**NASA DEVELOP National Program**



USGS at Colorado State University

*Fall 2015*

Colorado Agriculture

Mapping Forest Species Composition at the Colorado State Forest State Park using Landsat with Integrative Spatial Modeling

**Technical Report** 

Final Draft – November 19th, 2015

Brian Woodward (Project Lead)

Sarah Carroll

Nicholas Kotlinski

Eric Rounds

Dr. Paul Evangelista, Natural Resources Ecology Lab, Colorado State University (Science Advisor)

Tony Vorster, Bioenergy Alliance Network of the Rockies (Mentor)

# I. Abstract

State and National forests are actively managed for a variety of objectives including timber, recreation, wildlife, and livestock grazing. In Colorado, concerns about potential fire severity, aesthetics, and falling dead trees, as a result of the recent mountain pine beetle epidemic, are shared by our partners at Colorado State Forest Service (CSFS) and communities across the Rocky Mountain region. In addition, the Bioenergy Alliance Network of the Rockies (BANR) is assessing the feasibility of using beetle-kill wood as a source of biofuels. These diverse objectives and concerns highlight the need for current and accurate species composition maps to inform management decisions. Past species composition mapping has shown gaps in the ability to accurately highlight individual species at a fine scale over large forested landscapes, and these maps remain relatively coarse in terms of resolution and their usefulness to forest managers. This project utilizes scenes from Landsat 8 (OLI), 7 (ETM+), and 5 (TM), NAIP imagery, forest survey data, and an integrative model developed by researchers at Montana State University in order to create a detailed and highly accurate map of species composition in the Colorado State Forest State Park. Using this combined methodology, this project produced species composition maps for dominant species in Colorado State Forest, including lodgepole pine (*Pinus contorta*), subalpine fir (*Abies lasiocarpa*), Engelmann spruce (*Picea engelmannii*), and aspen (*Populus tremuloides*).

**Keywords**

Species Composition, Colorado, Forest Management, Agriculture, Ecological Modeling, Remote Sensing, Zero-Inflated Model

# II. Introduction

The Colorado State Forest (CSF) is managed for multi-use objectives including timber harvest, recreation, and wildlife. Forest management decisions are based on an adaptive and active strategy that requires accurate and current data to inform management decisions. Presently, these decisions are based primarily on historical records, field observations, and National Agricultural Imagery Program (NAIP) imagery. Although forest managers do have a species composition map of the State Forest, the map is inaccurate due to the lack of revisions over time and the outdated methods used to map tree species. Moreover, the need for an up-to-date species composition map has been exacerbated by drastic changes in forest health caused by a pine-beetle epidemic in the early 2000s. As a result, an accurate map detailing forest species composition throughout the CSF would provide a baseline for ecological studies and timber harvest planning, while at the same time providing a reference for future studies of beetle-kill.

As a part of the Agriculture category of the DEVELOP National Application Area, this project utilizes Landsat imagery in addition to ancillary datasets from fieldwork as inputs into a novel zero-inflated regression tree model designed by researchers at the Spatial Sciences Center at Montana State University. The primary objective of this project is to improve upon current forest management datasets by creating an accurate, up-to-date, and comprehensive record of species distribution and percent canopy cover (PCC) throughout the study area. In addition to improving available data for forest managers, the results should prove useful for ecological studies, as species composition information has important impacts on ecological functionality and resilience (Peterson et al., 1998).

Previous studies have shown some success at mapping PCC in small study areas without reference data (Alistair et al., 2009), while others have successfully used datasets such as LiDAR to map PCC across whole landscapes (Ahmed, 2014; Carreiras, 2006). However, this study solely relies upon freely available satellite imagery from the USDA and USGS. As such, the combination of timely and accurate field data, spectral remote sensing products, and zero-inflated modeling processes should allow for a highly accurate view of species composition in the Colorado State Forest, and in turn could be applied in larger-scale studies.

