MODELING ASPEN IN SOUTHEAST IDAHO USING LANDSAT 8 AND SENTINEL-2 IMAGERY: A COMPARISON

Brandy Nisbet-Wilcox, Graduate Research Assistant, Idaho State University GIS Training and Research Center, brandynisbetwilco@isu.edu

Keith Weber, GIS Director, Idaho State University GIS Training and Research Center, webekeit@isu.edu

1. ABSTRACT

This study compares two probabilistic decision forest land cover classification models focused on identifying aspen (*Populus tremuloides*) for the year 2023 across southeast Idaho. The Landsat 8-derived decision forest model (LDDFM) was based on 30-meter spatial resolution Landsat 8 imagery and the Sentinel-2-derived decision forest model (SDDFM) was based on 10-meter spatial resolution Sentinel-2 imagery. Comparisons were made between overall model accuracy, predicted aspen acreage, and a HUC08 watershed-level comparison of mean probability values at or above a 0.25 threshold. Based upon a standard error matrix, the LDDFM had higher accuracy (90% producer's accuracy) compared SDDFM (77% producer's accuracy). Kappa Index of Agreement (KIA) indicated very similar results, with LDDFM having a KIA of 0.92 and SDDFM having a KIA of 0.86. The percent difference between predicted acreage for the LDDFM and SDDFM models was 51% with the SDDFM also predicting 316,000 more acres of aspen within the study area compared to the LDDFM. Watershed basin analysis showed SDDFM predicted higher acreages of aspen in 20 of the 21 watersheds within the study area compared to the LDDFM. The findings of this study suggest the 30-m spatial resolution LDDFM performed well and is appropriate for studying landscape-level aspen populations, though Landsat may be unable to classify smaller stands. These results also suggest 10-m Sentinel-2 imagery does not provide fine enough spatial resolution to predict smaller stands either without gross overprediction. Future investigations into even finer spatial resolution commercial imagery to inventory aspen stands in areas of particular management and conservation concern are recommended.

2. INTRODUCTION

Aspen (*Populus tremuloides*) are the most widespread broadleaf tree in North America and are frequently the only broadleaf species in an otherwise conifer-dominated boreal landscape (Kitchen et al. 2019). Often referred to as a keystone species (Wilson 1992), aspen serve a disproportionately important role in the biodiversity and functioning of the ecosystems in which they appear (Kay 1997). They also provide a number of critical ecosystem services including nutrient cycling, carbon sequestration, and both food and shelter for many species of plants, insects, microbes, and animals (Kouki and Martikaenen, 2004). Aspen exists across diverse ecological settings and as a result, exhibit a diverse set of ecological roles, making generalizations challenging and context specific studies of aspen necessary for well-informed management (Romme et al. 2001).

While aspen has been reported to be declining across the western United States (Singer et al. 2019), few studies focus specifically on aspen populations in eastern Idaho. This study serves as one component of a larger study aimed at using remotely-sensed imagery to classify and map changes in aspen population and distribution across southeast Idaho (**Figure 1**). The focus of this study was to

compare two probabilistic decision forest models, one based on 30-m spatial resolution Landsat 8 imagery and the other based on 10-m spatial resolution Sentinel-2 imagery. The points of comparison were overall aspen model accuracy, predicted aspen acreage, and mean probability at each watershed basin.

