NASA DEVELOP National Program

Colorado
Spring 2023

Front Range Wildland Fires

Evaluating the Efficacy of Remote Sensing Imagery in Monitoring Forest Fuels Treatment Methods

DEVELOP Technical Report

Nora Carmody (Project Lead)
Lillian Gordon
Nathan Teich
Josh Virene

Advisors:
Nicholas Young, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)
Dr. Tony Vorster, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)
Dr. Paul Evangelista, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)
Christopher Tsz Hin Choi, Colorado State University, Natural Resource Ecology Laboratory (Science Advisor)
Dr. Catherine Jarnevich, Colorado State University, United States Geological Survey (Science Advisor)

Fellow:
Sarah Hettema (Colorado)

1. Abstract

Over the last several decades, wildfire frequency and severity in forested areas along Colorado's Front Range have increased due to a buildup of fuels. This has led to an increase in forest treatments, as well as an increased need to evaluate the success of these treatments. Remote sensing products offer an efficient and cost-effective way to monitor forest treatments; however, not all remote sensing products and analysis techniques have been explored by Coloradan land managers. Specifically, project partners at the Colorado State Forest Service (CSFS) and the Colorado Forest Restoration Institute (CFRI) were interested in using an effective and streamlined method of mapping canopy cover to better monitor forest treatment success. To support their needs, the NASA DEVELOP Front Range Wildland Fires team explored National Agricultural Imagery Program (NAIP) imagery at different spatial resolutions and numbers of training points with NASA's Shuttle Radar Topography Mission (SRTM) Data Elevation Model (DEM) as a predictor in addition to NAIP imagery spectral predictors. From this analysis, we created classified canopy cover rasters, and compared accuracy metrics across model iterations. We also determined that the best performing model, with an overall accuracy of 0.900 uses 2021 NAIP imagery at 2-meter resolution, 800 training points, 200 testing points, does not use topographic predictors, and reclassifies shadow pixels via a pre-selected NDVI threshold.

Key Terms

Remote Sensing, NAIP, Google Earth Engine, Random Forest, ocular sampling, shadow

2. Introduction

2.1 Background

Changes in land use, grazing, and fire suppression practices have impaired ecosystem functions on Colorado's Front Range, leading to a build-up of fuel in ponderosa-pine and mixed conifer forests (Larson & Churchill, 2012). In response, organizations such as the Colorado State Forest Service (CSFS) and the Colorado Forest Restoration Institute (CFRI) increased the size and number of forest treatments and forest treatment monitoring with goals such as reducing fuel, enhancing landscape heterogeneity, and restoring pre-colonial forest conditions (Cannon et al., 2018). Forest monitoring has been vital in assessing the success of these treatments but is time consuming and costly. Remote sensing products, however, offer an efficient and cost-effective way to monitor different forest treatments. Pairing remote sensing imagery products with classification algorithms provided a user-friendly means of detecting changes in canopy cover pre and post treatment, but not all remote sensing products are created equal; imagery attributes such as resolution and extent are important considerations for project success. While current remote sensing classification methodologies derived from Cannon et al. (2021), Dickenson & Pelz (2012), and Ritika et al. (2021) have been beneficial, certain aspects could be improved upon. Specifically, the partners expressed that their current methods involve the use of discrete lidar collections, which have varying temporal availability, unlike a high spatial resolution, regularly released remote sensing dataset like the National Agricultural Imagery Program (NAIP). Dickenson & Pelz (2012) did provide an alternative methodology that utilized imagery with better spatial and temporal attributes but was not streamlined in terms of software utilization. Ritika et al. (2021) acknowledged that Google Earth Engine (GEE) as a powerful and efficient tool in accessing and processing remote sensing data to accomplish image classification with Random Forest, however, their methodologies were used to classify land use and land cover, while our project focused on exclusively canopy cover classification. This project explored how effectively NAIP imagery could be classified via a Random Forest (RF) algorithm across different years, scales, forest types, and shadow classifications exclusively in GEE. Specifically, we explored imagery for 2013 and 2021 in the greater Ben Delatour Scout Ranch (BDSR) region as well as 2019 imagery for the Upper Monument Creek (UMC; Figure 1) region along the Colorado Front Range at a 1m, 2m, and 3m resolution. By adapting methodologies such as those found in previous literature, we can provide partners with imagery that has improved spatial and temporal resolution, and spatial extent, while streamlining software use. This approach would assist Colorado's public land managers in more efficiently and accurately monitoring forest treatments, giving them more capacity in their monitoring capabilities that will improve their adaptive forest management practices.

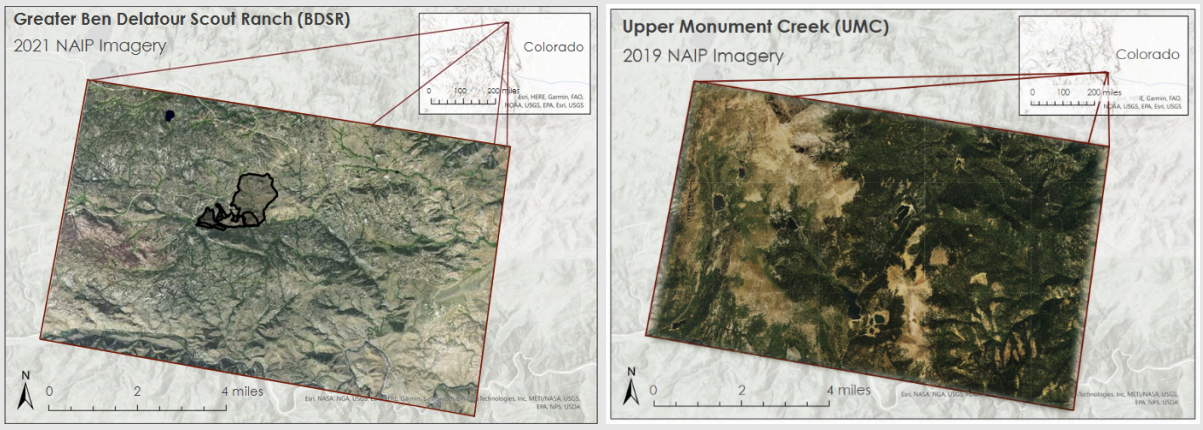


Figure 1: Greater Ben Delatour Scout Ranch (Left) and Upper Monument Creek (Right) study areas

2.2 Project Partners & Objectives

Our partners for this project include the CFRI and the CSFS. CFRI is a research institute that provides scientific knowledge to improve forest management practice for researchers, managers, and stakeholders. CSFS is a state governmental agency whose mission is aimed at providing technical expertise and finding new approaches to adaptive management in Colorado forests. CFRI currently has a robust remote sensing capability for forest treatment monitoring. Using data from the high spatial resolution satellites including WorldView-02, GeoEye-01, and Quickbird-02 to monitor canopy cover, they upload their data to ArcGIS Pro and rescale to a 3-m resolution before processing their imagery in R. While this approach provides meaningful results, CFRI has expressed needs for a more streamlined approach (i.e., completing the entire analysis within a single GIS software, like GEE) using imagery with higher spatial and temporal resolution and better spatial coverage. CSFS has limited use of remote sensing practices in their treatment monitoring and has expressed an interest in building capacity to use more geospatial data. In our partnership with these two groups, we supported their objectives by testing different imagery parameters such as image scale, image quality, and various pre-processing techniques. We provided a comparison of the strengths and weaknesses of these different parameters and evaluate their efficacy in classification. This yielded a tested and reproducible methodology that allows our partners to create imagery products that can be effectively classified for canopy cover and openings in forests.

