New York Ecological Forecasting

Utilizing NASA Earth Observations to Map Ash Distribution and Inform Emerald Ash Borer Control

 **Technical Report**

Final – August 11th, 2022

Liam Megraw (Project Lead)

Jesse Carlson

Samantha Kelly

Stefanie Mendoza

***Advisors:***

Juan Torres-Pérez, NASA Ames Research Center (Science Advisor)

Dr. Mark Friedl, Boston University (Science Advisor)

Joseph Spruce, Science Systems & Applications, Inc. (Science Advisor)

***Fellow:***

Britnay Beaudry (Ames Fellow)

# 1. Abstract

Since their first sightings in the U.S. in 2002, emerald ash borer beetles (*Agrilus planipennis*; EAB) have killed millions of native ash (*Fraxinus* spp.) trees across 35 states. Infected ash stands frequently exhibit complete mortality, with the predicted result being the functional extinction of native ash in U.S. forests. In August of 2020, EAB was discovered in the 6.1-million-acre Adirondack Park. The team’s partners at the Adirondack Park Invasive Plant Program (APIPP) desired ash tree distribution and EAB susceptibility information to help improve EAB bio-control efficiency and apply the methodology to future invasive programs. To assist, the team mapped ash tree distribution using NASA Earth observations from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Shuttle Radar Topography Mission (SRTM), plus hyperspectral imagery from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Field data from the Monitoring and Managing Ash (MaMA) project, iMapInvasives and iNaturalist databases, and the New York State Department of Environmental Conservation (NYSDEC) provided ground truthing for mapping and modeling. Results indicate that for ash detection, the team’s Spectral Angle Mapping (SAM) hyperspectral classification is slightly more sensitive but less accurate than multispectral Random Forest (RF) classification, though neither method was above a ~20% detection rate. End products include maps of ash extent derived from both imagery types, a model forecasting future spread scenarios based on current EAB presence, and outreach materials. These products inform APIPP’s management decisions and facilitate public awareness of EAB’s threat to communities within the region.

**Key Terms**

remote sensing, AVIRIS, Landsat, random forest, spectral angle mapping, spectral unmixing, hyperspectral, multispectral

# 2. Introduction

***2.1 Background Information***

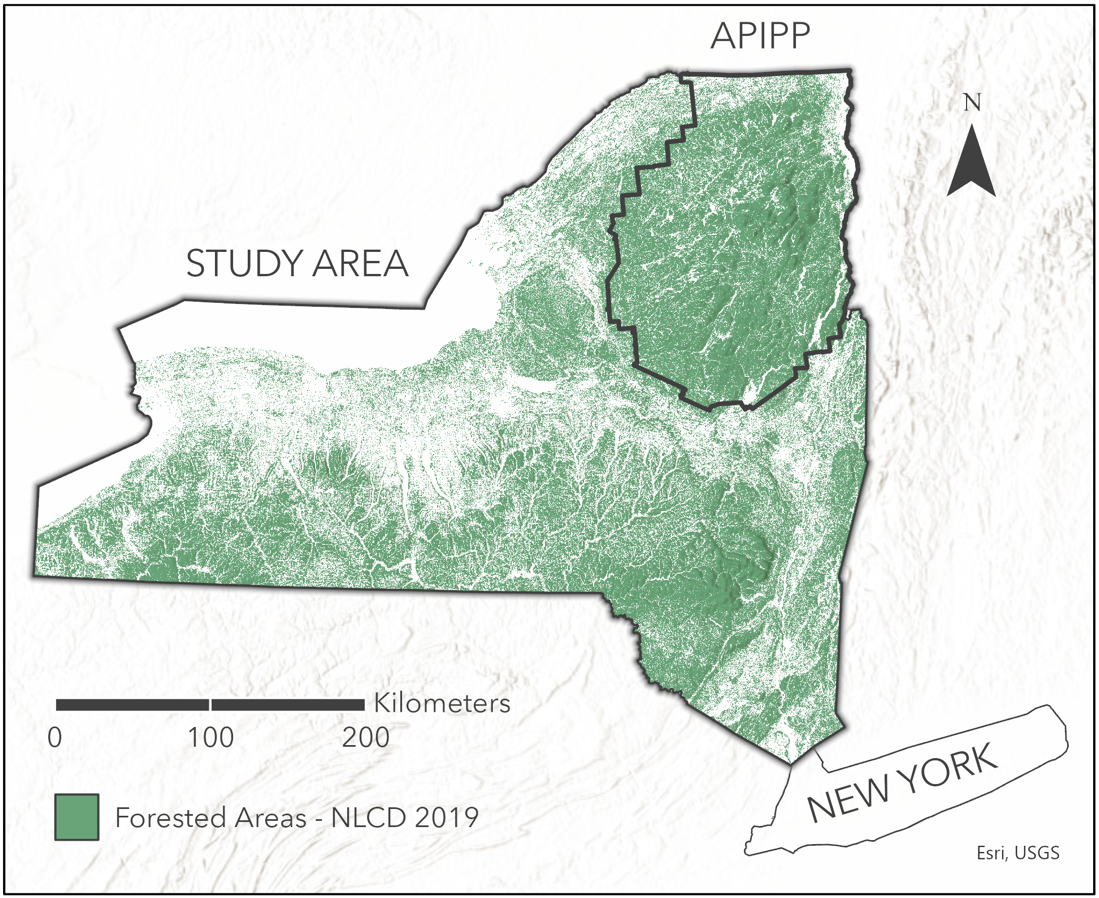
Across North America, ash trees (*Fraxinus* spp.) are threatened by the invasive emerald ash borer (EAB) (*Agrilus planipennis*). Discovered in the United States in 2002 (Haack et al., 2002), theEAB is now present in 35 states (U.S. Department of Agriculture, 2022). While in its four larval stages, EAB tunnels through ash bark, eating the xylem, cambium, and phloem (Cappaert et al., 2005). Once present in an area, trees exhibit near complete mortality within three to six years (Knight et al., 2012; Siegert et al., 2007). Kovacs et al. (2010) projected US $10.7 billion in EAB damages and treatment costs between 2009 and 2019.