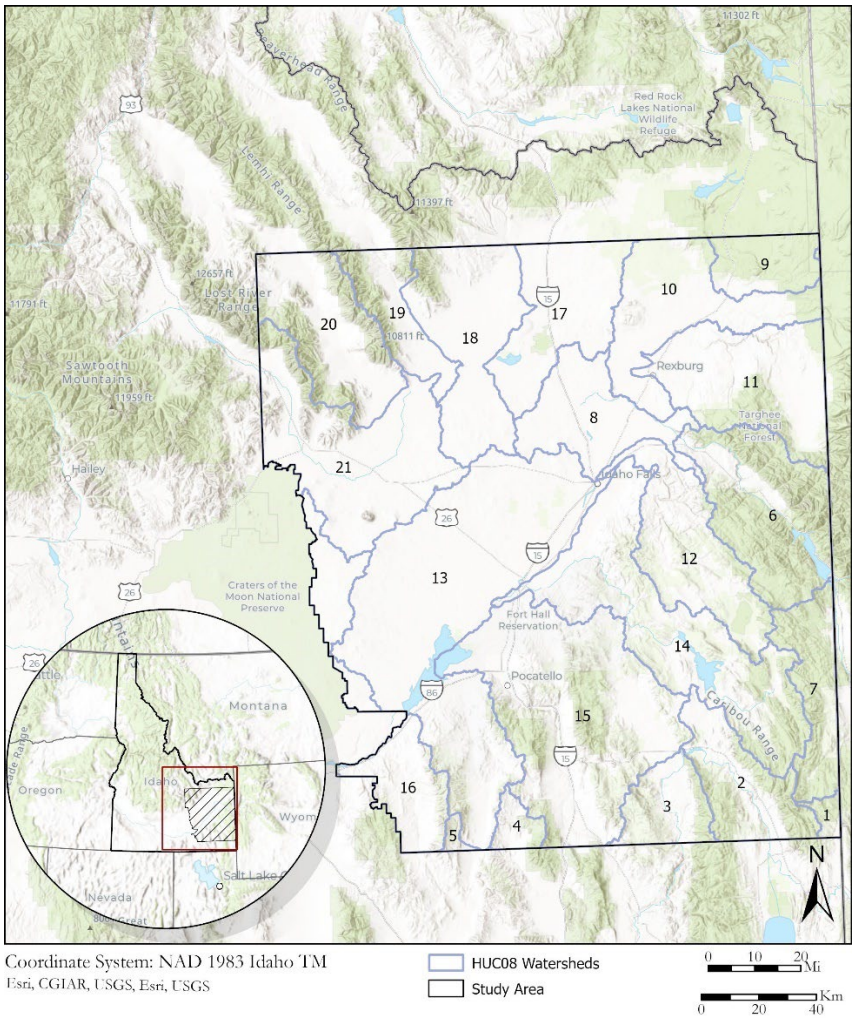

Figure 1. *Study area in southeast Idaho with 21 watershed boundaries (HUC08) used for analysis.*

This study sought to determine whether 10-m Sentinel-2 imagery would improve the accuracy and predictive power of a probabilistic decision forest model used to predict aspen in a section of southeast Idaho when compared to the same model based on 30-m Landsat imagery. Land managers at the Bureau of Land Management indicated that previous iterations of the Landsat-derived model were simultaneously overpredicting aspen in some watersheds within the study area while missing known stands in others. Because Landsat's 30-m spatial resolution is fairly coarse and some aspen stands may be too small, too sparse, or not spectrally homogenous enough to be detected at this resolution, we hypothesized that finer spatial resolution 10-m Sentinel-2 imagery might reduce overprediction while being able to detect smaller stands than the Landsat-derived model would allow.

3. METHODS

To test the hypothesis that finer spatial resolution 10-m Sentinel-2 imagery might perform better, two probabilistic decision forest models were built. The first used 30-m spatial resolution Landsat 8 data from the month of August, 2023 (LDDFM) and the second used 10-m spatial resolution Sentinel-2 data from the month of July, 2023 (SDDFM). Though there are more spectral bands available from Landsat 8 than for Sentinel-2, the blue, green, red, and near-infrared bands, plus the first component of a principal component analysis from each dataset were used to enable a more direct comparison. Both models were given the same set of landcover training points and ancillary datasets including topographic layers and a distance to streams layer.

Three primary aspects of model performance were investigated. First, validation against the same set of reserved known landcover points was performed on each model using a standard error matrix to compare overall model accuracy by looking at the Kappa Index of Agreement (KIA) for each model as well as agreement for the aspen class specifically. Second, a statistically-derived cutoff threshold (SDCT) was determined for each model. This was accomplished by extracting Decision Forest output probability values for each aspen validation point using the Extract Multivalues to Points geoprocessing tool in ArcGIS Pro. The mean probability value was then determined for each model (i.e., LDDFM and SDDFM) and is henceforth referred to as the SDCT. Pixels with values at or above the SDCT were assumed to be dense, homogeneous aspen stands. The area of aspen (in acres) was then calculated for each model across the study area. Third, zonal statistics were run in ArcGIS Pro using HUC08 watershed basins to test how model performance varied spatially and between the two models. To do this, a second probability threshold of 0.25 was used to remove pixels that likely were not aspen for the purpose of this analysis. Once this was completed, the mean probability value of aspen pixels within each watershed basin was determined, with higher mean values indicating greater model confidence for the presence of aspen.