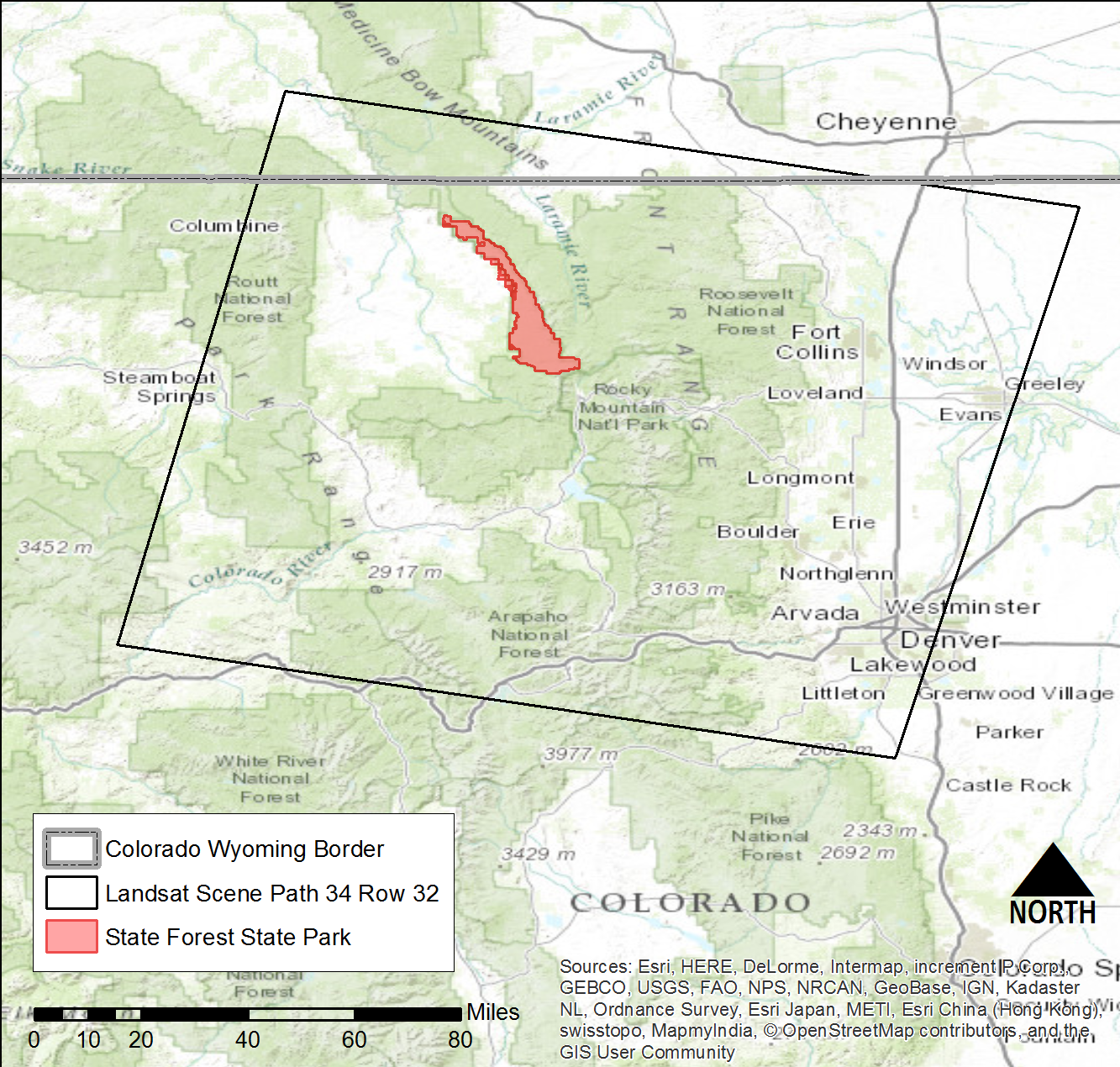


Figure 1 Colorado Agriculture study area in north-central Colorado

The study area focuses on the Colorado State Forest in north-central Colorado (Figure 1), which stretches across 70,000 acres through Jackson and Larimer counties. The area was established as a state park in 1970 and is now actively managed by Colorado Parks and Wildlife and the Colorado State Forest Service. Elevation in the park ranges from approximately 2,500 to nearly 4,000 meters and more than seventy percent of the acreage is covered with coniferous forests (Colorado Parks & Wildlife). Lodgepole pine *(Pinus contorta*) and fir-spruce communities dominate the subalpine zone. Other dominant species include: douglas fir (*Pseudotsuga menziesii)*, Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa)*, and quaking aspen (*Populus tremula*) (Wiken et al., 2011).

Our project partners and end-users include Tony Vorster of the Bioenergy Alliance Network of the Rockies (BANR) and John Twitchell of the Colorado State Forest Service (CSFS). Tony Vorster is part of the feedstock supply team with BANR, which is a large, collaborative project assessing the feasibility of using beetle-kill wood as a feedstock for biofuel. A map of species composition can be utilized by BANR to map live and dead biomass as they assess the location and quantity of potential biofuel feedstocks in beetle-kill areas. John Twitchell is a District Manager for the CSFS, the group charged with managing the diverse uses of the CSF. The CSFS is currently in need of an up-to-date, highly accurate, and comprehensive species composition map to aid in ecological understanding and management of the CSF. The methods and results presented in this report may prove useful far beyond the scope of the study area, as forest managers across the country can use these methods to carry out accurate, timely, and cost-effective assessments of tree species composition in a number of diverse ecological settings.

# III. Methodology

**Data Acquisition**

Three Landsat 8 OLI scenes for WRS path 34, row 32 were downloaded from USGS Earth Explorer: two summer scenes (July 2014 and June 2015), and one fall scene (September 2014). July and August scenes from 2002 were also downloaded in order to assess mapping results using pre-beetle kill imagery. The scenes were downloaded as Level 1 data products from USGS Earth Explorer. Shuttle Radar Topography Mission (SRTM) data was obtained from USGS Earth Explorer, providing the team with a 30-meter digital elevation model (DEM) of the study area. In addition, a 2014 NAIP image containing the study area was downloaded from the USDA.

Reference data on the CSF was collected during fieldwork undertaken by BANR and, in part, the summer DEVELOP team. The reference data provides forest composition information (e.g. species, status, diameter at breast height [DBH], and densiometer readings) for sixty-seven 30x30-meter plots within the CSF. The geographic center of these plots corresponds with the center of Landsat pixels. Other ancillary data utilized in this project include a CSF administrative boundary shapefile and a point shapefile with the locations of plots surveyed during the BANR fieldwork.

**Data Pre-processing**

*BANR forestry data*

To successfully replicate the methods documented by Savage et al. (2015), several pre-processing steps were necessary to format the forest plot data provided by BANR. First, we excluded all harvested plots from the data set, in order to avoid confusion in the model. Species composition data was used to create relative canopy cover percentages for the four focal species within all sixty-seven non-harvested plots surveyed by BANR. By deriving the sum of the total basal area for the given species within a plot and then converting those basal area values into percentages, we were able to determine the percent canopy cover for each species. As a final step, these canopy cover percentages were then scaled to account for open sky with the densiometer readings from each plot. We produced variants of reference data for the four focal species to test the effect of canopy cover variables on the accuracy of model outputs.

* All trees (live and dead canopy cover included)
* Live trees only (dead canopy cover excluded)
* Dead Lodgepole Pine only

*Imagery formatting*

Raster data was formatted to the specifications provided by Savage et al. (2015). All imagery acquired for the model was first converted from the standard TIFF (.tif) format to ERDAS Imagine (.img) file format. Slope and aspect rasters were then generated from a 30-meter DEM in ArcMap using the spatial analyst toolbox. Texture mean and minimum rasters were generated from the NAIP image using the focal statistics tool in ArcMap. These images were then resampled to a 30-meter pixel size to correspond with Landsat and SRTM data.