The loss of ash trees has implications for community dynamics, nutrient cycling, and ecosystem services. Ash trees support species including arthropods in several orders such as Lepidoptera and Coleoptera, with 43 species exclusively relying on ash trees (Gandhi & Herms, 2010). Ash trees also provide quality leaf litter, which is important for amphibian fitness (Stephens et al., 2013). Further, ash trees support specific soil microbial assemblages, which could change as EAB-induced mortality occurs thereby altering carbon and nutrient cycling (Ricketts et al., 2018). Ash tree mortality also results in a net increase in CO2 emissions even after accounting for replacement by other trees, which has implications for warming effects at least on short timescales (Flower et al., 2013).

There are several methods used to combat EAB and protect ash trees. Quarantines were an initial nation~~-~~wide strategy; although the USDA has since lifted them due to their ineffectiveness (U.S. Department of Agriculture, 2020), New York State still enforces a quarantine (N.Y. Comp. Codes R. & Regs. tit. 6 § 192.5 n.d.). Ash trees themselves can be protected from EAB using insecticides (Mercader et al., 2015). Unfortunately, insecticide application requires human labor, which is challenging at large scales. Biological control is a more recent intervention, with four parasitoids approved for use in the U.S. since 2012, each effective within specific geographic ranges (Gould et al., 2021).

Until 2020, one area that has been predominantly safe from EAB infestation is the Adirondack Park in north-eastern New York State (New York State Department of Environmental Conservation [NYS DEC], 2020). As of 2020, EAB has spread to at least seven of twelve counties within the park (U.S. Department of Agriculture, 2022). In addition to the environmental concerns of EAB spread, this spread threatens the state’s timber industry since ash trees comprise 15% of all timber production in New York State (U.S. Department of Agriculture, 2015).

Our project aimed to inform targeted, sustained management in the Adirondack Park by mapping ash extent during the state’s pre-infestation baseline of 2009 and in 2021 (Figure 1). One known approach to inform EAB management involves mapping ash stands and/or ash density to target locations likely to experience EAB infestation. Previous studies have demonstrated the feasibility of using both multispectral and hyperspectral imagery independently, however, accuracy between ash classification methods has not been directly compared in the same study to the best of our knowledge. Multispectral classification is feasible when utilizing a multi-sensor timeseries and incorporating Light Detection and Ranging (LiDAR) data; through this method, an overall accuracy of 64% is achievable (Host et al., 2020). Hyperspectral classification seems to be more accurate, with overall accuracies ranging from 76%–86% (Chan, 2020; Liu, 2017). Once locations of ash stands are known, EAB spread can be projected (Kovacs et al., 2010; Prasad et al., 2010).



*Figure 1.* Map of study area and Adirondack Park Invasive Plant Program (APIPP) jurisdiction within New York State. NLCD 2019 forested areas are overlaid within the study region.

***2.2 Project Partners & Objectives***

We partnered with the Adirondack Park Invasive Plant Program (APIPP), which was established in 2005 as the party responsible for overseeing both aquatic and terrestrial invasive species management in the park. The APIPP is interested in partnering with NASA DEVELOP to identify the best locations for bio-control-based eradication efforts of both the EAB and other invasive species. The APIPP is familiar with remote sensing and GIS but would like to expand their usage of these technologies using NASA Earth observations.

Our first objective was to assess ash tree distribution using multispectral and hyperspectral imagery. For our second objective we sought to compare the accuracies of multispectral and hyperspectral ash identification methods with field data. Finally, we set out to identify areas susceptible to future EAB outbreaks by using successive buffer and simplified SHIFT models to compare spread scenarios with ash stand locations. We delivered these objectives through two end products: Ash Tree Distribution Maps and Emerald Ash Borer Susceptibility Maps of the Adirondack Park region. Ash distribution maps will assist the APIPP with identifying the 2009 extent of ash within the Adirondack Park. Additionally, the EAB Susceptibility Maps were forecasted to 2027, which will be used to inform EAB management procedures such as biological control and trapping. These tools will support the continuance of APIPP’s efforts to reduce EAB presence within the park and will serve as a model for the management and treatment of other invasive species in the park.

# 3. Methodology

***3.1 Data Acquisition***

Earth observation data sets sourced included multispectral, hyperspectral, and topographic data (Table A1). We accessed Landsat 7 ETM+ images from collection 2 tier 1 surface reflectance via Google Earth Engine (GEE) for two time periods: 2000–2009 and 2000–2021. SRTM data from 2000 were also accessed via GEE and were used to determine topographic indices. We obtained hyperspectral images from the AVIRIS instrument, flown on a NASA ER-2 Jet (ER-2 AVIRIS) from NASA Jet Propulsion Laboratory’s file transfer protocol site. AVIRIS flightlines intersecting Adirondack Park were collected from July 5th–28th, 2009, as these were the latest available for the region.

We acquired ancillary datasets for EAB presence, forest inventory, campground, landcover, and jurisdictional boundaries (Table B1). The EAB datasets sourced included Monitoring and Managing Ash (MaMA) EAB Surveys data form 6/13/19, iMapInvasives invasive species database, and iNaturalist. These datasets were used to identify the current distribution of EAB. All MaMA data through 2021 (latest at time of analysis) were obtained from anecdata.org. Data from iMapInvasives were shared by our APIPP partners, including all confirmed and unconfirmed EAB reports until June 23, 2022. We accessed the National Land Cover Database from the Multi-Resolution Land Characteristics Consortium via GEE for mapping forested extent. We downloaded Department of Environmental Conservation (DEC) data from the NYS GIS Clearinghouse, including DEC Campgrounds points for EAB spread assessment, and a forest inventory dataset for use as ground truth data when training classifiers.

***3.2 Data Processing***

*3.2.1 Ground Truth Data*

To prepare our data, we conducted preprocessing before proceeding with our three layers of analysis: a hyperspectral classification, a multispectral classification, and an EAB spread model. Ground truth data were shared between classification methods, but we processed them to different resolutions and subset sizes. We performed all processing in the NAD 83 UTM Zone 18N projection, using snap rasters for the study area at 30 m and 20 m resolutions as appropriate for the multispectral and hyperspectral classifications, respectively. These resolutions were chosen to ensure the data matched the resolutions of our imagery (Host et al., 2020).

Processing was performed in ArcGIS Pro v3.0 unless otherwise noted.

Our preprocessing consisted of creating snap rasters. To create the snap rasters, we first outlined a minimum bounding rectangle by area. We then rasterized this rectangle into the two resolutions for additional processing, ensuring pixel alignment between data layers.