3. Methodology

3.1 Data Acquisition

To begin our data acquisition, we first created polygon boundaries in GEE for our two study regions: the greater BDSR region and UMC region. Next, we acquired NAIP imagery for the years 2013 and 2021 for the greater BDSR region and 2019 NAIP imagery for the UMC region. For 2013 and 2019, we acquired the NAIP imagery directly from the GEE data catalog. The 2021 NAIP imagery was downloaded for Larimer County from the United States Department of Agriculture (USDA) Geospatial Data Gateway. Additionally, we acquired a 30-meter resolution NASA Shuttle Radar Topography Mission (SRTM) elevation dataset for Colorado directly from GEE's data repository.

Given more time, this project would have additionally analyzed NASA and USGS Landsat 8 & 9 Collection 2 Tier 1 TOA Reflectance data in addition to NAIP imagery. Given our partner's desire for imagery with a high spatial and temporal resolution, we chose to prioritize NAIP imagery over the spaceborne multispectral data in order to focus the project on analysis that aligned with the objective of our partners. NAIP imagery has a spatial resolution of 0.8 m and is typically collected every 3 years. With additional time, we would have explored using multispectral Landsat 8 & 9 data, which has a 30 m resolution and makes a complete orbit of the Earth every 16 days.

3.2 Data Processing

3.2.1 Image Processing

To process our 2013 and 2019 NAIP imagery, we filtered the NAIP image collection to only include data from January 1 – December 31 of the desired years. For both years, we transformed the image collection into a single image using the ImageCollection.mosaic() function. After transforming to an image, we used the Image.clip(region) function to clip the 2013 and 2019 images to the greater BDSR and UMC regions respectively.

To process the 2021 data, we uploaded the entire 2021 true and false color NAIP imagery for Larimer County into ArcGIS Pro. We then clipped the data to our study region and changed the geographic coordinate system from NAD1983 to WGS1984 to be compatible with GEE. Next, we converted the true and false color raster files from a .sid to a .tiff file type in ArcPro as GEE can only ingest .tiff files. Additionally, GEE could only accept images under 10 GB; anything greater would require an additional Python API token. Rather than uploading the .tiff files for all of Larimer County with the Python API token, we decided to keep our process simple for the partners and only upload the 2021 imagery clipped to our study region. We uploaded the images as assets in GEE then imported the assets to our project code.

3.2.2 Generating Predictor Variables

Once we acquired all the necessary NAIP imagery, filtered the data to the correct years, and clipped to the study regions, we created spectral predictor variables using NAIP's red, green, blue, and near infrared (NIR) bands. We made variables out of each respective band, then used the red and NIR bands to calculate the Normalized Difference Vegetation Index (NDVI), a proxy for vegetation presence and vigor, and simple ratio. We also calculated the red:green ratio using the red and green NAIP bands. From the SRTM elevation dataset, we derived aspect, slope, northness, and eastness as additional topographic predictors and included them in our model. We tested model performance using both spectral and topographic predictors as well as spectral predictors alone. See Table 1 for a full list of predictor variables, predictor type, their corresponding sources, and the equations used to calculate them.

Table 1: Predictor Variables List

Predictor Variable	Predictor Type	Source	Equation	Reference
Slope	Topographic	Derived from the digital elevation model	Slope = ee.Terrain.slope(DEM)	N/A
Aspect	Topographic	Derived directly from the digital elevation model	Aspect = ee.Terrain.aspect(DEM).multiply(π/180)	N/A
Northness	Topographic	Derived from the aspect variable	Northness = aspect.cos()	N/A
Eastness	Topographic	Derived from the aspect variable	Eastness = aspect.sin()	N/A
NIR	Spectral	Derived directly from the NAIP imagery	NIR = NAIP_image.select('N')	Zeng, Y. *et al.*
Red	Spectral	Derived directly from the NAIP imagery	Red = NAIP_image.select('R')	N/A

Green	Spectral	Derived directly from the NAIP imagery	Green = NAIP_image.select('G')	N/A
Blue	Spectral	Derived directly from the NAIP imagery	Blue = NAIP_image.select('B')	N/A
NDVI	Spectral	Calculated using NIR and Red bands	$NDVI = \frac{(NIR - Red)}{(NIR + Red)}$	Kshetri, 2018
Simple Ratio	Spectral	Calculated using NIR and Red bands	$Simple\ Ratio = \frac{NIR}{Red}$	Hunt Jr., Daughtry, Eitel, Long, 2011
Red Green Ratio	Spectral	Calculated using the red and green bands	$Red\ Green\ Ratio = \frac{mean(Red)}{mean(Green)}$	Bannari, Morin, Bonn, Huete, 1995

3.2.3 Generating Different Image Resolutions

To test how different image resolutions affect model performance, we rescaled the 0.6m NAIP imagery to 1m, 2m, and 3m using the ImageProjection.atScale(x) function where x is an integer value denoting scale in meters. Then, we reprojected the 2013, 2019, and 2021 images to their new projections using the Image.reproject({crs, scale}) tool where crs is the coordinate reference system of the projection and scale is the scale in meters.

3.2.4 Generating Random Points and Building Appropriate Feature Classes

With all the predictor variables and image resolutions created, we generated 1,000 random points in the greater BDSR study region and 500 random points in the UMC study region using the ee.FeatureCollection.randomPoints() function in GEE. We created a 6m buffer around each point to avoid repeated sampling. Next, we created point feature collections titled canopy, gap, and shadow, and in each, we added a property called Class and assigned each feature collection a unique class value of 1-3, respectively. In order for the RF classifier to execute, all separate feature collections needed to be merged into one feature collection with different classes. We joined all feature collections together using the FeatureCollection1.merge(FeatureCollection2) function in GEE.