*Creating a composite image*

After all of the imagery was formatted properly, the next step was to create a single stack composite in ArcMap. The stack (see *Table 2 & 3*, Appendix) consists of select raw Landsat bands for each year’s scene, a DEM, slope and aspect rasters, and NAIP texture rasters. To limit the analysis to the study area, this stack was then restricted to the Colorado State Forest boundary using the extract by mask tool in ArcMap. Finally, a point shapefile containing the pixel locations of the corresponding reference data was used to extract the digital number values in Spatial Analyst (Extract Multi Values to Points) from each band in the composite for all sixty-seven plots surveyed. The resulting digital numbers were then compiled in a table with percent canopy cover values for each plot, creating a final spreadsheet of reference data for model input (see *Figure 3*, Appendix).

**Zero-Inflated Modeling**

To produce continuous species composition maps, we utilized a novel zero-inflated (ZI) regression model developed by the Spatial Sciences Center at Montana State University. Species distribution models are often based on presence/absence data, which frequently contain a high number of zero observations for a given species. Zero-heavy data have potential to bias model outputs and ZI models have been shown to be successful at eliminating or reducing these effects (Wenger, 2008). The ZI process involves modeling in two steps, in which the first step predicts presence or absence, and the second predicts spatially continuous values for presence only, thus eliminating the influence of zero data on continuous outputs. The ZI model developed by Savage et al. (2015) was developed to predict continuous cover in western mixed conifer forests and in turn, was applicable to our study area.

*Step 1: Model Testing and Training*

The first step of the ZI model trains the model with reference data and evaluates performance with a ten-fold cross validation, a Wilcoxon’s Signed-Rank test, and calculates an RMSE (root mean square error) statistic. The output from this step is a table displaying a p-value, mean difference, RMSE, and confidence interval for each fold. The ten values from each fold were then averaged to obtain a single value for the above statistics. Savage et al. (2015) determined that a combination of Random Forest and Support Vector Machine models performed best (i.e., a minimum confidence interval width and low RMSE). After analysis of the results from Step 1, and given the congruence between reference data utilized by Savage et al (2015) and our own, we opted to move forward with the same model combination (RF-SVM) in our analysis.

*Step 2: Spatial Prediction of Binary and Continuous Cover*

To successfully produce canopy cover maps for the four target species within the study area, Step 2 of the model required the input of a single image stack in conjunction with the reference data from Step 1. To predict presence and absence for a given species, the image stack and reference data were fed into the Random Forest model (RF), resulting in a binary (BIN) map of species presence (1) or absence (0). Next, the model utilizes all non-zero (species present) data in a Support Vector Machine (SVM) to create a continuous (CON) map of percent canopy cover. The two associated maps (BIN & CON) for each species were then combined using the raster calculator tool in ArcMap. The resulting map displays the continuous values where the binary map displayed species presence. Any pixel marked as absent (0) was determined to be no data for that particular species. This process was repeated for each focal species. In order to test for the influence of the mountain pine beetle epidemic on outputs, we ran Step 2 with both a 2002 and 2014/15 image stack.

# IV. Results & Discussion

**Results and Analysis**

*Lodgepole Pine (Pinus contorta)*

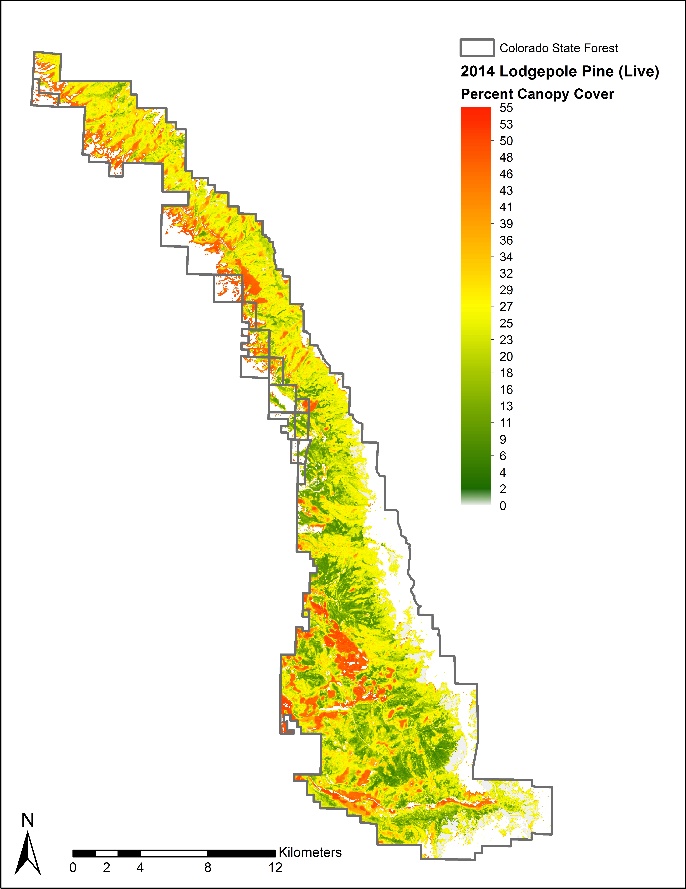
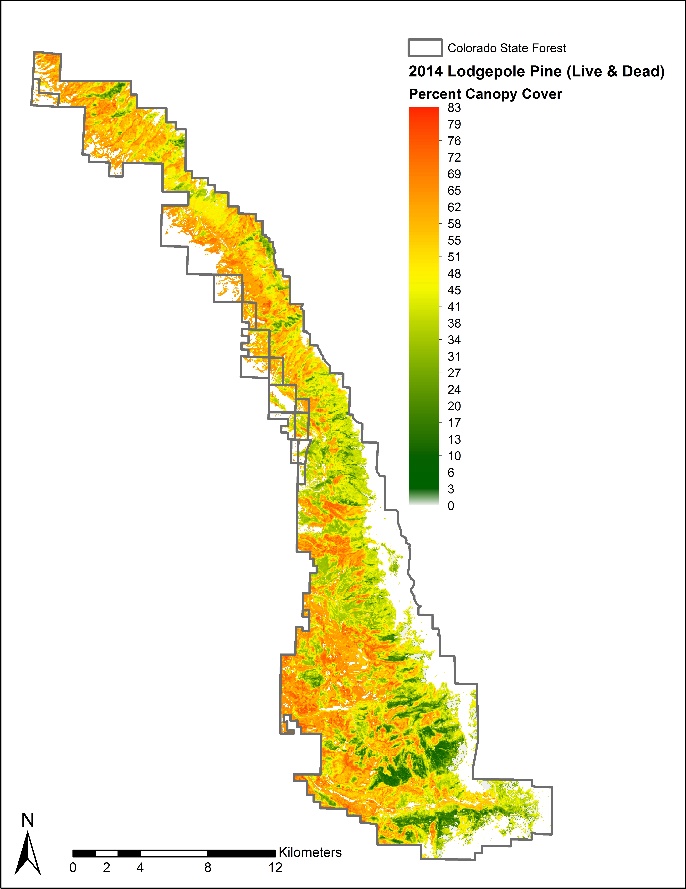