We then created training and testing sets for each classification. First, we prepared our ground truth data of forest stands in New York State by converting stands with 100% of one tree type to raster and resampling using bilinear interpolation to match the two resolutions of our Earth observations. Next, we converted these rasters to point sets. For the hyperspectral classification, using the "select by location" tool, we identified points that intersected with the chosen flightline polygons (downloaded with the imagery). We then a created square buffers of 10 m, essentially creating ROIs (regions of interest) for individual pixels. Afterward, we selected a random training/testing split of 75% and 25% of these pixel ROIs. For the multispectral classification, a random subset of the ground truth data points was selected to create training and testing sets of roughly 5,000 points each (the maximum allowed when importing into GEE). All random selections were performed using the “RAND()” function in Microsoft Excel Online Build 16.0.15601.37952.

*3.2.2 Hyperspectral Imagery*

To process the hyperspectral imagery from AVIRIS, we used the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) atmospheric correction tool in ENVI v5.6.2 to calculate surface reflectance values and remove bad bands (i.e., redundant bands or bands with low atmospheric transmittance; L3 Harris Geospatial, 2022). FLAASH identified bad bands in the ranges of 107–116 (1353.0890–1442.8140 nm), 153-169 (1811.4750-1947.2920 nm) and 223-224 (2486.3169–2496.2361 nm). Although some flightlines had fewer bad bands than others, we used flightline f090728t01p00r05, which had the most identified, as a template for the remaining flightlines to maintain band count consistency across our analysis. Removal of these bands brought the total band count of the AVIRIS imagery down from the original 224 to 195. Reducing the band count not only decreased processing time, but also improved the accuracy of our hyperspectral classification by reducing the noise signature in the data (Tane et al., 2018).

An additional output of the FLAASH tool included a cloud mask, which we applied to our surface reflectance imagery to exclude pixels containing cloud cover. We then used ENVI’s Vegetation Index Calculator to calculate the Normalized Difference Vegetation Index (NDVI; Equation 1) for each flightline (Rouse et al., 1974). This NDVI raster provided the input for our non-vegetation mask, which we built using the ENVI Mask Builder and used to exclude pixels containing bare soil, asphalt, and other non-vegetative materials. We then set a masking threshold for NDVI values of 0.4 or lower, excluding any pixels below this value (Liu, 2017). However, visual inspection of the imagery and initial masking results led us to increase the masked values to 0.5. While a masking threshold value of 0.4 resulted in areas that were visibly unforested remaining unmasked, the 0.5 threshold provided vegetation masking that more closely adhered to the visually apparent vegetation levels in the imagery. The necessity to increase the threshold was likely due to the dense, broadleaf forests that predominate in the Adirondack Park region which can saturate the NDVI score, leading to higher than predicted NDVI values (Valtonen et al., 2021).

*3.2.3 Multispectral Imagery*

We pre-processed multispectral imagery from Landsat 7 ETM+ in GEE. Images from each sensor were already atmospherically corrected to produce surface reflectance. For all Landsat images within our study periods (2000–2009 and 2000–2021), we masked out cloud shadows and clouds by selecting only clear conditions (those with a value of zero) in the pixel quality assessment band bits 3 and 5, respectively. We then calculated the NDVI for each image (Equation 1) and added a timestamp band.

After preprocessing images, we calculated several indices in GEE as predictors of ash presence on a yearly basis, following the methodology of Host et al. (2020). Variables included NDVI amplitude (Equation 3) and NDVI phase (Equation 4) derived from an NDVI harmonic regression (Equation 2) as well as median NDVI (Host et al., 2020). In equation 2, *ω* is the period (one year in milliseconds), and *t* is the image timestamp (in milliseconds). We then isolated the regression coefficients and used the arrayFlatten() method to output the yearly coefficients to a new image for subsequent calculation.

We also calculated two layers from SRTM data. These included the Topographic Position Index (TPI) and Compound Topographic Index (CTI). These indices account for local differences in elevation and flow accumulation, respectively (Moore et al., 1991). We calculated the TPI by subtracting the mean elevation within a 3x3 neighborhood from the elevation of the cell itself (Equation 5; Wilson & Gallant, 2000). We calculated the CTI (Equation 6; Host et al. 2022) by first masking out any negative values in the SRTM data. Next, we calculated beta by finding the local slope gradient of the masked SRTM data. We calculated alpha by finding the flow direction of the masked SRTM data, and then used that data to calculate flow accumulation. We fixed alpha and beta to a constant value and then performed individual functions to calculate the CTI (Host et al., 2020). We then resampled these SRTM products using bilinear interpolation to a 30 m resolution to match our Landsat imagery.

***3.3 Data Analysis***

*3.3.1 Hyperspectral Imagery*

After determining areas where AVIRIS coverage overlapped with the 100% dominant tree species polygons from the NYS DEC field surveys, and discounting species that were represented by fewer than 100 pixels, we were left with 12 species classes to use as endmembers in building a spectral library. We used the Spectral Library Builder tool in ENVI v5.6.2 to create a spectral library containing mean spectra of each endmember class in our training ROIs. We then used ENVI’s Spectral Angle Mapping tool to compare the angle of our imagery derived spectra to the endmember spectra in our spectral library (Equation 7; Kruse et al., 1993). Using an iterative approach, we determined that the optimum maximum angle threshold equaled 0.15 radians, meaning that image spectra whose comparative line angles exceeded this threshold would not be considered as close matches and therefore remain unclassified (Kruse et al., 1993; L3 Harris Geospatial, 2022). This produced a spatially coherent, classified dataset while reducing the potential for errors of commission to occur.

Where:

* **T** is the threshold of deviance in radians from the reference spectra
* **i** is the band number
* **nb** is the total number of bands
* **t** is the test spectrum
* **r** is the reference spectrum

Following the classification of all flightlines, we used ENVI’s Confusion Matrix Using Ground Truth ROIstool to compute the overall accuracy of our classification by summing the number of correctly classified pixels and dividing them by the total number of pixels in the test ROIs (L3 Harris Geospatial, 2022). The tool also calculated the Kappa Coefficient for our dataset (Equation 8), which produces a value between 0 and 1, representing the level of agreement between a classified image and its ground truthing data. While a Kappa value of 1 indicates complete agreement between classified imagery and ground truth data, a value of 0 indicates complete disagreement (Cohen, 1960).