3.2.5 Building a Grid and Ocular Sampling

To classify all random points systematically, we created a grid around our study region using the StudyRegion.geometry().coveringGrid() tool. The grid helped us ensure that all random points received a classification. We moved through each grid cell and classified the randomly generated points into the canopy, gap, and shadow classes via ocular sampling. Our ocular sampling method involved classifying each random point using the GEE base satellite map and NAIP true-color and false-color imagery to help determine class. To calibrate our eyes as a team, we classified 50 training data points together. The remaining 950 data points were then classified by individual team members.

3.2.6 Random Forest Classifier and Training/Testing Data

Once we generated all classified points for both study regions, we divided the data into training and testing points, and fed them into the RF classifier. RF is a machine learning algorithm that classifies image pixels into designated classes using the training data and predictor variables defined by the user. We ran the RF classifier for 2013, 2019, and 2021 at 1m, 2m, and 3m resolutions varying the number of training points. For 2013 and 2021, we tested model performance using 100-800 training points, and for 2019, we tested model performance using 100-400 training points. The remaining 20% of our classified points were withheld from model creation and used as test data.

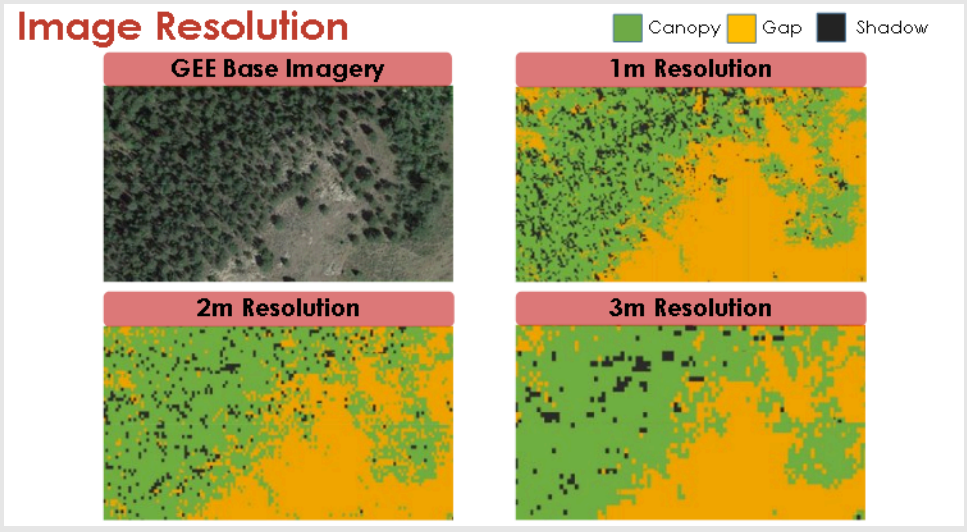


Figure 2: Examples classification model output derived from 2013 imagery at 1m, 2m, and 3m resolution. GEE base satellite imagery included for comparison.

3.2.7 Shadow Reclassification

One of the parameters we explored was the method of shadow reclassification. To accomplish this, we held the image resolution and number of training points constant and tested four different shadow classification methods for 2013, and five different shadow classification methods for 2019 and 2021. For 2013, we ran one model with no shadow reclassification, retaining canopy, gap, and shadow classifications. We ran another model for 2013 where all shadow pixels were reclassified as canopy to produce a two-class model output. We ran a third model where all shadow pixels were reclassified as gap, and we ran a final model where all shadow pixels were reclassified as either canopy or gap based on their NDVI value. For the final method of reclassifying shadow, we generated a histogram of NDVI values for shadow points, determined the mid-point in the bimodal distribution of the data, and utilized that value as our threshold for shadow reclassification. All shadow points with an NDVI value below that threshold were reclassified as gap and all shadow points above that value were reclassified as canopy. CFRI's existing methodology informed this approach. For 2019 and 2021, in addition to the four methods of shadow reclassification, we created a fifth classification method: manual reclassification. This consisted of manual reclassifying all shadow pixels as canopy or gap using visual cues in the imagery.

3.3 Data Analysis

Using the test data, we generated confusion matrices to evaluate the accuracy of the classified raster maps produced by the RF classifier. Once the confusion matrices were generated, we could use the test data accuracy metric to compare model performance at different numbers of training points, scale resolutions, shadow reclassification methods, predictor variables, and image qualities. Since the clarity, contrast, and image brightness increased as the imagery became more current, the 2013 image represents our lowest image quality and 2021 represents our highest image quality.

We then created a table that outlines accuracy metrics for each model iteration to systematically compare their performance. Accuracy metrics include training accuracy and test accuracy. We used training accuracy as a perfunctory check and test accuracy as the metric to indicate model performance. Additionally, we compared our 2013 model outputs to a CFRI LiDAR derived raster via percent overall agreement and class agreement. While neither model was compared to field validation data, comparing our 2013 model outputs to the LiDAR derived raster, we provided our partners with information about how our modeling approach differed from their current methodology.

4. Results & Discussion

4.1 Analysis of Results

4.1.1 Comparing Image Resolutions and Numbers of Training Points

We determined that image resolution had a minimal effect on the model test accuracy; differences between model performance at the 1m, 2m, and 3m resolutions were marginal. Increasing the number of training points increased the model test accuracy for both years at all resolutions. For 2013, the 3m resolution models performed slightly better than the 1m and 2m models, but at 800 training points, all model resolutions had a test accuracy of approximately 0.85 (Figure 3). For 2021, the 2m resolution model performed slightly better than the 1m and 3m models, but at 800 points, all model resolutions had a test accuracy of approximately 0.88 (Figure 3). For 2013 and 2021, increasing the number of training points improved model accuracy. Moreover, as the number of training points increased, model test accuracy continued to increase linearly without plateauing, suggesting that we had not yet reached the optimal number of training points.