Figure 2 Lodgepole pine percent canopy cover. On the right, only live canopy cover is represented. On the left, live and dead canopy cover are represented for the species.

Spatial outputs from the model show that *Pinus contorta* (Lodgepole pine) is the dominant forest species at Colorado State Forest State Park (Figure 2). The maps (displayed above) were produced under two scenarios: 1) using live and dead percent canopy cover data for the species, and 2) using only live percent canopy cover data for the species. Live lodgepole percent canopy cover ranged from a low of 0 to a high of 55 percent. Live and dead lodgepole percent canopy cover combined ranged from a low of 0 to a high of 83 percent. Since the mountain pine beetle epidemic killed a large percentage of *Pinus contorta* in the study area, it is logical that a live and dead combination would result in a higher percent canopy cover. Dead canopy cover is known to make up a large percentage of overall canopy cover at the CSFSP, so the model output that combines live and dead canopy cover should be used by forest managers and partners.

*Quaking Aspen, Subalpine Fir, Engelmann Spruce*

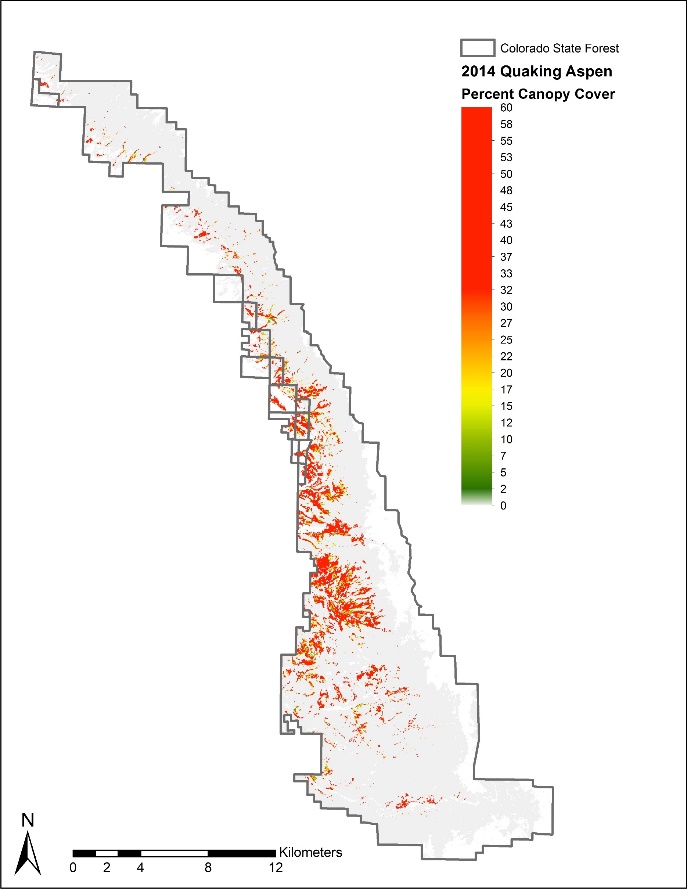
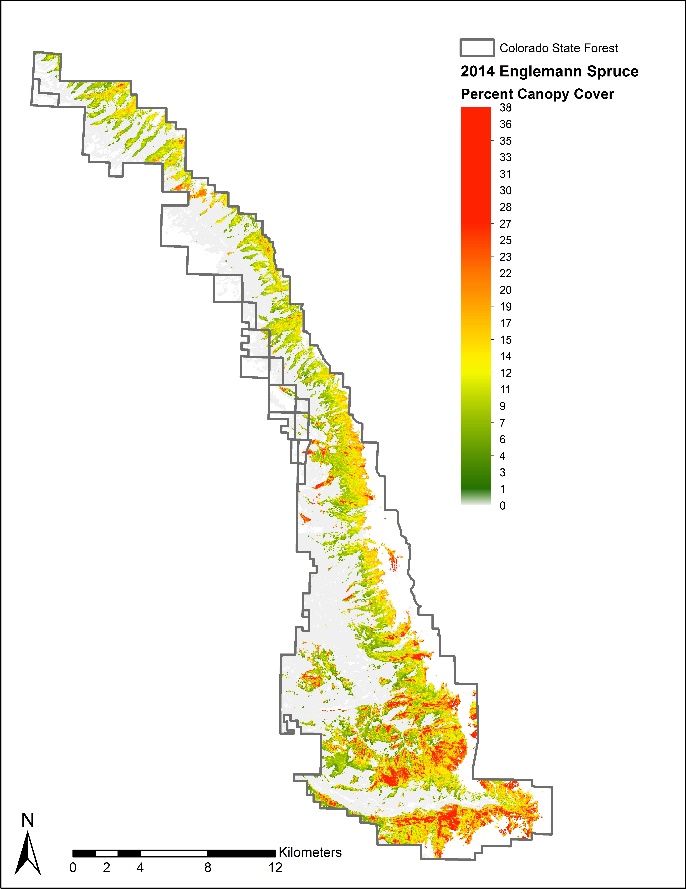
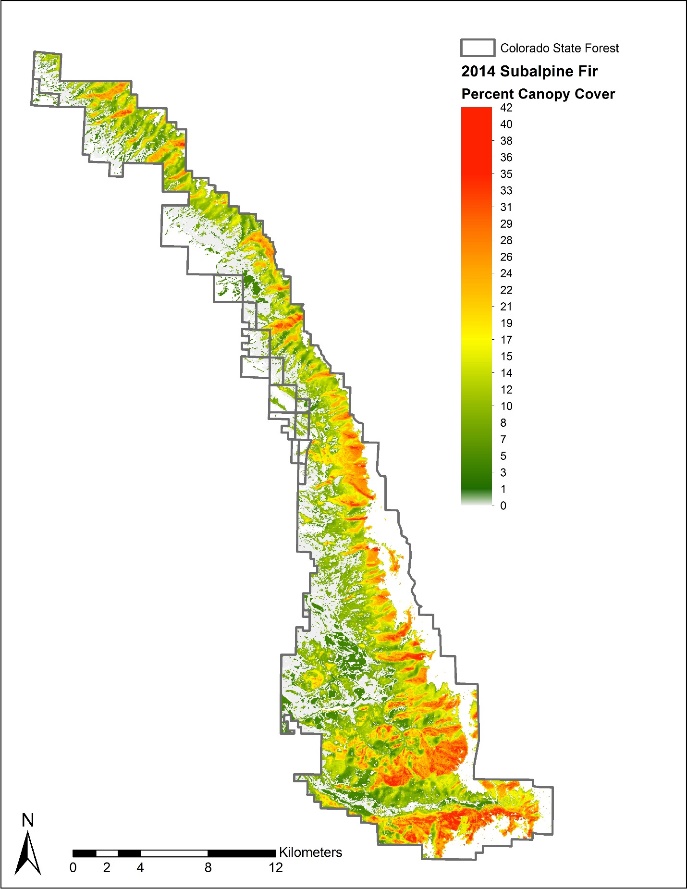