3.1 Data Acquisition

Landsat 8 Optical Land Imager (OLI) 30-meter spatial resolution imagery was used for this study. Each scene contained less than ten-percent cloud cover and was acquired via the United States Geological Survey (USGS) Earth Explorer data portal. Images were selected from the Landsat Collection 2, Level-1 dataset. Images with the least amount of visible cloud cover, snow/ice, smoke contamination were selected. Four overlapping scenes were necessary to cover the entire study area.

In addition, Sentinel-2 Multispectral Imager (MSI) 10-meter spatial resolution imagery was also used for this study. Similarly, each scene contained less than ten percent cloud cover and was acquired via the European Space Agency (ESA) Copernicus Browser data portal using the Sentinel-2 MSI L2A collection. Images from July, August, and September were investigated, but images collected on 4 July 2023 were selected as these had the least amount of cloud, snow/ice, and smoke contamination overall. Four overlapping images were required to cover the entire study area.

For a more direct sensor-to-sensor comparison, only the blue, green, red, and near-infrared bands were used from both Landsat 8 OLI and Sentinel-2 MSI. The wavelengths and bandwidths for each instrument are quite similar in the blue, green, and red, and vary only slightly in the near-infrared (NIR) band (**Table 1**).

Table 1. *Comparison of bands used from Landsat 8 OLI and Sentinel-2 MSI.*

	Landsat 8 (L8)			Sentinel-2 (S2)		
	Band ID	**Start wavelength (µm)**	**End wavelength (µm)**	**Band ID**	**Start wavelength (µm)**	**End wavelength (µm)**
Blue	**2**	0.45	0.51	**2**	0.46	0.53
Green	**3**	0.53	0.59	**3**	0.54	0.58
Red	**4**	0.64	0.67	**4**	0.65	0.68
NIR	**5**	0.85	0.88	**8**	0.78	0.89

Topographic layers were acquired from the NASA RECOVER database. These layers include aspect, elevation, landforms, maximum curvature, slope, and topographic shape. A distance to streams layer was also created using the Distance module in Idrisi TerrSet and the USGS NHD Rivers, streams, and flowlines dataset to calculate and assign a Euclidean distance to each pixel from the nearest perennial and intermittent stream. For use with Landsat imagery, these layers were resampled from 10-meter spatial resolution to 30-meter spatial resolution using the resample geoprocessing tool in ArcGIS Pro with bilinear interpolation, then clipped to the extent of the study area to feed into the decision forest model.

3.2 Data Preparation

3.2.1 Atmospheric Correction

Landsat 8 OLI images were atmospherically corrected in Idrisi TerrSet's Landsat archive import module where multispectral bands were converted to reflectance using the Cosine(t) model of reflectance correction (Chavez, P.S. 1996). Sentinel-2 images were already atmospherically corrected at the point of download, and were simply imported into TerrSet.

3.2.2 Normalized Difference Vegetation Index

The red and near infrared bands of the eight total Landsat 8 and Sentinel-2 images were used to derive a normalized difference vegetation index (NDVI) image for each tile using the following equation (Pettorelli et al. 2005):

Equation 1. $NDVI = (NIR - Red)/(NIR + Red)$

For Landsat 8 images, NDVI was equal to (B5 - B4)/(B5 + B4). For Sentinel-2 images, NDVI was equal to (B08 - B04)/(B08 + B04). The resulting images all had values between -1 and 1, where low values indicate pixels with lower vegetative vigor and higher values indicate pixels with higher vegetative vigor.