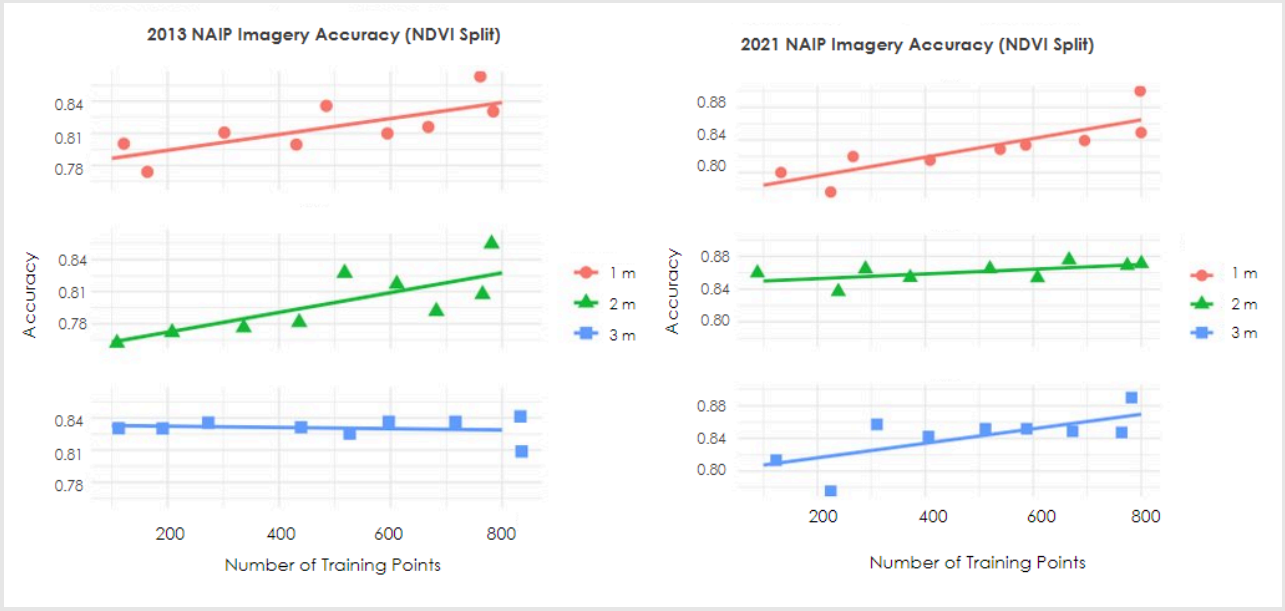


Figure 3. Model performance across scale and number of training points

4.1.2 Comparing Methods of Shadow Reclassification 2013, 2019, and 2021

For 2013, reclassifying shadow using an NDVI threshold and reclassifying all shadow pixels as canopy performed the best and performing no reclassification performed the worst (Figure 4). For 2021, using an NDVI threshold performed the best and performing no shadow reclassification performed the worst (Figure 4). For 2019, reclassifying all shadow pixels to canopy performed the best (0.83 test accuracy) while classifying all shadow pixels to gap and performing no shadow reclassification performed the worst (0.60 test accuracy) (Figure 4). There is over a 20-percentage point difference between the best and worst performing models, which suggests that shadow reclassification is more significant in forests with more rugged topography such as Upper Monument Creek. Note that performing no shadow reclassification retains three classes in the model output which inherently has a lower performance compared to two class model output.

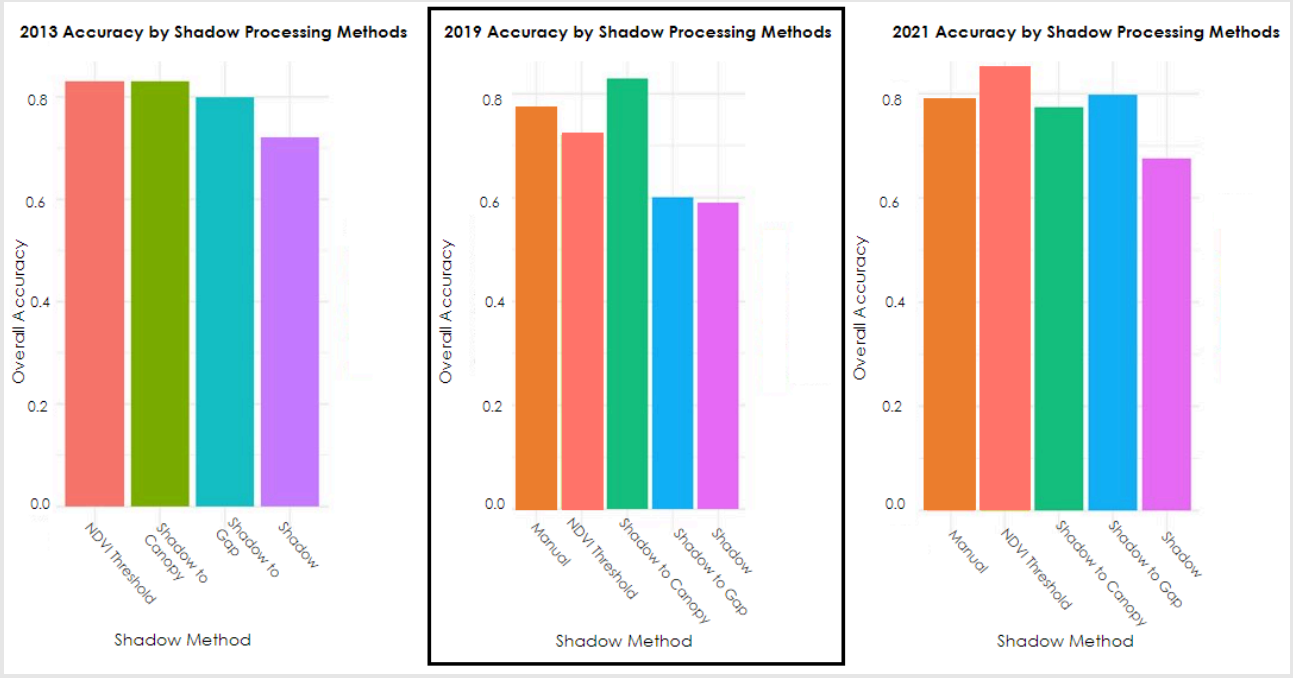


Figure 4. 2013, 2019, and 2021 model performance with varying shadow reclassification methods

4.1.3 Comparing Predictor Variables Used

For 2013 and 2021, there are minimal differences between the test accuracies of the spectral only models and the models run with spectral and topographic predictors (Figure 5). However, for 2019, there are larger differences between the test accuracies of the spectral only models and the models run with spectral and topographic predictors (Figure 5). The spectral-only models perform better which may suggest that topographic predictors may become confounding in regions with more variable topography, such as UMC.