Where:

* **i** is the class number
* **N** is the total number of classified values compared to truth values
* **mi,i** is the number of values belonging to the truth class **i** that have also been classified as class **i** (i.e., values found along the diagonal of the confusion matrix)
* **Ci** is the total number of predicted values belonging to class **i**
* **Gi** is the total number of truth values belonging to class **i**

*3.3.2 Multispectral Imagery*

For the multispectral classifications, we trained a random forest classifier and assessed its accuracy in GEE. Each classification used an image with bands for CTI, TPI, and yearly NDVI amplitude, phase, and median as the input. The 2009 classification used 23 bands, while the 2000–2021 classification used 68 bands. Each classification used the same training and testing datasets. We trained the smileRandomForest() (Google Earth Engine, 2022) classifier (hereafter referred to as the RF classifier) with 1,000 trees instead of the ten used by Host et al. (2020) to decrease generalization error (Breiman, 2001). By default, the RF classifier defines the variables per split as equal to the standard square root of input variables, which matches the approach of Host et al. (2020). After training, we calculated a confusion matrix and overall accuracy based on the test set.

*3.3.3 EAB Spread Modeling*

For the EAB spread risk assessment, we modeled three simple scenarios from 2022 to 2027 and compared campground distribution with known EAB populations. All evaluations and scenarios assumed that both confirmed and unconfirmed EAB presences were true presences. For the campground (i.e., human-dispersal) evaluation, we selected campgrounds within the 80.5 km (50 mile) quarantine zone of currently known EAB records. The first spread scenario considers current EAB presence knowledge alone, while the second and third consider abundance estimates. We modeled the first scenario as a series of expanding buffers, each 20 km larger than the previous year, based on the insect flight distance reported by Iverson et al. (2008) and used by Prasad et al. (2010). In the second and third scenarios, we used a flight distance of nearly 10 km to account for variation in EAB flight distance reported by other studies (Taylor et al., 2010).

We modeled the second and third scenarios using a 270 m resolution cell-based approach. To do this, we spatially joined EAB presence data to the grid, then calculated the years elapsed since the year of first discovery. Abundance values were then assigned based on a right-skewed distribution for years one to ten after first discovery (Prasad et al., 2010). We then converted the grid features to a raster dataset to use focal statistics to sum abundance values within a nearly 10 km (37 cell) radius. The second scenario kept abundance values constant after the first year, while in the third scenario we calculated unique abundance values for every cell each year. In the second scenario, we also only considered new cells “infested” if their abundance sum was higher than the 50th percentile of abundance sums for all cells in that year as a sensitivity measure.

# 4. Results & Discussion

***4.1 Analysis of Results***

*4.1.1 Hyperspectral Classification*

Based on the confusion matrix created from our classified flightlines (Figure 2), we achieved an overall accuracy of 46.31% for all species, though that accuracy varies greatly between species. For Ash, our target species in this study, the producer's accuracy (i.e., detection rate) reached 21.71% with a user’s accuracy (i.e., percent of the classified pixels that are true positives) of 52.20%. European Larch had the lowest producer’s accuracy, with only 16.18% of the pixels within its test ROI being correctly classified. In comparison, Norway Spruce had the greatest producer’s accuracy with 68.2% of its test ROI pixels being correctly classified.

Diagram

Description automatically generated

*Figure 2.* Map of Ash tree distribution within APIPP after SAM classification of AVIRIS imagery. While Ash appears to occur throughout the park, they tend to group together in groves as shown in the magnified inset.

The Kappa value for our dataset was 0.3768, which falls within the range of minimal agreement ( = 0.21–0.39). This means that between 4 and 15% of the classified data is reliable (McHugh, 2012). Two types of error contributed to reducing overall accuracy and lowering the Kappa score during the classification process: errors of commission and errors of omission (Table C1). In the context of our study, commission errors occurred when pixels from outside of a target species’ ROI were classified as belonging to the target species. Errors of omission occurred when pixels from within a target species’ ROI were classified as another species (L3 Harris Geospatial, 2022.). Errors of commission are located along the rows of Table 1 while errors of omission are in the columns. Note, that this excludes cells along the main diagonal highlighted in green.

Table 1

*Hyperspectral classification confusion matrix*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SAM Classification Confusion Matrix Species Class | **Ash** | **Aspen spp** | **Cherry; Black** | **Hemlock; Eastern** | **Larch; European** | **Larch; Japanese** | **Maple; Red** | **Maple; Sugar** | **Pine; Red** | **Pine; White** | **Shrub; Tall** | **Spruce; Norway** | **Total** |
| **Unclassified** | 7.1 | 8.06 | 41.92 | 55.56 | 21.08 | 10.79 | 21.03 | 0 | 1.77 | 0 | 4.07 | 7.59 | 10.95 |
| **Ash** | 21.71 | 1.34 | 16.17 | 0 | 0.49 | 2.16 | 13.77 | 1.32 | 2.21 | 6.56 | 22.22 | 1.73 | 7.85 |
| **Aspen spp** | 9 | 64.88 | 0.6 | 0 | 3.43 | 2.81 | 3.5 | 28.95 | 7.52 | 3.28 | 3.33 | 1.46 | 8.22 |
| **Cherry; Black** | 13 | 6.53 | 26.95 | 0.74 | 0 | 0.84 | 5.13 | 7.89 | 1.33 | 9.84 | 4.44 | 3.47 | 5.73 |
| **Hemlock; Eastern** | 1.19 | 0.19 | 0 | 18.52 | 2.94 | 0 | 0.13 | 0 | 0 | 0 | 0 | 0.54 | 0.85 |
| **Larch; European** | 5.34 | 0.19 | 0 | 6.67 | 16.18 | 3.19 | 6.51 | 0 | 0 | 1.64 | 0 | 3.82 | 4.04 |
| **Larch; Japanese** | 6.04 | 4.8 | 0.6 | 0 | 18.63 | 42.59 | 2.13 | 5.26 | 14.6 | 6.56 | 2.59 | 3.05 | 9.92 |
| **Maple; Red** | 3.58 | 1.54 | 3.59 | 18.52 | 14.71 | 1.13 | 24.66 | 0 | 0 | 0 | 2.22 | 1.85 | 5.08 |
| **Maple; Sugar** | 5.9 | 0.77 | 0.6 | 0 | 0 | 0.19 | 2.13 | 36.84 | 1.33 | 0 | 1.48 | 0.19 | 1.96 |
| **Pine; Red** | 6.54 | 1.15 | 4.79 | 0 | 2.45 | 10.98 | 0.25 | 3.95 | 54.87 | 9.84 | 1.11 | 5.05 | 6.6 |
| **Pine; White** | 0.84 | 0.58 | 0.6 | 0 | 8.82 | 17.35 | 1 | 0 | 1.77 | 52.46 | 0.74 | 0.31 | 3.62 |
| **Shrub; Tall** | 12.79 | 9.79 | 3.59 | 0 | 3.92 | 4.6 | 18.02 | 10.53 | 2.21 | 9.84 | 51.48 | 2.74 | 8.87 |
| **Spruce; Norway** | 6.96 | 0.19 | 0.6 | 0 | 7.35 | 3.38 | 1.75 | 5.26 | 12.39 | 0 | 6.3 | 68.2 | 26.31 |
| **Total** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Columns represent true species classes based on the test ROIs and rows represent the predicted classification of image pixels. Cells highlighted in green along the diagonal represent the percentage of predicted classification pixels that were classified correctly.