3.2.3 Principal Component Analysis

For each Landsat 8 scene and each Sentinel-2 tile, the blue, green, red, and NIR bands were added to a raster group file in TerrSet. Using the principal component analysis (PCA) module in TerrSet, the first principal component for each scene/tile was extracted. For the Landsat-derived DF model, 83% of overall dataset variability was explained by the first component. For the Sentinel-derived DF model, 91% of overall dataset variability was explained by the first component. PCA was run using the Forward T-Mode analysis type and an unstandardized covariance matrix. PCA was used to pull out the largest

amount of spectral variability within the original datasets while removing redundancy between band files (Richards, 2013).

3.2.4 Mosaic Datasets to Raster Group Files

All image files were reprojected from UTM 11N to NAD 1983 Idaho Transverse Mercator (TM) and mosaiced in ArcGIS Pro using bilinear resampling to create continuous layers for each band, NDVI, and PCA layers for both Landsat 8 and Sentinel-2 imagery. These mosaiced layers were then clipped to the minimum bounding box around the study area for further analysis. The topography and distance to streams layers were continuous over the state of Idaho, so they were also clipped to the same dimensions as the satellite imagery-derived layers. These layers were then imported again into Idrisi TerrSet and compiled into two raster group files (RGF) where one RGF contained the mosaiced and clipped Landsat 8 band layers, NDVI, PCA, and 30-m topographic variables and the other RGF contained the mosaiced and clipped Sentinel-2 band layers, NDVI, PCA, and 10-m topographic variables and distance to streams layer.

3.2.5 Field Sites

To train and validate the models, a set of field sites was created using a combination of field observations, interpretation of aerial imagery, and the LANDFIRE Existing Vegetation Type (EVT) model (USDOI, USGS, USDA, 2023). While this study focused on aspen in particular, the field sites for model training and validation were grouped into ten land cover classes shown in **Table 2** below.

Table 2. *Ten land cover classes were used for classification with Decision Forest in Idrisi TerrSet.*

Class ID	Class Name	Samples (*n*)
1	Aspen	191
2	Sagebrush steppe	145
3	Conifer	25
4	Agriculture	134
5	Water	290
6	Impervious surfaces	53
7	Basalt	125
8	Riparian	41
9	Cottonwood	76
10	Maple	132

The points for each landcover class were randomly subset in ArcGIS Pro using a 50-50 split. In other words, half of the points were used for training the decision forest models and the other half were reserved for independent validation of the models. The training and validation point feature classes were converted to raster layers with 30-m resolution to match the Landsat 8 imagery and with 10-m spatial resolution to match the Sentinel-2 imagery.

3.3 Decision Forest Modeling

Using the Decision Forest module in Idrisi TerrSet, the RGF files described above in section 3.2.4 were used as inputs for the Decision Forest models. The parameters for each model run were as follows: 3

variables allowed at each split, 100 trees (model iterations), output hard classification image, and output probability images. The Decision Forest was trained using the rasterized training field sites layer described above

3.4 Model Validation

Both models were validated using the Error Matrix (ErrMat) module in TerrSet. The output hard classification model was used as the categorical map image to be tested against the raster layer containing reserved validation field sites as the input ground truth image. ErrMat creates an error matrix that tabulates the different land cover classes assigned in the ground truth image. It determines how many instances of each class were classified in agreement between the output model and the validation sites. Similarly, it also calculates the number of instances where the classified model and validation data do not agree and are thus, are considered in error. Furthermore, ErrMat calculates the error of omission and the error of commission for each class, as well as overall error, and the kappa index of agreement (KIA) for each class and for the overall model.

3.5 Statistically-Derived Probability Thresholds for Aspen Presence

To calculate and compare predicted aspen acreage, the statistically-derived cutoff threshold (SDCT) was used. The mean probability value of reserved points was used as the SDCT for each model resulting in a mean probability of 0.67 for the LDDFM and 0.58 for the SDDFM. Pixels with values greater than or equal to the SDCT for each model were assumed to contain pure, dense aspen and were subsequently used for acreage calculations.

For the watershed-level analysis of probability values, only pixels with a value of 0.25 or higher were used to help produce meaningful, non-zero mean probability values (NOTE: the vast majority of pixels in every watershed in the study area were not predicted to contain aspen and thus have an aspen probability of zero).