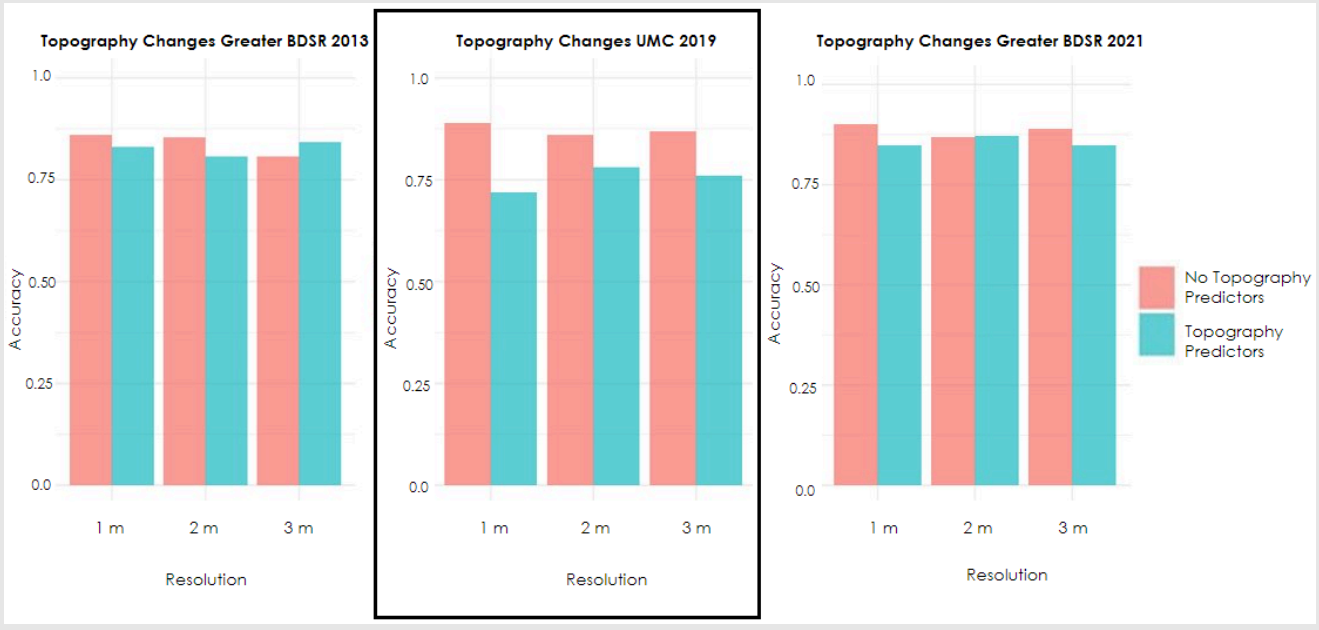


Figure 5. 2013, 2019, and 2021 model performance using different predictor variables

4.1.4 Comparing Model Outputs Between Image Qualities (2013 and 2021)

By comparing the 2021 and 2013 classification outputs for the greater BDSR region, we determined that image quality has minimal effect on model accuracy. 2021 NAIP imagery is higher quality, which means that canopy pixels can be more easily distinguished from gap pixels. This improved image quality led to less difficult decision making when manually classifying points as canopy, gap, or shadow, leading to higher quality training and test data. Comparing the 2013 and 2021 model outputs to each other, we determined that there is not a significant difference between 2013 and 2021 test accuracy; however, 2021 has a higher average test accuracy (0.82) and a higher maximum test accuracy (0.90). It should be noted that the range of model performance of 2021 (0.25) is greater than that of 2013 (0.14) (Figure 6).

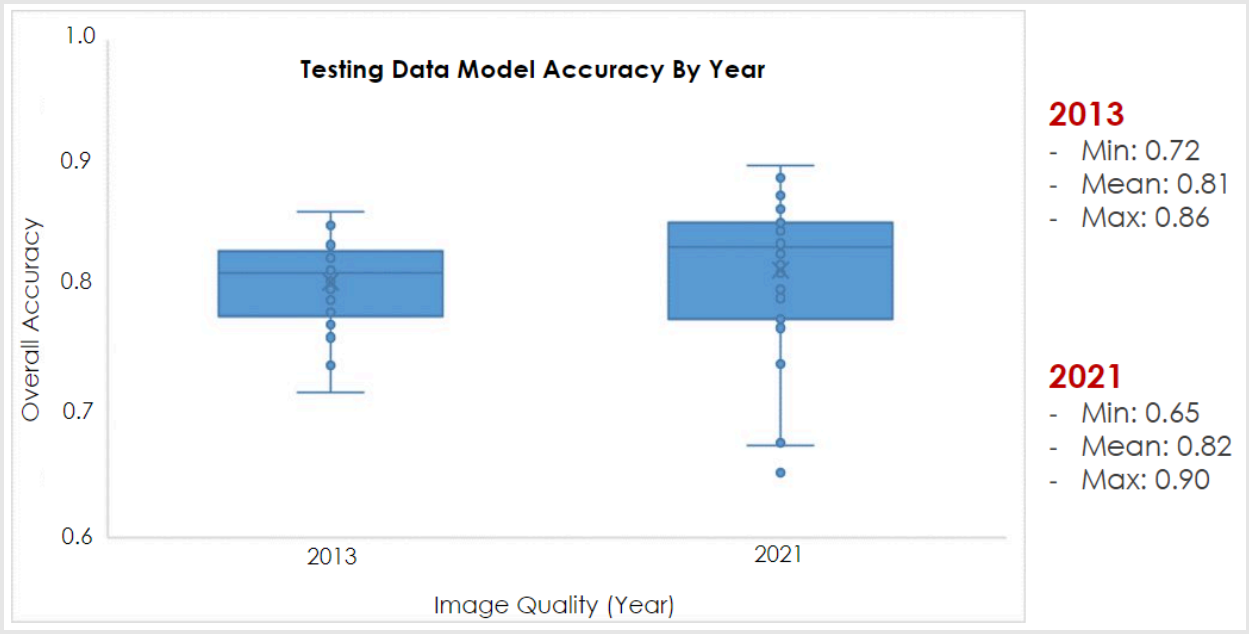


Figure 6. Model performance across different years

4.1.5 Best Performing Models

The model with the highest test accuracy for 2013 BDSR was the 1-meter resolution model classified with 800 training points, an NDVI threshold shadow reclassification, and only spectral predictors. The overall test accuracy was 0.863 (Figure 7a). The model with the highest test accuracy for 2021 BDSR was the 1-meter resolution model classified with 800 training points, an NDVI threshold shadow reclassification, and only spectral predictors. The overall test accuracy was 0.900 (Figure 7b). The model with the highest test accuracy for 2019 UMC was the 2-meter resolution model classified with 400 training points, a shadow reclassification from shadow to canopy, and both spectral and topographic predictors. The overall test accuracy was 0.790 (Figure 7c). See Appendix A – Appendix C for all 2013, 2019, and 2021 model iterations and results.

