using Boosted Classification Trees and Landsat ETM+ Data. *Remote Sensing of Environment, 112*(5), 2324-2331. https://remotesensing.montana.edu/documents/Landenburger_et_al_2008.pdf

Moreira, E., Valeriano, M., Sanches, I., & Formaggio, A. (2016). Topographic effect on spectral vegetation indices from Landsat TM data: Is topographic correction necessary? *Boletim de Ciências Geodésicas, 22*(1), 95–107. https://doi.org/10.1590/S1982-21702016000100006

NASA Shuttle Radar Topography Mission (SRTM; 2013). Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography. https://doi.org/10.5069/G9445JDF. Accessed: June 2024.

Omernik, J.M. (1987). Ecoregions of the conterminous United States. Map (scale 1:7,500,000). *Annals of the Association of American Geographers* 77(1), 118-125.

Palmer, C., & Larson, B. M. H. (2014). Should We Move the Whitebark Pine? Assisted Migration, Ethics and Global Environmental Change. *Environmental Values*, *23*(6), 641–662. http://www.jstor.org/stable/43695193

Pebesma, E., & Bivand, R. (2023). Spatial Data Science: With Applications in R. *Chapman and Hall/CRC*. https://doi.org/10.1201/9780429459016

Rossi, C., & Gholizadeh, H. (2023). Uncovering the hidden: Leveraging sub-pixel spectral diversity to estimate plant diversity from space. *Remote Sensing of Environment*, *296*, 113734. doi:https://doi.org/10.1016/j.rse.2023.113734.

Tomback, D.F., & Achuff, P. (2010). Blister rust and western forest biodiversity: ecology, values and outlook for white pines. *Forest Pathology, 40*: 186-225. https://doi.org/10.1111/j.1439-0329.2010.00655.x

Tomback, D. F., Chipman, K. G., Resler, L. M., Smith-McKenna, E. K., & Smith, C. M. (2014). Relative abundance and functional role of whitebark pine at treeline in the Northern Rocky Mountains. *Arctic, Antarctic, and Alpine Research, 46*(2), 407–418. Retrieved from https://bowvalleynaturalists.org/wp-content/uploads/Tomback-et-al-2014.pdf

U.S. Geological Survey. Landsat 9 Operational Land Imager (OLI-2) Collection 2, Level 2, Tier 2 Surface Reflectance [Dataset]. Earth Engine Catalog/USGS. Retrieved June 2024, from https://doi.org/10.5066/P9OGBGM6

U.S. Fish and Wildlife Service. (2022). Endangered and Threatened Wildlife and Plants; Threatened Species Status with Section 4(d) Rule for Whitebark Pine (Pinus albicaulis). *Federal Register, 87*, 76882 – 76917 https://www.govinfo.gov/content/pkg/FR-2022-12-15/pdf/2022-27087.pdf

Wu, M., Yang, C., Song, X., & et al. (2018). Monitoring cotton root rot by synthetic Sentinel-2 NDVI time series using improved spatial and temporal data fusion. *Scientific Reports, 8*, Article 2016. https://doi.org/10.1038/s41598-018-20156-z

9. Appendix

Appendix A.