Figure 3 Percent canopy cover end-products for subalpine fir, Englemann spruce, and Quaking aspen.

The less dominant species at the CSFSP include Engelmann spruce, subalpine fir, and quaking aspen (Figure 3). Subalpine fir and Engelmann spruce thrive in similar environments, and are often found together in mixed-species forest stands (Alldritt-McDowell, 1998). Our model outputs for the two species show that fir and spruce are often present in the same areas of the forest, although Engelmann spruce is restricted to mostly higher elevation areas of the forest, while Subalpine fir is not. Subalpine fir canopy cover ranged from 0 to 42%, while Engelmann spruce percent canopy cover ranged from 0 to 38%. Quaking aspen is often found in small stand clusters below the subalpine zone, which our model outputs display (Elliot, 2004). When present, aspen is often the dominant species in the small stand clusters that it inhabits, which makes our model results displaying fairly consistent percent canopy cover for the species of ~60% to be quite plausible.

**Error and Uncertainty**

The RMSE statistic and confidence interval were assessed for each species in order to interpret the accuracy and precision of the model. The RMSE values ranged from 11-14%, suggesting that the model predicted canopy cover with reasonable accuracy, and that there were no extreme under or over estimates for any of the four species. Overall, the 95 % confidence interval widths were narrower for live only canopy cover reference data signifying that dead trees may be confusing the model. This is particularly evident in the results for lodgepole pine, where the confidence interval width increased by 6% on average when dead canopy cover was included. However, the level of precision achieved by the model is a significant improvement on existing species composition maps and is highly informative to forest managers who make decisions over larger landscape scales.

|  |  |
| --- | --- |
| SPECIES | Approximate Confidence Interval Width |
| Lodgepole pine | ± 9 % (live only) / 15% (live & dead) |
| Subalpine fir | ± 14 % |
| Engelmann spruce | ± 13 % |
| Aspen | ± 19 % |

*Table 1. Approximate confidence interval widths of the four study focal species*

At this stage, model outputs have not been verified against independent datasets, as the group did not have sufficient data to do so. To further assess mapping accuracy, the outputs could be validated by a comparison with high-resolution imagery to identify possible discrepancies between true cover and the model’s distribution outputs. Another critical consideration is that the collection of additional field data would increase the sample size of training data, likely bolstering model performance and improving accuracy of outputs. Both of these validation techniques would be worthwhile pursuits in the future.

# V. Conclusions

The zero-inflated model designed by Savage et al. (2015) proved to be both a simple and an effective means of mapping percent canopy cover in the CSF. Even with fairly limited field reference data, the model outputs were mapped with a high level of accuracy. The data provided by the model and the resulting PCC maps are a significant improvement when compared to existing datasets. Our maps provide an estimate of PCC for each species on pixel basis, thus providing multiple unique canopy values for a given pixel. The end products provided to our project partners include a total of nine distinct canopy cover maps for each focal species and two overall species composition maps displaying the dominant canopy cover by species for 2002 and 2014.

The model and methodology utilized provided a simple and effective means to map percent canopy cover with limited field reference data. A major advantage of the model is that it can be applied to existing datasets that are regularly collected by forests managers, meaning that this study can be replicated and expanded to include larger study areas in ecologically diverse regions. Due to the flexibility of the model, it will likely be of great assistance to forest managers throughout the country who are in need of more detailed species composition maps.

# VI. Acknowledgments

This project was made possible through support and mentorship of select individuals and organizations. We extend a kind thanks to the following:

* Dr. Paul Evangelista, Colorado State University (CSU), Natural Resource Ecology Laboratory (NREL)
* Tony Vorster, CSU, NREL, Bioenergy Alliance Network of the Rockies (BANR)
* Shannon Savage, Montana State University (MSU)

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 NNL11AA00B and cooperative agreement NNX14AB60A.

# VII. References

Ahmed, O.S., Franklin, S.E., and Wulder, M.A. 2014. Integration of Lidar and Landsat

data to estimate forest canopy cover in coastal British Columbia.