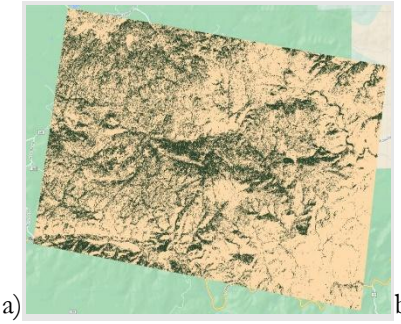
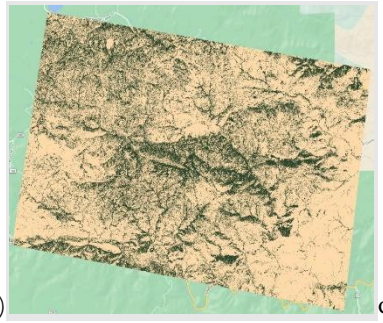
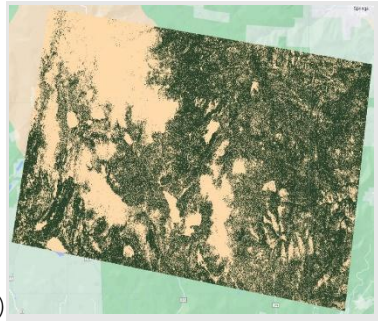
a)  b)  c) 

Figure 7. Highest performing models by year and study region a) 2013 BSDR b) 2021 BDSR c) 2019 UMC

4.1.6 Comparison to CFRI LiDAR Data:
The 2013 model with the best CFRI LiDAR agreement is the 2013 1m model classified using 800 training points where all shadow pixels were reclassified to canopy and both spectral and topographic predictors were used (Figure 8). It had a 68% overall agreement with the LiDAR classifications and an overall test accuracy of 0.72. Figure 9 shows the comparison between our 2013 BSDR model and CFRI's LiDAR data raster map for the same area. The cream color represents gap agreement between models, the green represents canopy agreement between models, and the black represents disagreement between models (Figure 9).

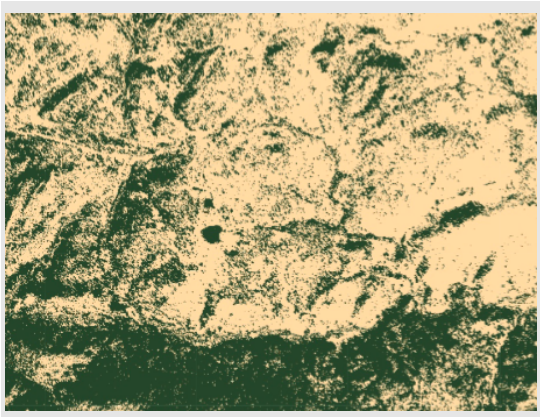


Figure 8. 2013, 1m, 800 training points NDVI Shadow Reclassification Model Output

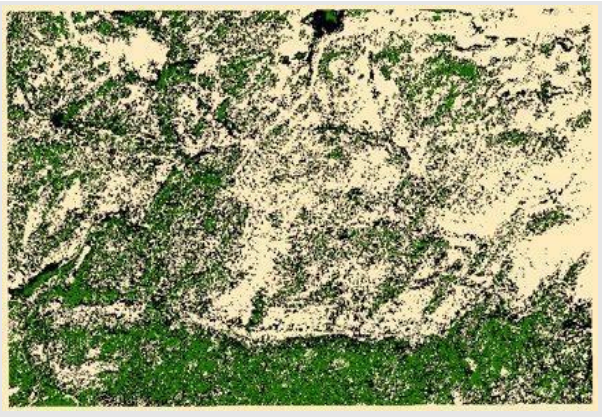


Figure 9. CFRI and 2013 Model Agreement/Disagreement

4.1.7 Errors and Uncertainties
While our methodology produced meaningful results that we could draw conclusions from, we still encountered errors and uncertainties. One of our most prominent sources of error was our assumption that the NAIP imagery used was consistent from year to year and site to site. NAIP mosaics are created from aerial images whose sensors vary in acquisition times, regions surveyed, and viewing geometries. This affects their ability to make consistent and useful classification products.

Another prominent error in our analysis was our treatment of NAIP bands as reflectance. NAIP images do not have units. They are simply digital numbers (DNs) that may not be proportional to reflectance and can vary from image to image. Our analysis, however, assumed that these bands were consistent and proportional to reflectance. This means that the indices applied on different dates are different from each other and that none of them are likely to match literature established performance of the indices.

Finally, there were several sources of error and uncertainty pertaining to our classifications, which were done manually. First, the 2013 NAIP imagery was poorer quality as defined by the lower contrast, brightness, and clarity compared to the 2021 image. The lower image quality led to difficulty when manually classifying points as canopy, gap, or shadow. Because the 2013 image had lower contrast, it was occasionally challenging to distinguish between classes. We examined the true and false-color NAIP imagery as well as satellite base imagery in GEE to aid in point classification, and while this process was effective, the satellite base imagery had a limited zoom extent which led to some uncertainty classifying at the pixel level. There is additional uncertainty in our classifications because we classified our training data manually without field validation data. While some points had an obvious classification, other points were far less obvious and required a best guess from our team when classifying. This means that some points could have been misclassified which would lead to inaccuracies in the model.

4.2 Future Work

With additional time, we would have continued our analysis in several ways. First, we would explore object-based classifications in which an algorithm groups areas of an image by likeness. Object-based classifications could lead to more accurate model classifications compared to the pixel-based classification our current methodology uses. We would also test for collinearity and perform a regression analysis on our predictors to determine the optimal combination of variables for classification. Next, we would explore reclassifying shadows as a post-processing technique rather than a pre-processing technique. Our current methodology explored shadow reclassification prior to our images being classified, so we would be interested to see how shadow reclassification performs after a three-class model output is produced. We would also further explore the areas where our model outputs disagree with CFRI's LiDAR classification outputs. For example, we noticed that disagreement is more common in riparian areas. We would also compare our 2019 and 2021 model outputs to LiDAR data as an additional accuracy metric. Finally, we would test how our model performs at a watershed level to see if the random forest classifier can still produce accurate classifications at a larger scale.

5. Conclusions

This project analyzed the impact that image quality, image resolution, number of training points, methods of shadow reclassification, and predictor variables have on model accuracy. We determined that image quality had a minimal effect on model performance, with 2021 models performing slightly better. We found minimal differences between the test accuracies produced by 1m, 2m, and 3m models, suggesting that image resolution had minimal impact on model test accuracy. For the number of training points, we detected a positive correlation between model test accuracy and number of training points used for model classification. This suggested that higher numbers of training points correspond to a more accurate model. However, we did not detect a plateau in model performance as training points increased, suggesting that more training points could be added to improve model performance. For shadow reclassification, we determined that any method of shadow reclassification is better than no reclassification at all, since all tested methods of shadow reclassification performed better than the model where no reclassification was done. Finally, we determined that our models performed the best when only spectral variables were used as opposed to both spectral and topographic predictors. By exploring how each of these parameters affected model performance, our partners will be able to create effective classification rasters. This will save our partners time and effort in the early stages of model classification and will allow them to quickly advance into more complex analyses.