*4.1.2 Multispectral Classification*

Overall accuracies for the test set were high for each time series duration, at roughly 97% (Table 2). Ash tree producer’s accuracies were 96–99% in the training set but were both below 10% in the test set (Table 2) which likely reflects overfitting and/or the imbalance between classes in the sets. Therefore, each of these classifications had high errors of omission for the ash class. There were no errors of commission for the ash classes (Table 2), though class numbers may be too low to draw definitive conclusions.

The accuracy of the multispectral classification was much lower than expected. It is important to note that there are several differences between the classification performed here and the one performed by Host et al. (2020) for which we based this work upon. Host et al. (2020) found that when using approximately twenty years of Landsat 7 imagery, ash producer’s accuracies of 43% were possible. Two key differences were that we omitted the canopy height model as a predictor and also did not perform canopy height or NDVI masking. Canopy height was not used due to the much lower spatial sampling rate of Global Ecosystem Dynamics Investigation (GEDI) data compared to the LiDAR data used by Host et al. (2020), while NDVI masking even at low thresholds omitted desired forested areas. Another notable difference is that we used stands that were 100% one tree species for training and testing, instead of using stands that were either ash-dominated or not ash-dominated. Host et al. (2020) also only selected stands verified within several years of the end of their imagery timeseries. Most of the stands in our ground truth dataset were updated between 2011–2021, indicating they were likely accurate for the 2009 classification but may not have been for the 2021 classification.

Table 2

*Multispectral classification confusion matrix*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **2000 – 2009** | | **2000 – 2021** | |
| **Ash** | **Other** | **Ash** | **Other** |
| **Ash** | 2 | 130 | 10 | 116 |
| **Other** | 0 | 4396 | 0 | 4124 |
| **Overall Accuracy** | 97.1% | | 97.3% | |
| **Ash Producer’s Accuracy** | 1.5% | | 7.9% | |
| **Ash User’s Accuracy** | 100% | | 100% | |

Matrix for two periods of imagery used to classify ash stands. Columns indicate the classification assigned, while rows represent the actual class.

*4.1.3 EAB Spread Model*

Of the three scenarios, scenario one shows the greatest geographic spread potential, while scenario two shows the least. Scenario one suggests that when assuming a maximum range expansion of 20 km per year, EAB could fully establish in APIPP by 2026 based on current presence knowledge (Figure 3). In comparison, scenario two shows that the 2027 extent of EAB presence would roughly be between scenario one’s projection for 2023 and 2024, while scenario three roughly matches its projection for 2024 (Figure 4).

|  |  |
| --- | --- |
| Inserting image... |  |

*Figure 3.* Successive buffer spread model of EAB spread through 2027, based on observations at time of analysis.

|  |  |
| --- | --- |
| Inserting image...(a) | (b) |
|  |  |

*Figure 4.* Cellular EAB spread risk scenarios for 2027 modeled using abundance values estimated from time since discovery. Scenario two (a) uses a constant abundance value after the initial year and a 50th percentile threshold of cellular abundance sums for introduction to occur. Scenario three (b) uses a variable abundance estimated from time since discovery with no threshold.

It is important to note that these scenarios only account for insect flight and omit many factors that modify EAB establishment likelihood. An omitted key factor used in other EAB spread models is wall-to-wall ash density (Prasad et al., 2010). Even though the hyperspectral classification was more sensitive than the multispectral classification, its low accuracy and limited spatial extent resulted in a product that could not accurately capture ash distribution for our spread modeling purposes. Even when simply looking at insect dispersal, there is also a decay in likelihood of spread with distance to consider (Prasad et al., 2010), which is also unaccounted for in these models.

From currently known populations, all public campgrounds are within the 80.5 km (50 mile) quarantine zone, indicating that introduction could occur within a year at any campsite if individuals transported infested firewood. However, models of human-mediated transport account for factors such as human population, campsite characteristics, and “gravity” behavior (Muirhead et al., 2006; Prasad et al., 2010); consideration of these factors was outside of the scope of this project.

***4.2 Future Work***

There are several limitations of note in our analysis and areas where future work could be done. The ground truth data collected by the DEC had very few stands in the APIPP jurisdiction, and these data were collected over a ten-year period. Within the timeframe of our analysis, EAB was spreading across the state, killing ash trees. Therefore, the temporal mismatch between ground truth data and the conditions that might be present in the imagery complicates making a more recent classification baseline. Running these analyses again with data in the Adirondack Park and utilizing only more recent data may improve results.

In the future, a sufficient amount of time should be focused on developing the EAB spread model. Due to time constraints, we created only simplified models of EAB spread risk. Realistically, EAB spread is much more complicated and includes many additional factors, such as human population, randomness, and varying dispersal likelihood (Muirhead et al., 2006; Prasad et al., 2010). Given the low sensitivity of the multispectral method and only having partial coverage of hyperspectral data for our study region, we could not model EAB spread using wall-to-wall ash distribution, which would be optimal. If a more accurate and comprehensive ash classification were achieved, a proper EAB spread model for this project could be performed. This improved model could provide three output maps of risk: risk of the EAB to fly into new zoned areas of the park, risk of the EAB to ride with humans, and overall risk as a combination of the flight and ride models. With a more comprehensive analysis of the EAB spread risk throughout the Adirondack Park, we could have better addressed our objective of forecasting emerald ash borer susceptibility to assist project partners with invasive species management practices.