3.6 Zonal Statistics by Watershed and Acreage Calculations

Using the HUC08 watersheds basins layer, zonal statistics were run to calculate the acreage of predicted aspen presence at or above the overall probability threshold value for each model (0.67 and 0.58 for LDDFM and SDDFM, respectively). Assuming the area of each pixel predicted to be aspen was composed entirely of aspen, the calculation of acreages followed Equation 2:

Equation 2: $Acreage = pixel\ count * pixel\ area * 0.000247105$

4. RESULTS

Output from the decision forest included a probability layer for each land cover class and a hard classified land cover layer. For the purposes of this study, the aspen probability layer is of most interest along with the hard classified image from each of the two models. The aspen probability layers had pixel values ranging from 0 to 1 where values closer to zero indicate a low probability of aspen presence at a given pixel and values closer to one indicate a higher probability of aspen presence at a given pixel. These were used to derive the SDCT layers for calculating and comparing predicted aspen acreage between the two models. The LDDF aspen SDCT layer was multiplied by the SDDF aspen SDCT layer to determine where both models predicted aspen, then the two layers were subtracted to show where one model predicted aspen and the other did not (**Figure 2**).

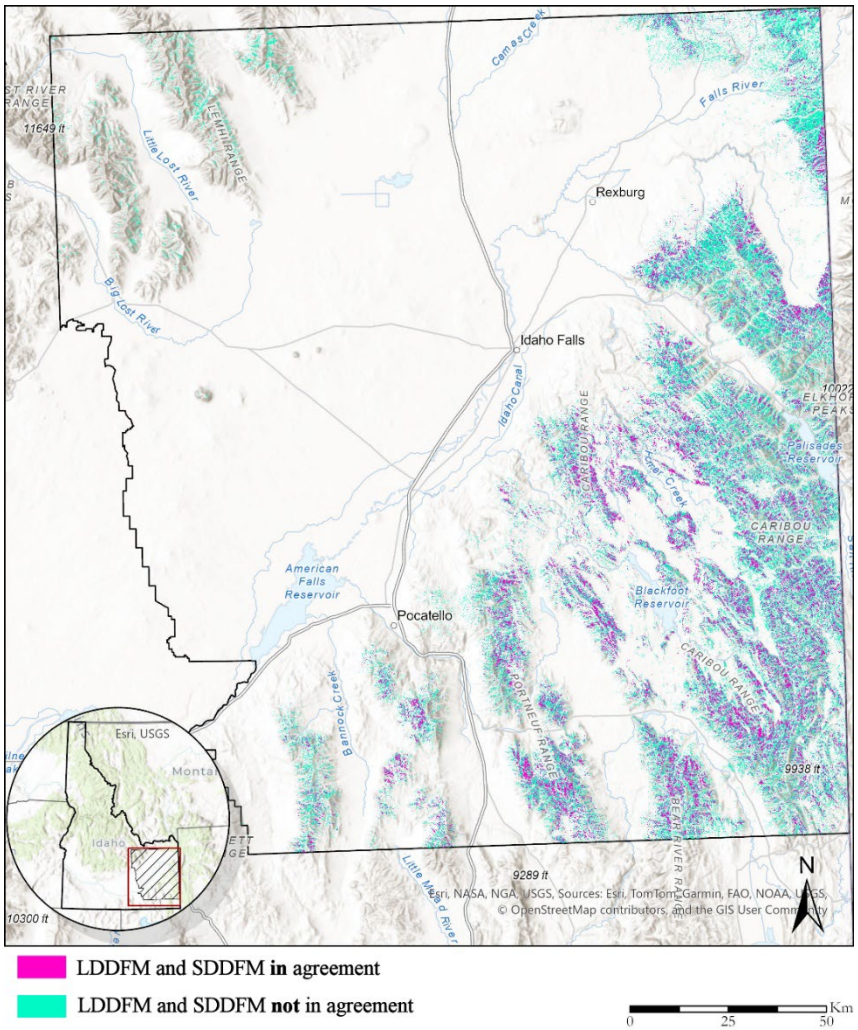

Figure 2. *Comparison of agreement in model-predicted aspen presence above the mean probability threshold from the LDDFM and SDDFM. Pink pixels show where both models predicted aspen and cyan pixels show where one model predicted aspen while the other did not. The SDDFM predicted aspen more often than the LDDFM particularly in the northwest and eastern corners of the study area.*

Total acreage at or above each SDCT shows the SDDFM predicted 51% more aspen across the study area, (783,000 acres) compared to the LDDFM which predicted approximately 467,000 acres (**Table 3**).