6. Acknowledgments

The Front Range Wildland Fires team extends our gratitude to all the individuals who made our project possible. We thank our node fellow Sarah Hettema for supporting our project from beginning to end. We thank our brilliant Science Advisors Nick Young, Dr. Tony Vorster, Christopher Tsz Hin Choi, Dr. Paul Evangelista, and Dr. Catherine Jarnevich for their scientific expertise and their continuous guidance throughout the term. We thank our devoted project partners Jackie Edinger, Stephanie Mueller, Amanda Fordham, and Ethan Bucholtz. Finally, we thank Bob Sturtevant, the land manager of the Ben Delatour Scout Ranch region, who took us on a tour of our study site and answered our many questions. Our team was grateful for the work each of you contributed to the project!

This material contains modified Shuttle Radar Topography Mission data (2008) and modified National Agricultural Imagery Program data (2013, 2019, 2021).

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

NAIP - National Agricultural Imagery Program

DEM - Digital Elevation Model

SRTM - Shuttle Radar Topography Mission

GEE - Google Earth Engine

RF - Random Forest Classifier

8. References

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1–2), 95–120. https://doi.org/10.1080/02757259509532298

Birth, G., and G. McVey. "Measuring the Color of Growing Turf with a Reflectance Spectrophotometer." *Agronomy Journal* 60 (1968): 640-643.

Cannon, J. B., Barrett, K. J., Gannon, B. M., Addington, R. N., Battaglia, M. A., Fornwalt, P. J., Aplet, G. H., Cheng, A. S., Underhill, J. L., Briggs, J. S., & Brown, P. M. (2018). Collaborative restoration effects on forest structure in ponderosa pine-dominated forests of Colorado. *Forest Ecology and Management*, *424*, 191–204. https://doi.org/10.1016/j.foreco.2018.04.026

Cannon, J. B., Warnick, K. J., Elliott, S., & Briggs, J. S. (2021). Low- and moderate-severity fire offers key insights for landscape restoration in ponderosa pine forests. *Ecological Applications, 32*(2), e2490. https://doi.org/10.1002/eap.2490

Colorado Forest Restoration Institute (CFRI). (2023). *LiDAR based classification model of Ben Delator Boy Scout Ranch*. [Model]. CFRI

Dickinson, Y.L., & Pelz, K.A. (2012). Proposed methods for monitoring "groupy-clumpy" forest cover characteristics at the stand scale. 2012_FR-CFLR_SpatialHeterogeneityMethod_2012Nov_FINAL.pdf (colostate.edu)

Gamon, J., and J. Surfus. "Assessing Leaf Pigment Content and Activity With a Reflectometer." *New Phytologist* 143 (1999): 105-117.

Hunt Jr., E. Raymond; Daughtry, C. S. T.; Eitel, Jan U. H.; Long, D. S. (2011). Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index

Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara. 2008. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database: https://srtm.csi.cgiar.org.

Kshetri, Tek. (2018). NDVI, NDBI & NDWI Calculation Using Landsat 7, 8.

Larson, A. & Churchill, D. (2012). Tree spatial patterns in fire-frequent forests of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. *Forest Ecology and Management, 267,* 74–92. https://doi.org/10.1016/j.foreco.2011.11.038

Prasai, R., Schwertner, T. W., Mainali, K., Mathewson, H., Kafley, H., Thapa, S., Adhikari, D., Medley, P., & Drake, J. (2021). Application of Google Earth Engine Python API and NAIP imagery for land use and land cover classification: A case study in Florida, USA. *Ecological Informatics*, *66*, 101474. https://doi.org/10.1016/j.ecoinf.2021.101474

Rouse, J., R. Haas, J. Schell, and D. Deering. *Monitoring Vegetation Systems in the Great Plains with ERTS*. Third ERTS Symposium, NASA (1973): 309-317.

U.S. Department of Agriculture Farm Service Agency. (2013). *National Agriculture Imagery Program* (NAIP). [Data set]. Google Earth Engine.

U.S. Department of Agriculture Farm Service Agency. (2019) *National Agriculture Imagery Program* (NAIP). [Data set]. Google Earth Engine.

U.S. Department of Agriculture Farm Service Agency. (2021). *National Agriculture Imagery Program* (NAIP). [Data set]. US Department of Agriculture Geospatial Data Gateway.

Zeng, Y. *et al.* (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nature Reviews Earth and Environment.* https://doi.org/10.1038/s43017-022-00298-5

9. Appendices

Appendix A

Table A1. *All 2013 model iterations and accuracies*

Site	Year	Resolution	Number of Training Points	Number of Training Classes	Post Processing Type	Training Accuracy	Test Accuracy	Overall CFRI LiDAR Product Agreement	Canopy CFRI LiDAR Product Agreement	Gap CFRI LiDAR Product Agreement
BSR	2013	1m	800	3	No reclass	0.998	0.716	0.679	0.196	0.483
BSR	2013	1m	800	2	Shadow to Gap	0.995	0.795	0.664	0.172	0.491
BSR	2013	1m	800	2	Shadow to Canopy	1	0.830	0.669	0.235	0.433
BSR	2013	1m	800	2	NDVI threshold	0.993	0.862	0.663	0.194	0.468
BSR	2013	1m	800	2	NDVI threshold	0.995	0.830	0.665	0.192	0.472
BSR	2013	1m	700	2	NDVI threshold	0.995	0.815	0.664	0.191	0.472
BSR	2013	1m	600	2	NDVI threshold	1	0.810	0.663	0.187	0.476
BSR	2013	1m	500	2	NDVI threshold	0.998	0.835	0.666	0.191	0.474
BSR	2013	1m	400	2	NDVI threshold	1	0.8	0.665	0.190	0.474
BSR	2013	1m	300	2	NDVI threshold	0.996	0.810	0.666	0.196	0.470
BSR	2013	1m	200	2	NDVI threshold	0.994	0.774	0.668	0.192	0.476
BSR	2013	1m	100	2	NDVI threshold	1	0.8	0.665	0.174	0.490
BSR	2013	2m	800	3	No reclass	0.997	0.740	0.674	0.211	0.463
BSR	2013	2m	800	2	Shadow to Gap	0.996	0.851	0.661	0.161	0.500
BSR	2013	2m	800	2	Shadow to Canopy	0.998	0.765	0.671	0.238	0.432
BSR	2013	2m	800	2	NDVI threshold	0.996	0.854	0.664	0.208	0.455
BSR	2013	2m	800	2	NDVI threshold	0.997	0.807	0.666	0.200	0.465
BSR	2013	2m	700	2	NDVI threshold	0.997	0.791	0.665	0.200	0.465