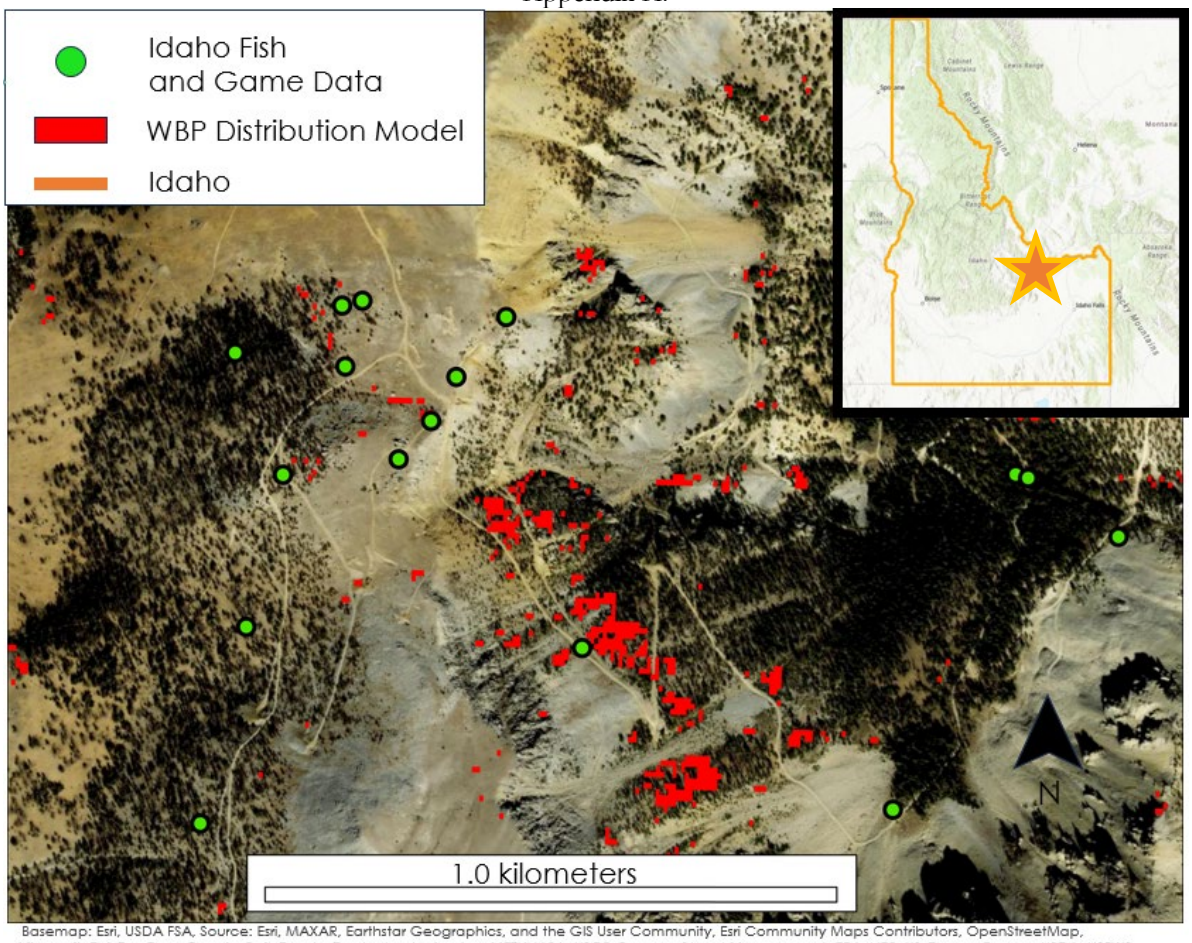


Figure A1. A visual validation of classified WBP (in red) with the IDFG data points of WBP (in green) from Lemhi Range in Spring Mountain.

Appendix B.

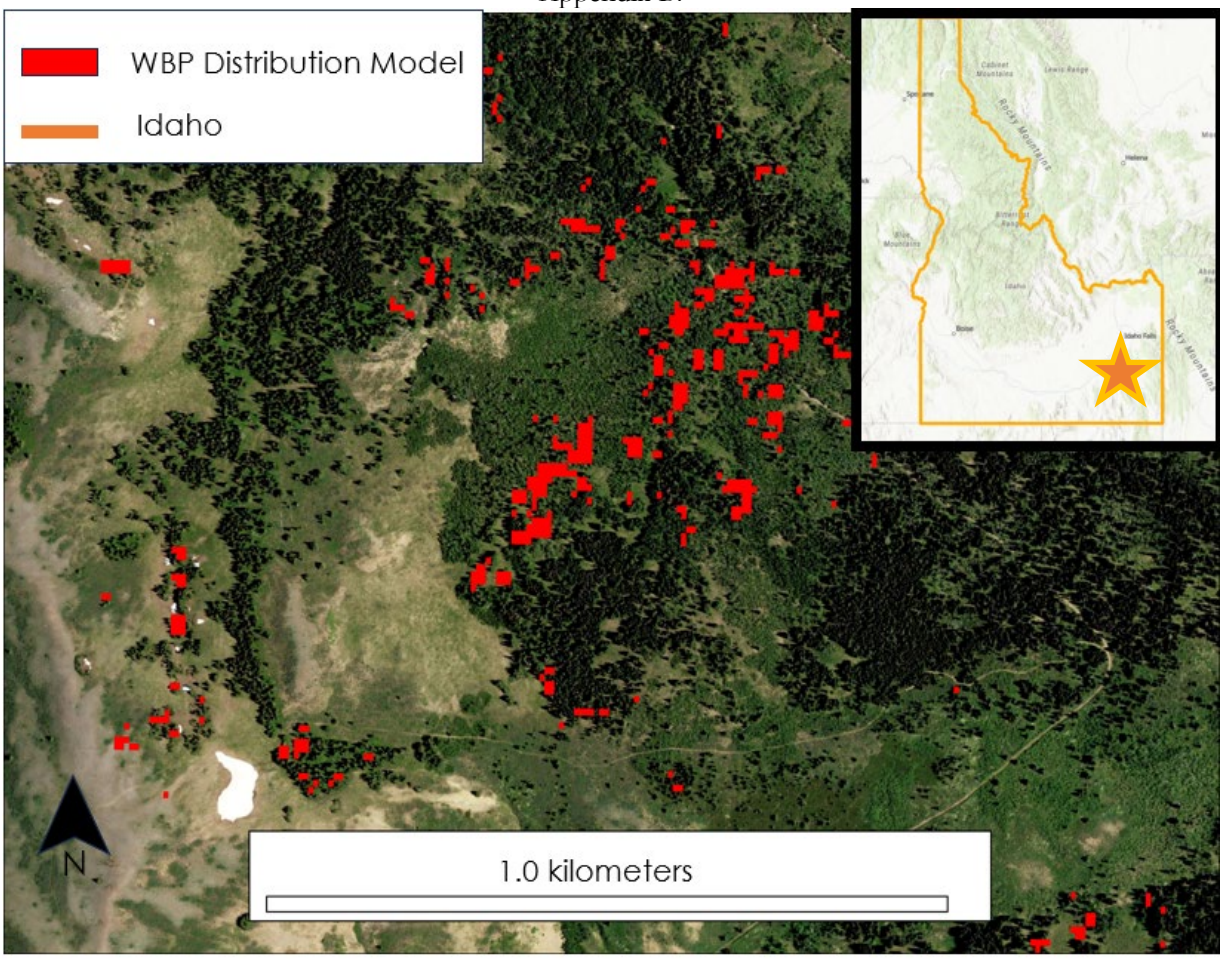


Figure B1. Possible whitebark pine clusters in Caribou-Turgee National Forest

Appendix C.

Table C1
Confusion Matrix

	WBP	**LP**	**Others**	**Row Total**
WBP	**8**	0	2	10
LP	0	**2**	0	2
Others	0	0	**10**	10
Column Total	8	2	10	**22**

Total sample = 22
WBP sample = 10
LP sample = 2
Others = 10

Overall accuracy (OA) = $\frac{8+2+10}{22} \times 100 =$ = 91%

WBP producer's accuracy (PA) = $\frac{8}{8} \times 100 =$ 100%

WBP user's accuracy = $\frac{8}{10} \times 100 =$ 80%