When each hard classified model output was validated against the reserved validation field sites, the LDDFM resulted in higher accuracy compared to the SDDFM (**Table 4**). The LDDFM had the fewest number of aspen validation pixels classified incorrectly, the fewest number of other cover type validation pixels classified as aspen, and a higher KIA for both the aspen class and overall model, relative to the SDDFM.

Table 3. *Predicted acreage of aspen within each watershed at or above the study area's mean probability threshold for the Landsat 8 decision forest model (LDDFM) and the Sentinel-2 decision forest model (SDDFM). The percent difference between the predicted acreage of aspen from each model is shown in the column on the right. The predicted acreage and percent difference for the overall study area is shown in the bottom row.*

		Acres		Difference	
ID	**HUC08 Watershed Code**	**LDDFM**	**SDDFM**	**Acreage**	**Percent**
1	16010102	3,198	7,762	4,564	83
2	16010201	34,535	64,602	30,067	61
3	16010202	19,524	32,756	13,232	51
4	16010204	3,890	5,152	1,262	28
5	16020309	2,103	3205	1,102	42
6	17040104	80,244	140,177	59,933	54
7	17040105	54,445	953,56	40,911	55
8	17040201	1,983	2,553	570	25
9	17040202	14,850	33,581	18,731	77
10	17040203	6,482	19,937	13,455	102
11	17040204	37,694	79,216	41,522	71
12	17040205	54,462	66,330	11,868	20
13	17040206	7,417	9,163	1,746	21
14	17040207	82,312	117,117	34,805	35
15	17040208	58,544	81,868	23,324	33
16	17040209	4,866	5,092	226	5
17	17040214	5	0	-5	-200
18	17040215	7	2,853	2,846	199
19	17040216	3	7,519	7,516	200
20	17040217	28	6,835	6,807	198
21	17040218	18	1,819	1,801	196
	OVERALL	**466,614**	**782,894**	**316,280**	**51**

Table 4. *Validation of hard classified model outputs against reserved field sites including Kappa Index of Agreement (KIA) for both the Landsat 8 decision forest model (LDDFM) and Sentinel-2 decision forest model (SDDFM).*

	Aspen Class Accuracy		KIA	
	Users'	**Producer's**	**Aspen**	**Overall**
LDDFM	0.82	0.90	0.87	0.92
SDDFM	0.78	0.77	0.71	0.86

When values from the decision forest model probability output layers for the aspen class were extracted to the validation points for both models, SDCT was 0.67 for the LDDFM and 0.58 for SDDFM suggesting the predictive power of the LDDFM was greater than that of the SDDFM. Results of the watershed-level analysis in modeled probability of pixels with a value of 0.25 or higher showed more

mixed results with two watersheds having the same mean probability value for both models, nine watersheds had higher mean probability values with the LDDFM, and ten watersheds having higher mean probability values with the SDDFM (**Table 5**).

Table 5. *Table of zonal statistics results for the mean of probability values greater than or equal to 0.25 for each HUC08 watershed in the study area.*

		Mean probability ≥ 0.25	
ID	**HUC08 Code**	**LDDFM**	**SDDFM**
1	16010102	0.53	0.55
2	16010201	0.55	0.55
3	16010202	0.53	0.56
4	16010204	0.52	0.49
5	16020309	0.55	0.50
6	17040104	0.52	0.54
7	17040105	0.57	0.57
8	17040201	0.54	0.51
9	17040202	0.47	0.48
10	17040203	0.46	0.50
11	17040204	0.53	0.56
12	17040205	0.54	0.53
13	17040206	0.51	0.46
14	17040207	0.56	0.55
15	17040208	0.51	0.50
16	17040209	0.52	0.46
17	17040214	0.43	0.30
18	17040215	0.31	0.47
19	17040216	0.31	0.48
20	17040217	0.30	0.42
21	17040218	0.30	0.40

5. CONCLUSIONS

Based on validation results, the LDDFM had higher accuracy than the SDDFM. The LDDFM also had a higher mean probability value at validation field sites than the SDDFM. The percent difference between LDDFM-predicted aspen acreage and SDDFM-predicted aspen acreage at or above the mean probability value for the total study area was 51%. At the HUC08 watershed level, the SDDFM predicted higher acreage of aspen than the LDDFM in 20 of the 21 watersheds within the study area.