BSR	2013	2m	600	2	NDVI threshold	0.998	0.817	0.668	0.193	0.474
BSR	2013	2m	500	2	NDVI threshold	0.997	0.827	0.667	0.181	0.485
BSR	2013	2m	400	2	NDVI threshold	1	0.781	0.668	0.191	0.476
BSR	2013	2m	300	2	NDVI threshold	1	0.776	0.664	0.190	0.474
BSR	2013	2m	200	2	NDVI threshold	1	0.771	0.659	0.199	0.460
BSR	2013	2m	100	2	NDVI threshold	1	0.761	0.661	0.209	0.452
BSR	2013	3m	800	3	No reclass	0.991	0.738	0.678	0.198	0.479
BSR	2013	3m	800	2	Shadow to Gap	0.995	0.744	0.664	0.177	0.487
BSR	2013	3m	800	2	Shadow to Canopy	0.996	0.829	0.670	0.240	0.429
BSR	2013	3m	800	2	NDVI threshold	0.992	0.808	0.664	0.197	0.467
BSR	2013	3m	800	2	NDVI threshold	0.996	0.841	0.667	0.193	0.473
BSR	2013	3m	700	2	NDVI threshold	0.997	0.835	0.666	0.194	0.472
BSR	2013	3m	600	2	NDVI threshold	0.995	0.835	0.669	0.193	0.475
BSR	2013	3m	500	2	NDVI threshold	0.997	0.825	0.668	0.201	0.467
BSR	2013	3m	400	2	NDVI threshold	0.997	0.830	0.666	0.194	0.472
BSR	2013	3m	300	2	NDVI threshold	1	0.835	0.668	0.186	0.481
BSR	2013	3m	200	2	NDVI threshold	1	0.830	0.665	0.173	0.492
BSR	2013	3m	100	2	NDVI threshold	1	0.830	0.670	0.164	0.505

Appendix B

Table B1: *All 2021 model iterations and accuracies*

Site	Year	Resolution	Number of Training Points	Number of Training Classes	Post Processing Type	Training Accuracy	Test Accuracy
BSR	2021	1m	800	3	No reclass	0.990	0.677
BSR	2021	1m	800	2	Shadow to Gap	0.996	0.798
BSR	2021	1m	800	2	Shadow to Canopy	0.999	0.773
BSR	2021	1m	800	2	Manual Reclassification	0.996	0.781
BSR	2021	1m	800	2	NDVI threshold	0.991	0.900
BSR	2021	1m	800	2	NDVI threshold	0.997	0.849
BSR	2021	1m	700	2	NDVI threshold	0.999	0.839
BSR	2021	1m	600	2	NDVI threshold	0.996	0.834
BSR	2021	1m	500	2	NDVI threshold	0.998	0.829
BSR	2021	1m	400	2	NDVI threshold	0.998	0.815
BSR	2021	1m	300	2	NDVI threshold	1.000	0.820
BSR	2021	1m	200	2	NDVI threshold	1.000	0.776
BSR	2021	1m	100	2	NDVI threshold	1.000	0.800
BSR	2021	2m	800	3	No reclass	1.000	0.653
BSR	2021	2m	800	2	Shadow to Gap	0.996	0.775
BSR	2021	2m	800	2	Shadow to Canopy	0.998	0.769
BSR	2021	2m	800	2	Manual Reclassification	0.998	0.740
BSR	2021	2m	800	2	NDVI threshold	0.994	0.869
BSR	2021	2m	800	2	NDVI threshold	0.996	0.871
BSR	2021	2m	700	2	NDVI threshold	0.999	0.876
BSR	2021	2m	600	2	NDVI threshold	1.000	0.854
BSR	2021	2m	500	2	NDVI threshold	0.997	0.865
BSR	2021	2m	400	2	NDVI threshold	1.000	0.854
BSR	2021	2m	300	2	NDVI threshold	1.000	0.865
BSR	2021	2m	200	2	NDVI threshold	0.990	0.837
BSR	2021	2m	100	2	NDVI threshold	1.000	0.860
BSR	2021	3m	800	3	No reclass	1.000	0.675
BSR	2021	3m	800	2	Shadow to Gap	0.999	0.773

BSR	2021	3m	800	2	Shadow to Canopy	0.999	0.793
BSR	2021	3m	800	2	Manual Reclassification	0.998	0.816
BSR	2021	3m	800	2	NDVI threshold	0.995	0.890
BSR	2021	3m	800	2	NDVI threshold	1.000	0.847
BSR	2021	3m	700	2	NDVI threshold	0.997	0.849
BSR	2021	3m	600	2	NDVI threshold	0.996	0.852
BSR	2021	3m	500	2	NDVI threshold	0.994	0.852
BSR	2021	3m	400	2	NDVI threshold	0.998	0.842
BSR	2021	3m	300	2	NDVI threshold	0.994	0.857
BSR	2021	3m	200	2	NDVI threshold	0.995	0.775
BSR	2021	3m	100	2	NDVI threshold	1.000	0.813

Appendix C

Table C1. *All 2019 (Upper Monument Creek) model iterations and accuracies*

Site	Year	Resolution	Number of Training Points	Number of Training Classes	Post Processing Type	Training Accuracy	Test Accuracy
UMC	2019	1m	400	3	No reclass	0.995	0.586
UMC	2019	1m	400	2	Shadow to Gap	1	0.604
UMC	2019	1m	400	2	Shadow to Canopy	1	0.83
UMC	2019	1m	400	2	Manual Reclassification	0.990	0.766
UMC	2019	2m	400	3	No reclass	0.997	0.602
UMC	2019	2m	400	2	Shadow to Gap	0.998	0.663
UMC	2019	2m	400	2	Shadow to Canopy	1	0.79
UMC	2019	2m	400	2	Manual Reclassification	1	0.789
UMC	2019	3m	400	3	No reclass	0.998	0.646
UMC	2019	3m	400	2	Shadow to Gap	0.992	0.575
UMC	2019	3m	400	2	Shadow to Canopy	0.995	0.865
UMC	2019	3m	400	2	Manual Reclassification	0.989	0.909
UMC	2019	3m	400	2	NDVI threshold	1	0.759
UMC	2019	2m	400	2	NDVI threshold	0.992	0.776
UMC	2019	1m	400	2	NDVI threshold	0.995	0.723
UMC	2019	3m	400	2	NDVI threshold	0.995	0.873
UMC	2019	2m	400	2	NDVI threshold	0.997	0.86
UMC	2019	1m	400	2	NDVI threshold	0.99	0.894
UMC	2019	2m	300	2	NDVI threshold	0.996	0.757
UMC	2019	2m	200	2	NDVI threshold	1	0.776
UMC	2019	2m	100	2	NDVI threshold	1	0.786