*Photogrammetric Engineering & Remote Sensing*: Volume 80, pp. 953-961.

Alldritt-McDowell, Judith. Ecology of the Engelmann Spruce-Subalpine Fir Zone. (1998).

Alistair, M.S., Falkowski, M.J., Hudak, A.T., Evans, J.S., Robinson, A.P., and Caiti M. Steele.

2009. A cross-comparison of field, spectral, and lidar estimates of forest canopy

cover. *Canadian Journal of Remote Sensing*: Vol. 35, No. 5. pp. 447-459.

Carreiras, J.M.B., Pereira, J.M.C., and Pereira, J.S. 2006. Estimation of tree canopy cover

in evergreen oak woodlands using remote sensing. *Forest Ecology and*

*Management*: Volume 223, pp. 45-53.

Colorado Parks & Wildlife. Colorado Parks & Wildlife. http://cpw.state.co.us/placestogo/

parks/ StateForest/Pages/default.aspx.N.p.,n.d. Web. 01 Oct. 2015.

Elliott, Grant P., and William L. Baker. "Quaking aspen (Populus tremuloides Michx.) at

treeline: a century of change in the San Juan Mountains, Colorado, USA." *Journal of Biogeography* 31.5 (2004): 733-745.

Peterson, G., Allen, C.R., and Holling, C.S. 1998. Ecological Resilience, biodiversity, and

scale. *Ecosystems,* 1, pp. 6-18.

Savage, S.L., Lawrence, R.L., and John R. Squires. 2015. Predicting Relative Species

Composition within Mixed Conifer Forest Pixels Using Zero-Inflated Models and Landsat Imagery. Unpublished Manuscript (*In Review*).

Wenger, Seth J., and Mary C. Freeman. "Estimating species occurrence, abundance,

and detection probability using zero-inflated distributions." Ecology 89.10 (2008): 2953-2959.

Wiken, Ed, Francisco Jiménez Nava, and Glenn Griffith. 2011. North American Terrestrial

Ecoregions—Level III. Commission for Environmental Cooperation, Montreal, Canada. pp. 88-92.

# VIII. Content Innovation

**Glossary**

**Basal Area –** a measurement of the cross-sectional area of a tree’s trunk at breast height that is easily calculated from the diameter at breast height (DBH).

**Composite Band –** ArcMap data management tool that creates a single raster stack from a selection of multiple bands.

**95 % Confidence Interval –** Displays theamount of uncertainty associated with a predicted statistic in an estimated interval, providing that if the same sampling method was repeated numerous times, we could expect the true value to be contained within the estimated interval 95% of the time.

**Digital Numbers** **–** Individual pixel values that correspond to levels of reflectance intensity.

**Focal Statistics –** ArcMap spatial analyst tool that calculates statistics (mean, max, min, range, standard deviation) for a specified neighborhood of pixels within an image.

**Generalized Linear Models (GLM) –** An indiscriminate method of linear regression, in which the response variable being assessed is allowed a discrete distribution of errors as opposed to a normal distribution. Using link functions, GLM’s allow for flexibility by predicting response variables that might be related to independent variables non-linearly.

**Landsat –** Started in the 70s, Landsat is a joint USDA and NASA program that provides multi-spectral moderate resolution (30m) satellite imagery for the whole globe.

**National Agriculture Imagery Program (NAIP) –** A USDA program starting in 2003 that gathers 1-meter resolution aerial imagery of the continental US during the agricultural growing season. NAIP imagery is acquired and made available in 3-year cycles.

**Raster Calculator –** A spatial analyst tool that makes it possible to perform Map Algebra calculations on raster pixel values.

**Root Mean Squared Error –** A statistic that highlights accuracy of a model by measuring difference between values predicted by the model and observed values. It is calculated by taking the square root of the mean square error or the variance.

**Shuttle Radar Topography Mission (SRTM) –** A joint project between the NGA and NASA that gathered topographic digital elevation data for the majority of the globe.

**Support Vector Machine (SVM) –** A method used in classification and regression analyses that utilizes learning algorithms to interpret data.

**Wilcoxon’s Signed-Rank Test –** A statistical test for paired datasets that lack a normal or parametric distribution. The test assesses the difference between an observed median of a data set and a predicted or hypothetical median to determine if the distributions are comparable.

**Zero-Inflated –** A method used to model data that has a high number of zero values or absence data so that to reduce or remove the bias effects of zero heavy data ie presence absence data.

**ArcGIS Online Interactive Map**

<http://www.arcgis.com/apps/Viewer/index.html?appid=1c9262794fca4e19b2488f73c17a8c10>