These results suggest that for studying general trends in aspen population distribution, Landsat's 30-meter spatial resolution is adequate. However, detecting individual aspen trees or small stands for precise inventory is not possible at this resolution. The 10-meter spatial resolution of the SDDFM did not improve model accuracy and predicted nearly twice as much aspen overall. This suggests Sentinel-2 imagery is also not able to provide small stand detection.

A likely source of uncertainty in this study is the way acreage was calculated. The area of each pixel at or above the selected probability threshold value (0.67 for the LDDFM and 0.58 for the SDDFM) was assumed to represent homogeneous, dense stands of aspen. However, given the size of pixels analyzed in the study (900 square meters for Landsat 8, 100 square meters for Sentinel-2), and the proximity of aspen to other cover classes, many --if not most-- of the predicted aspen pixels likely contained other types of vegetation along with aspen. In fact, while aspen is likely present, it may not be considered dominant.

Given the wide range of stand density observed during field work, and visible, but variable understory, Sentinel-2 10-m spatial resolution does not appear to have fine enough spatial resolution to accurately predict individual aspen trees on the landscape. Additionally, a continuous archive of 30-meter spatial resolution Landsat imagery is available from the year 1972 to present, whereas 10-m Sentinel imagery is only available from 2015 to present, giving Landsat-based modeling the advantage of allowing for landscape-level trend analysis of aspen populations across a greater temporal range than would be possible using Sentinel imagery. If land managers are interested in producing a more accurate inventory of existing aspen stands, exploring finer spatial resolution commercial imagery is recommended.

6. REFERENCES

Chavez, P. S. (1996). Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing*, *62*(9), 1025-1035. Retrieved 5 August 2025, from https://www.asprs.org/wp-content/uploads/pers/1996journal/sep/1996_sep_1025-1036.pdf

Kay, Charles E. 1997. Is aspen doomed? *Journal of Forestry.* 95(5): 4–11. Retrieved 5 August 2025, from https://stopthespraybc.com/wp-content/uploads/2013/08/is-aspen-doomed.pdf

Kitchen, S. G., Behrens, P. N., Goodrich, S. K., Green, A., Guyon, J., O'Brien, M., and Tart, D. (2019). Guidelines for aspen restoration in Utah with applicability to the Intermountain West. Retrieved 5 August 2025 from https://digitalcommons.usu.edu/aspen_bib/7812/

Kouki, J., Arnold, K., and Martikainen, P. (2004). Long-term persistence of aspen–a key host for many threatened species–is endangered in old-growth conservation areas in Finland. *Journal for Nature Conservation*, *12*(1), 41-52. https://doi.org/10.1016/j.jnc.2003.08.002

U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture (2023). LANDFIRE, 2023, Existing Vegetation Type Layer, LANDFIRE 2.0.0. Accessed May 2024 at http://www.landfire/viewer.

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution*, *20*(9), 503-510. https://doi.org/10.1016/j.tree.2005.11.006

Richards, J. A. (2013). *Remote sensing digital image analysis* (Vol. 5, pp. 256-258). Berlin/Heidelberg, Germany: springer. http://doi.org/10.1007/978-3-642-30062-2

Romme, W. H., Floyd-Hanna, L., Hanna, D. D., & Bartlett, E. (2001). Aspen's ecological role in the West. *USDA Forest Service Proceedings.* RMRS-P-18, 2001. Retrieved 5 August 2025 from https://research.fs.usda.gov/treesearch/35832

Singer, J. A., Turnbull, R., Foster, M., Bettigole, C., Frey, B. R., Downey, M. C., and Ashton, M. S. (2019). Sudden aspen decline: a review of pattern and process in a changing climate. *Forests*, *10*(8), 671. https://doi.org/10.3390/f10080671

Wilson, Edward O. 1992. The diversity of life. New York: W. W. Norton and Company, 48.