If this project were to be continued for another term, it would be beneficial to use a similar method of analysis that we did with some adjustments to produce more comprehensive results. For instance, Sentinel-2 data could be incorporated into the multispectral analysis to improve temporal coverage, which would solve the issue of a temporal mismatch between ground truth data and imagery and therefore improve accuracy in our results. Incorporating Landsat 8 and 9 data would also be beneficial to find better temporal coverage and fewer imagery gaps in our analysis. With better temporal coverage, NDVI masking should produce fewer errors as well, thereby aligning classification methods more closely with Host et al. (2020). A future team could also incorporate lower coverage percentages of ash trees in a stand and balance training sets, which may also increase sensitivity of the multispectral model.

To improve the hyperspectral approach, DLR Earth Sensing Imaging Spectrometer (DESIS) data could be incorporated to create a more current classification of ash stands with greater spatial coverage throughout the Adirondack Park. A greater extent of imagery would also overlap with more ground truth data, which would increase the unique samples available thereby potentially improving accuracy. Ground truth data could also be reviewed to find likely present ash stands for a more recent classification than 2009. Further, it may be beneficial to pursue other classification methods other than SAM.

It would also be beneficial to include ash tree density in addition to distribution within the Adirondack Park in future analysis. This is another area of study that we were unable to deliver on due to time and data constraints, but it was desired by our partners to determine which areas in the park have at least a 25% density threshold within at least 40 acres that is also connected to other ash stands. This could potentially be completed using data from the DEC’s forest inventory polygons on our ash distribution analysis. Future work could also run a regression of calculated density versus measured basal area on the multispectral and hyperspectral classifications to help determine ash density in the park; this would most likely work best with the hyperspectral data, but the regression could be run on whichever product seems more suitable. This regression would assess how well our classifications matched the basal area, providing another demonstration of accuracy for our results.

# 5. Conclusions

Our team produced three different maps and compared the accuracy of the different methods, which demonstrated that hyperspectral imagery can be useful for preliminary ash tree classification. The team also demonstrated that multispectral imagery may not be sufficient for ash tree classification when using relatively short periods of imagery and single sensors. Further, these results may have been limited by the lack of high-resolution LiDAR data. The results indicate that for detecting ash presence, the team’s hyperspectral classification using Spectral Angle Mapping (SAM) is slightly more sensitive, however it was less accurate than multispectral Random Forest (RF) classification and neither method was above a ~20% detection rate. The lack of forest ground truth data within APIPP also complicated our ability to assess accuracy within the region itself. While the methods employed had their limits, the final products will still be useful to APIPP for identifying the distribution of ash tree extent in 2009 within Adirondack Park. These products have the potential to inform EAB management practices such as biological control and trapping. Overall, these tools could help APIPP as they strive to continue their efforts in reducing EAB presence within the park as well as improve management and raise public awareness of EAB’s threat to communities within the region.

# 6. Acknowledgments

The team would like to thank our partner representative from the Adirondack Park Invasive Plant Program, Rebecca Bernacki for her collaboration and input on this project. We also would like to give a special thank you to our science advisors Dr. Juan Torres-Pérez, Dr. Kenton Ross, Dr. Mark Friedl and Joseph Spruce, and also our Node Fellows Britnay Beaudry and Hayley Pippin for their continued support throughout this term. Finally, we thank the DEVELOP Project Coordination Fellows Tamara Barbakova and Robert Cecil Byles for their guidance and editing.

Maps throughout this work were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. All rights reserved.

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 NNL16AA05C.

# 7. Glossary

**APIPP** – Adirondack Park Invasive Plant Program

**AVIRIS** – Airborne Visible and Infrared Imaging Spectrometer; hyperspectral sensor aboard a NASA ER-2 Jet

**CTI** – Compound Topographic Index; a measure of water flow accumulation

**DEC** – Department of Environmental Conservation

**EAB** – Emerald ash borer

**Earth observations** – Satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**ETM+** – Enhanced Thematic Mapper Plus; multispectral sensor on NASA’s Landsat 7 satellite

**FLAASH** – Fast Line-of-sight Atmospheric Analysis of Hypercubes; ENVI tool used for atmospheric correctance and cloud masking

**GEE** – Google Earth Engine

**LiDAR** – Light Detection and Ranging

**NDVI** – Normalized Difference Vegetation Index

**NYS** – New York State

**Parasitoids** – Insect with a larval stage that develops in or on another insect host

**PRISM** – Partnership for Regional Invasive Species Management; title for management organizations within New York State responsible for coordinating invasive species management

**SAM** – Spectral Angle Mapping; image classification method that uses angular differences between spectra to differentiate between endmember classes

**TPI** – Topographic Position Index; a measure of relative terrain elevation change

# 8. References

Breiman, L. (2001). *Random Forests*. Kluwer Academic Publishers. <https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf>

Cappaert, D., McCullough, D. G., Poland, T. M., & Siegert, N W. (2005). Emerald ash borer in North America: a research and regulatory challenge. *American Entomologist,* *51*(3), 152-165. <https://www.nrs.fs.fed.us/pubs/jrnl/2005/nc_2005_cappaert_001.pdf>

Chan, C. (2020). *Airborne hyperspectral data application in stress detection of blueberry fields and ash trees.* [Master’s thesis, University of Maine]. <https://digitalcommons.library.umaine.edu/etd/3307>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46 <https://doi.org/10.1177/001316446002000104>

‌

Dewitz, J. (2021). *National land cover database (NLCD) 2019 products* (version 2.0) [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54

Flower, C. E., Knight, K. S., & Gonzalez-Meler, M. A. (2013). Impacts of the emerald ash borer (agrilus planipennis fairmaire) induced ash (fraxinus spp.) mortality on forest carbon cycling and successional dynamics in the Eastern United States. *Biological Invasions*, *15*(4), 931–944. <https://doi.org/10.1007/s10530-012-0341-7>

Gandhi, K. J., & Herms, D. A. (2010). North American arthropods at risk due to widespread fraxinus mortality caused by the alien emerald Ash Borer. *Biological Invasions*, *12*(6), 1839–1846. https://doi.org/10.1007/s10530-009-9594-1

Google Earth Engine. (2022). ee.Classifier.smileRandomForest [Source code]. Retrieved July 26, 2022 from https://developers.google.com/earth-engine/apidocs/ee-classifier-smilerandomforest.