**KML Interactive Map Viewer**

<https://drive.google.com/file/d/0B_rR-6Z_Sx19dEhmM0xaX2NWd0E/view?usp=sharing>

# IV. Appendices

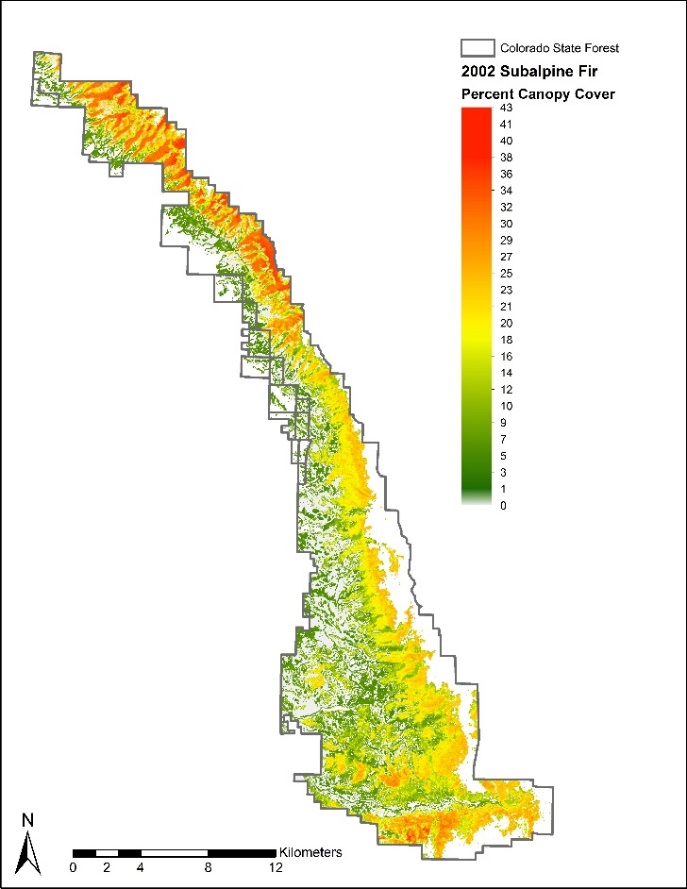
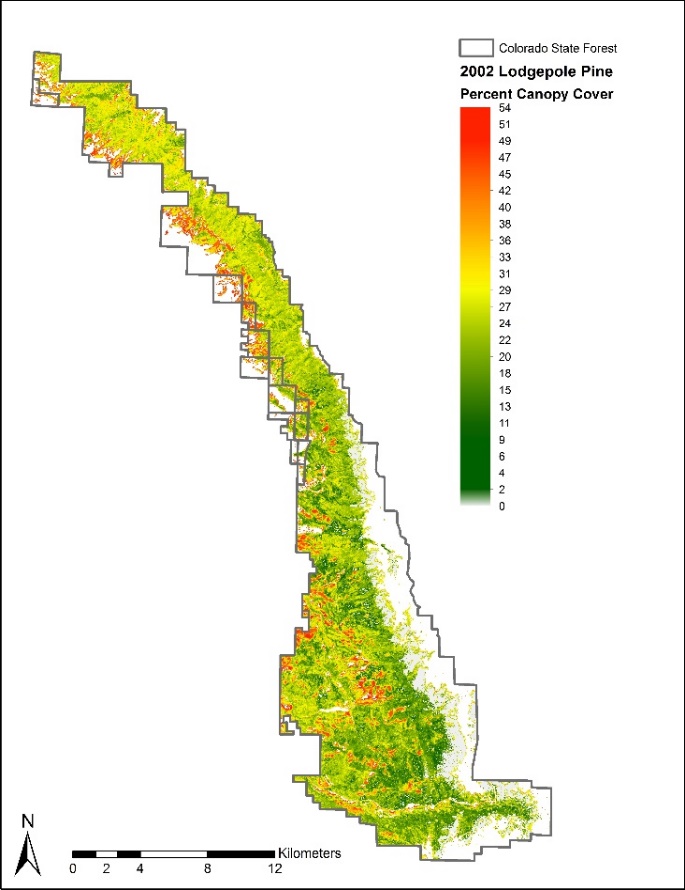
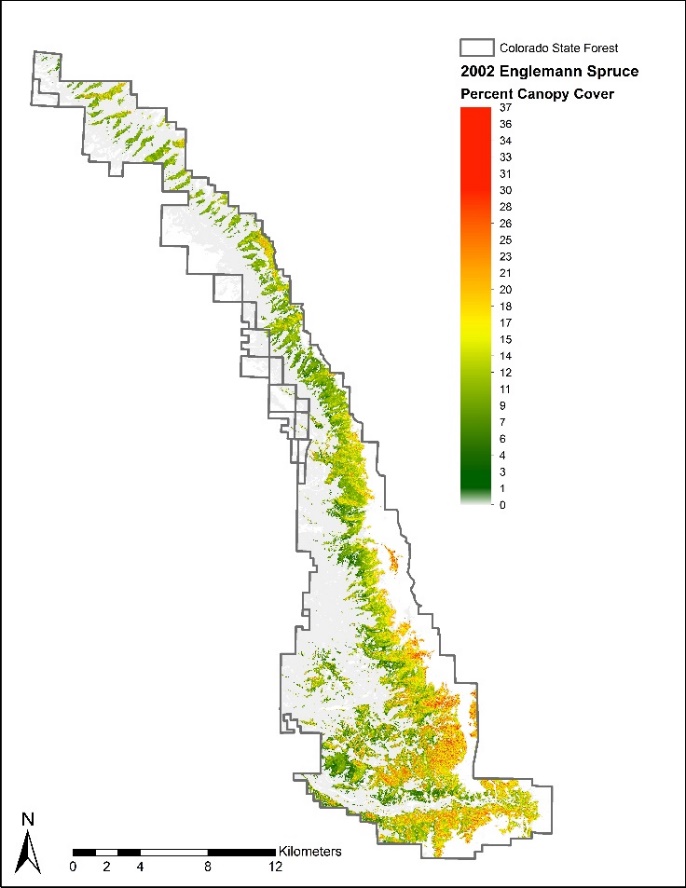
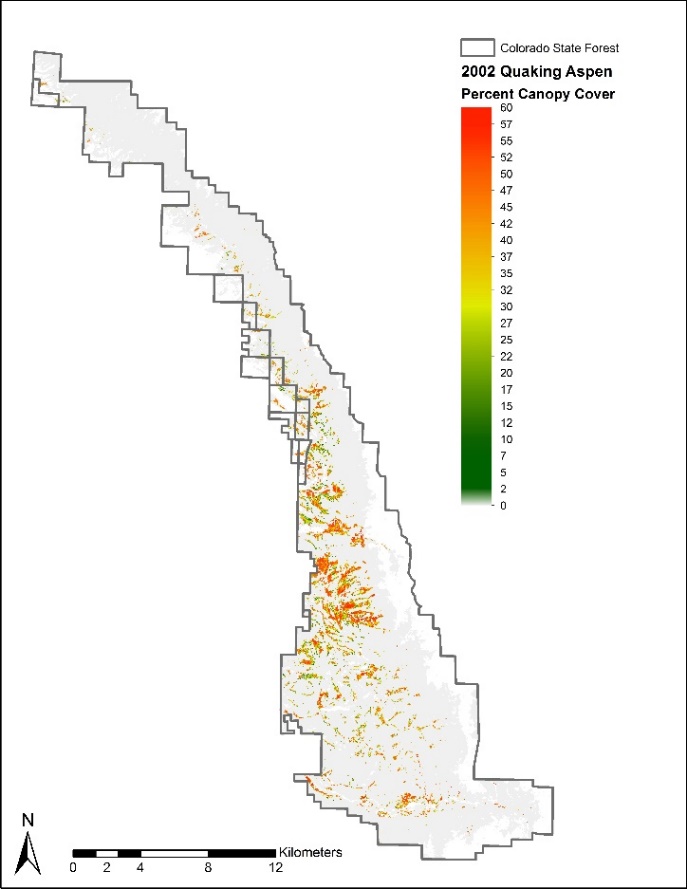