Gould, J., Slager, B., Bauer, L., Duan, J., & Petrice, T. (2021). *Emerald ash borer,* Agrilus planipennis *(fairmaire), biological control release and recovery guidelines 2021.* U.S. Department of Agriculture. <https://www.aphis.usda.gov/plant_health/plant_pest_info/emerald_ash_b/downloads/eab-field-release-guidelines.pdf>

Haack, R., Jendek, E., Liu, H., Marchant, K. R., Petrice, T. R., Poland, T. M., & Ye, H. (2002, September). The emerald ash borer: A new exotic pest in North America. *Newsletter of the Michigan Entomological,* 1–5. <https://www.researchgate.net/publication/228544207_The_Emerald_Ash_Borer_A_New_Exotic_Pest_in_North_America>

Host, T. K., Russell, M. B., Windmuller-Campione, M. A., Slesak, R. A., & Knight, J. F. (2020). Ash presence and abundance derived from composite landsat and sentinel-2 time series and LIDAR Surface Models in Minnesota, USA. Remote Sensing, 12(8), 1341. https://doi.org/10.3390/rs12081341

iNaturalist. (2022). *Emerald ash borer presence records, research-grade.* [Data set]. iNaturalist Network. https://www.inaturalist.org /

Iverson, L. R., Prasad, A., Bossenbroek, J., Sydnor, D., & Schwartz, M. W. (2008). *Advances in Threat Assessment and Their Application to Forest and Rangeland Management*. <https://www.nrs.fs.fed.us/pubs/jrnl/2010/nrs_2010_iverson_003.pdf>

Knight, K. S., Brown, J. P., & Long, R. P. (2012). Factors affecting the survival of ash (*Fraxinus spp.*) trees infested by emerald ash borer (*Agrilus planipennis). Biological Invasions*, *15*(2), 371–383. <https://doi.org/10.1007/s10530-012-0292-z>

Kovacs, K. F., Haight, R. G., McCullough, D. G., Mercader, R. J., Siegert, N. W., & Liebhold, A. M. (2010). Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. *Ecological Economics*, *69*(3), 569–578. <https://doi.org/10.1016/j.ecolecon.2009.09.004>

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., & Goetz, A. F. H. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, *44*(2), 145–163. <https://doi.org/10.1016/0034-4257(93)90013-N> ‌

L3 Harris Geospatial. (2022). Spectral Angle Mapper. www.l3harrisgeospatial.com. Retrieved July 25, 2022, from <https://www.l3harrisgeospatial.com/docs/spectralanglemapper.html> ‌

Liu, H. (2017). *Ash tree identification based on the integration of hyperspectral imagery and high-density lidar data* [Doctoral dissertation, University of Wisconsin-Milwaukee]. <https://dc.uwm.edu/cgi/viewcontent.cgi?article=2511&context=etd>

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>

Mercader, R. J., McCullough, D. G., Storer, A. J., Bedford, J. M., Heyd, R., Poland, T. M., & Katovich, S. (2015). Evaluation of the potential use of a systemic insecticide and girdled trees in area wide management of the emerald ash borer*. Forest Ecology and Management, 350*, 70–80. <https://doi.org/10.1016/j.foreco.2015.04.020>

Monitoring and Managing Ash (MaMA) (2022). *MaMA Ash\_EAB Surveys – data form 6\_13\_19*

[Data set]*.* Ecological Research Institute. <http://www.monitoringash.org/ash-eab-surveys>

Moore, I.D., Grayson, R.B. and Ladson, A.R. (1991), Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, *5*, 3-30. <https://doi.org/10.1002/hyp.3360050103>

Muirhead, J. R., Leung, B., van Overdijk, C., Kelly, D. W., Nandakumar, K., Marchant, K. R., & MacIsaac, H. J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions*, *12*(1), 71–79. <https://doi.org/10.1111/j.1366-9516.2006.00218.x>

NASA Jet Propulsion Laboratory (2016). *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)* [Data set]. National Aeronautics and Space Administration. <https://doi.org/10.5067/R92TRY26ZN2Q>

NASA Shuttle Radar Topography Mission. (2013). *Shuttle Radar Topography Mission (SRTM) Global* [Data set]. OpenTopography. <https://doi.org/10.5069/G9445JDF>

National Land Cover Database. (2019). *Multi-Resolution Land Characteristics Consortium* [Data set]. Multi-Resolution Land Characteristics Consortium. <https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover>

*N.Y. Comp. Codes R. & Regs. Tit. 6 § 192.5.* (n.d).

<https://govt.westlaw.com/nycrr/Document/I21efc065c22211ddb7c8fb397c5bd26b?viewType=FullText&originationContext=documenttoc&transitionType=CategoryPageItem&contextData=(sc.Default)&bhcp=1>

New York State Department of Environmental Conservation. (2020, August). *DEC Confirms Emerald Ash Borer in Warren County*. https://www.dec.ny.gov/press/121040.html

New York State Department of Environmental Conservation. (2021a). *DEC lands and campgrounds* [Data set]. NYS GIS Clearinghouse. <https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1114>

New York State Department of Environmental Conservation. (2021b). *DEC state land forest stands* [Data set]. NYS GIS Clearinghouse. <https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1356>

New York State Department of Environmental Conservation. (2015, October). *New York forests 2012.* https://www.dec.ny.gov/docs/lands\_forests\_pdf/nyforests2012.pdf

New York iMapInvasives. (2022, June 23). *Emerald ash borer confirmed and unconfirmed records* [Data set]. New York Natural Heritage Program. [https://www.nyimapinvasives.org](https://www.nyimapinvasives.org/)

New York Natural Heritage Program. (2022). *PRISM boundaries* [Data set]. Retrieved June 20, 2022 from <https://services6.arcgis.com/DZHaqZm9cxOD4CWM/ArcGIS/rest/services/SG_PRISM_Boundary/FeatureServer>

Prasad, A. M., Iverson, L. R., Peters, M. P., Bossenbroek, J. M., Matthews, S. N., Sydnor, T. D., & Schwartz, M. W. (2010). Modeling the invasive emerald ash borer risk of spread using a spatially explicit cellular model. *Landscape Ecology*, *25*(3), 353–369. <https://doi.org/10.1007/s10980-009-9434-9>

Ricketts, M., Flower, C., Knight, K., & Gonzalez-Meler, M. (2018). Evidence of ash tree (*Fraxinus spp.*) specific associations with soil bacterial community structure and functional capacity. *Forests*, *9*(4), 187. <https://doi.org/10.3390/f9040187>

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Third Earth Resources Technology Satellite-1 Symposium*, *1*. <https://ntrs.nasa.gov/citations/19740022614>

Siegert, N. W., McCullough, D. G., Liebhold, A. M., & Telewski, F. W. (2007). *Resurrected from the ashes: a historical reconstruction of emerald ash borer dynamics through dendrochronological analysis.* United States Department of Agriculture, Forest Service, Northern Research Station. <https://www.nrs.fs.fed.us/pubs/jrnl/2007/nrs_2007_siegert_001.pdf>

Stephens, J. P., Berven, K. A., & Tiegs, S. D. (2013). Anthropogenic changes to leaf litter input affect the fitness of a larval amphibian. *Freshwater Biology*, *58*(8), 1631–1646. https://doi.org/10.1111/fwb.12155

Tane, Z., Roberts, D., Veraverbeke, S., Casas, Á., Ramirez, C., & Ustin, S. (2018). Evaluating Endmember and band selection techniques for multiple endmember spectral mixture analysis using post-fire imaging spectroscopy. *Remote Sensing*, *10*(3), 389. https://doi.org/10.3390/rs10030389

Taylor, R. A. J., Bauer, L. S., Poland, T. M., & Windell, K. N. (2010). Flight performance of *agrilus planipennis* (Coleoptera: Buprestidae) on a flight mill and in free flight. *Journal of Insect Behavior*, *23*(2), 128–148. <https://doi.org/10.1007/s10905-010-9202-3>

U.S. Census Bureau. (2021). TIGER/line state boundary shapefiles [Data set]. <https://www2.census.gov/geo/tiger/TIGER2021/STATE/>

U.S. Department of Agriculture. (2022). *Emerald ash borer (EAB) known infested counties* [Map]. Retrieved June 16, 2022 from <https://www.aphis.usda.gov/aphis/maps/plant-health/eab-map>

U.S. Department of Agriculture. (2020, December 15). *Removal of emerald ash borer domestic quarantine regulations.* Regulations. <https://www.regulations.gov/document/APHIS-2017-0056-015>

U.S. Department of Agriculture. (2015, October). *New York forests 2012.* United States Department of Agriculture, Forest Service, Northern Research Station. <https://www.dec.ny.gov/docs/lands_forests_pdf/nyforests2012.pdf>

U.S. Geological Survey Earth Resources Observation and Science Center. (1999-2022). *Landsat 7 ETM+ Collection 2 Tier 1 Data Products* [Data set]. U.S. Geological Survey. <https://doi.org/10.5066/P9TU80IG>

Valtonen, A., Korkiatupa, E., Holm, S., Malinga, G. M., & Nakadai, R. (2021). Remotely sensed vegetation greening along a restoration gradient of a tropical forest, Kibale National Park, Uganda. *Land Degradation & Development*, *32*(18). <https://doi.org/10.1002/ldr.4096>

Wilson, J. P., & Gallant, J. C. (2000). Primary topographic attributes. In J. P. Wilson & J. C. Gallant (Ed.), *Terrain Analysis: Principles and applications* (pp. 51–85). John Wiley & Sons. <https://www.researchgate.net/publication/303543730_Primary_topographic_attributes>

‌

# 9. Appendices

Table A1

*Remote sensing data*

|  |  |  |
| --- | --- | --- |
| **Platform & Sensor** | **Collection Dates** | **Collection IDs** |
| ER-2 AVIRIS | 2009-07-05 | f090705t01p00r10 (Ground Truthing) |
| ER-2 AVIRIS | 2009-07-10 | f090710t01p00r05 |
| ER-2 AVIRIS | 2009-07-15 | f090715t01p00r08, f090715t01p00r12 |
| ER-2 AVIRIS | 2009-07-28 | f090728t01p00r05 (Ground Truthing),  f090728t01p00r06, f090728t01p00r08 |
| Landsat-7 ETM+ | 2000-01-01 – 2009-12-31  2000-01-01 – 2021-12-31 | Path/Row: 13/30, 13/31, 13/32, 14/29, 14/30, 14/31, 14/32, 15/29, 15/30, 15/31, 16/29, 16/30, 16/31, 17/29, 17/30, 17/31, 18/30, 18/31 |
| SRTM | 2000 | N/A |

Table B1

*Ancillary data*

|  |  |  |
| --- | --- | --- |
| **Product Title** | **Product Date** | **Use** |
| iMapInvasives: Invasive Species Database | 2022 | Identify EAB presence for spread model |
| Monitoring and Managing Ash (MaMA) Ash/EAB Surveys: Ash & EAB presence/absence | 2021 | Identify EAB for spread model |
| iNaturalist: Biodiversity Database | 2022 | Identify EAB for spread model |
| DEC State Land Forest Stands 2021: Forest Database | 2021 | Train and validate hyperspectral and multispectral classifiers |
| Multi-Resolution Land Characteristics (MRLC) National Land Cover Database (NLCD) 2019: Land Cover Collection | 2019 | Map forest extent |
| DEC Campgrounds 2021: Public campground points and polygons | 2021 | Evaluate human-mediated EAB spread potential |
| 2021 TIGER/Line® Shapefiles: States (and equivalent) | 2021 | Define study area within in New York State |
| PRISM Boundaries | 2022 | Display APIPP boundary within maps and reference when selecting AVIRIS flightlines |

Table C1

*SAM Classification Errors of Commission/Omission*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | **Commission (%)** | **Omission (%)** | **Commission (Pixels)** | **Omission (Pixels)** |
| **Ash; White** | 47.8 | 78.29 | 283/592 | 1114/1423 |
| **Aspen spp** | 45.48 | 35.12 | 282/620 | 183/521 |
| **Cherry; Black** | 89.58 | 73.05 | 387/432 | 122/167 |
| **Hemlock; Eastern** | 60.94 | 81.48 | 39/64 | 110/135 |
| **Larch; European** | 89.18 | 83.82 | 272/305 | 171/204 |
| **Larch; Japanese** | 39.3 | 57.41 | 294/748 | 612/1066 |
| **Maple; Red** | 48.56 | 75.34 | 186/383 | 602/799 |
| **Maple; Sugar** | 81.08 | 63.16 | 120/148 | 48/76 |
| **Pine; Red** | 75.1 | 45.13 | 374/498 | 102/226 |
| **Pine; White** | 88.28 | 47.54 | 241/273 | 29/61 |
| **Shrub; Tall** | 79.22 | 48.52 | 530/669 | 131/270 |
| **Spruce; Norway** | 10.84 | 31.8 | 215/1984 | 825/2594 |