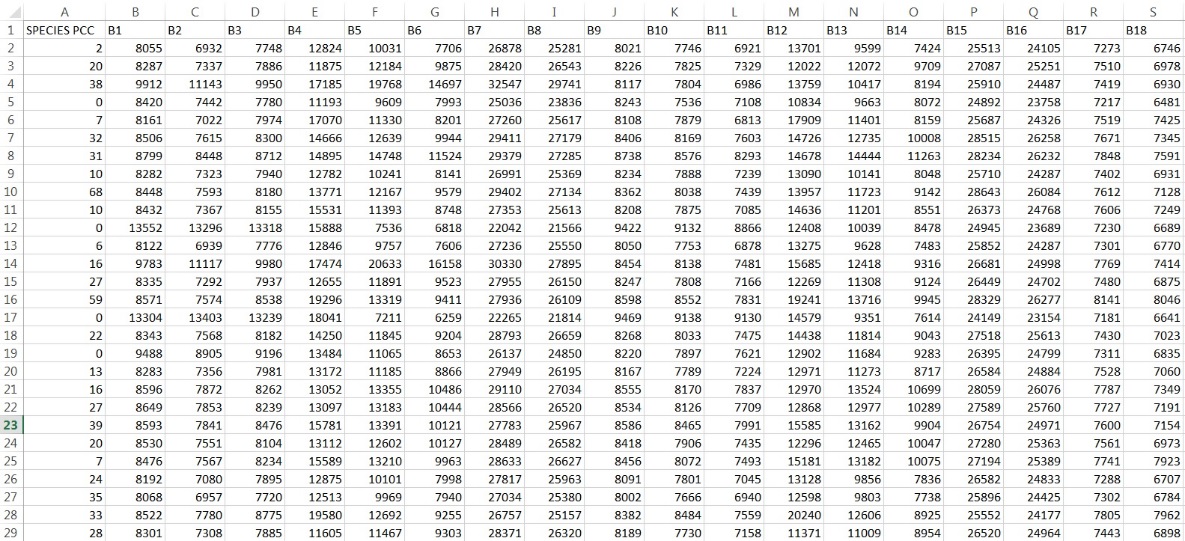
Figure 2. Species composition maps for Year 2002. Each of the maps displays percent canopy cover at a 30m resolution for each of the focal species at the Colorado State Forest State Park. Year 2002 end products were produced as a comparison between pre and post mountain pine beetle outbreak species composition.

|  |  |
| --- | --- |
| Model Image Stack  Band Number | Landsat (Year - Band) / Raster Data Source  July (TM) /August (ETM+) |
| 1 | July 2002 - Blue |
| 2 | July 2002 - Green |
| 3 | July 2002 - Red |
| 4 | July 2002 - NIR |
| 5 | July 2002 - MIR1 |
| 6 | July 2002 - MIR2 |
| 7 | July 2002 - Thermal |
| 8 | August 2002 - Blue |
| 9 | August 2002 - Green |
| 10 | August 2002- Red |
| 11 | August 2002- NIR |
| 12 | August 2002 - MIR1 |
| 13 | August 2002 - MIR2 |
| 14 | August 2002 - Thermal |
| 15 | SRTM DEM (30m) |
| 16 | Slope |
| 17 | Aspect |
| 18 | NAIP Texture Mean |
| 19 | NAIP Texture Minimum |

Table 2. Stacked image combination for 2002 based on specifications from Savage et al. (2015)

|  |  |
| --- | --- |
| Model Image Stack  Band Number | Landsat 8 (Year - Band) / Raster Data Source |
| 1 | July 2014 - Blue |
| 2 | July 2014 - Green |
| 3 | July 2014 - Red |
| 4 | July 2014 - NIR |
| 5 | July 2014 - MIR1 |
| 6 | July 2014 - MIR2 |
| 7 | July 2014 - TIR1 (Thermal) |
| 8 | July 2014 - TIR2 (Thermal) |
| 9 | September 2014 - Blue |
| 10 | September 2014 - Green |
| 11 | September 2014 - Red |
| 12 | September 2014 - NIR |
| 13 | September 2014 - MIR1 |
| 14 | September 2014 - MIR2 |
| 15 | September 2014 - TIR1 (Thermal) |
| 16 | September 2014 - TIR2 (Thermal) |
| 17 | June 2015 - Blue |
| 18 | June 2015 - Green |
| 19 | June 2015 - Red |
| 20 | June 2015 - NIR |
| 21 | June 2015 - MIR1 |
| 22 | June 2015 - MIR2 |
| 23 | June 2015 - TIR1 (Thermal) |
| 24 | June 2015 - TIR2 (Thermal) |
| 25 | SRTM DEM (30m) |
| 26 | Slope |
| 27 | Aspect |
| 28 | NAIP Texture Mean |
| 29 | NAIP Texture Minimum |

Table 3. Stacked image combination for 2014/15 based on specifications from Savage et al. (2015)

*Figure 3: Sample reference data sheet. The left-most column includes percent canopy cover data on a 0-100 scale. The remaining columns contain spectral values extracted for each band of the image composite.*