


2022

Using Landsat-Based Phenology Metrics, Terrain Variables, and Machine Learning for Mapping and Probabilistic Prediction of Forest Community Types in West Virginia

Faith M. Hartley
West Virginia University, fmh00001@mix.wvu.edu

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

 Part of the Biodiversity Commons, Data Science Commons, Forest Management Commons, Natural Resources Management and Policy Commons, Other Earth Sciences Commons, and the Other Forestry and Forest Sciences Commons

Recommended Citation

Hartley, Faith M., "Using Landsat-Based Phenology Metrics, Terrain Variables, and Machine Learning for Mapping and Probabilistic Prediction of Forest Community Types in West Virginia" (2022). *Graduate Theses, Dissertations, and Problem Reports*. 11375.
<https://researchrepository.wvu.edu/etd/11375>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Using Landsat-Based Phenology Metrics, Terrain Variables, and Machine Learning for Mapping and Probabilistic Prediction of Forest Community Types in West Virginia

Faith Hartley

Thesis submitted to the Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements for the degree of
Master of Arts in Geography

Aaron Maxwell, Ph.D., Chair

Rick Landenberger, Ph.D.

Zachary Bortolot, Ph.D.

Department of Geology and Geography

**Morgantown, West Virginia
2022**

Keywords: forest type classification, forests, machine learning, West Virginia, community type, Phenology, digital terrain analysis
Copyright 2022 Faith Hartley

ABSTRACT

Using Landsat-Based Phenology Metrics, Terrain Variables, and Machine Learning for Mapping and Probabilistic Prediction of Forest Community Types in West Virginia

Faith Hartley

This study investigates the mapping of forest community types for the entire state of West Virginia, USA using Global Land Analysis and Discovery (GLAD) Phenology Metrics analysis ready data (ARD) derived from the Landsat time series and digital terrain variables derived from a digital terrain model (DTM). Both classifications and probabilistic predictions were made using random forest (RF) machine learning (ML) and training data derived from ground plots provided by the West Virginia Natural Heritage Program (WVNHP). The primary goal of this study is to explore the use of globally consistent ARD data for operational forest type mapping over a large spatial extent. Mean overall accuracy calculated from 50 model replicates for differentiating seven forest community types using only variables selected from the 348 GLAD Phenology Metrics used in the study resulted in an overall accuracy (OA) of 53.36% (map-level image classification efficacy (MICE) = 0.42). Accuracy increased to a mean OA of 73.0% (MICE = 0.62) when the Oak/Hickory and Oak/Pine classes were combined to an Oak Dominant class. Once selected terrain variables were added to the model, the mean OA for differentiating the seven forest types increased to 61.58% (MICE = 0.52). Our results highlight the benefits of combining spectral data and terrain variables and also the enhancement of the product's usefulness when probabilistic prediction are provided alongside a hard classification. The GLAD Phenology Metrics did not provide an accuracy comparable to those obtained using harmonic regression coefficients; however, they generally outperformed models trained using only summer or fall seasonal medians and performed comparably to spring medians. We suggest further exploration of the GLAD Phenology Metrics as input for other spatial predictive mapping and modeling tasks.

Acknowledgements

I would like to thank the West Virginia Division of Natural Resources (WVDNR) and the associated Natural Heritage Program for providing the field plot data used in this study. I would also like to thank the Global Land Analysis & Discovery program at the University of Maryland for providing phenology data and associated processing script. Graduate Teaching and Research Assistant funding and support was provided by the West Virginia University Department of Geology and Geography and AmericaView/West Virginia View. WV View is supported by AmericaView and the U.S. Geological Survey under Grant/Cooperative Agreement No. G18AP00077. I would also like to thank my advisor, Dr. Aaron Maxwell for all his guidance and support on this project along with the rest of my Thesis Committee, Dr. Rick Landenberger and Dr. Zachary Bortolot, for their time and effort. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Table of Contents

Title Page	i
Abstract	ii
Acknowledgments	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Chapter 1: Introduction and Research Questions	1
1.1 Introduction	1
1.2 Research Questions	2
Chapter 2: Background	2
2.1 Remote Sensing and Forest Type Mapping	2
2.2 Machine Learning	3
2.3 Input Data Considerations	5
2.4 Results from Prior Studies	6
Chapter 3: Study Area	8
Chapter 4: Data	8
Chapter 5: Methods	11
Chapter 6: Results	15
Chapter 7: Discussion	26

Chapter 8: Conclusion29

Chapter 9: References..... 29

List of Tables

Table 1: Subset of forest community type mapping projects	7
Table 2. Phenological variables used in analysis from GLAD	10
Table 3. Topographic variables used in analysis	11
Table 4. Final West Virginia forest type count from the WVDNR	11
Table 5: Key packages used for analysis in RStudio	13
Table 6: Assessment metrics used for assessment with their description	15
Table 7. Full set of variables used in the models	10
Table 8. Overall model accuracy rates	17
Table 9. Additional model accuracy rates	18
Table 10: Confusion Matrix from kNN model using all predictor variables	22
Table 11. Confusion Matrix from kNN model using only terrain predictor variables	22
Table 12. Confusion Matrix from kNN model using only spectral predictor variables	23
Table 13. Confusion Matrix from RF model using all predictor variables	23
Table 14: Confusion Matrix from RF model using only terrain predictor variables	24
Table 15. Confusion Matrix from RF model using only spectral predictor variables	24
Table 16. Confusion Matrix from SVM model using all predictor variables	25
Table 17. Confusion Matrix from SVM model using only terrain predictor variables	25
Table 18. Confusion Matrix from SVM model using only spectral predictor variables	26
Table 19. Confusion Matrix from RF model with the Oak/Hickory and Oak/Pine classes combined to an Oak/Hickory/Pine class	27

List of Figures

Figure 1: Map of point locations of forest type observations from WVDNR	12
Figure 2: List of top 25 important variables estimated using conditional permutation importance ...	17
Figure 3: Probability raster grids for each of the forest classes in West Virginia	20
Figure 4: Final map of classification results based on RF using spectral and terrain variables	21

Chapter 1. Introduction

Forests serve as an important source of land-based carbon and as biological resource reservoirs (Liu et al. 2018). Dominant species composition, which defines forest type, impacts forests' ecological characteristics and workings, making understanding this key to quantifying a forest communities' function (Costanza et al. 2018). Key functions of forest communities include providing habitat for a variety of organisms, acting as a carbon pool for the global carbon cycle, regulating the flow of water through the hydrological cycle, and providing a number of ecosystem goods and services for human and societal uses (Liu et al. 2018, Lu et al. 2017). Forest management practices have matured over centuries as humans and society have adapted to live within, make use of, and manage the natural environment. This can be seen historically through indigenous community practices and more recently in state-centric management such as prescribed fires and forest thinning. Shifting from traditional forest management focused on increasing timber production, sustainability-based approaches have arisen to meet the challenges of protecting and sustaining forests and the communities they support (Mery et al. 2005).

Sustainable forest management began in the 1960s and evolved to address deforestation and biodiversity loss resulting from widescale timber production. This transition was further spurred by media coverage of deforestation and global poverty (Mery et al. 2005). A crucial aspect of sustainable forest management is mapping forest community types, their abundances, and their distribution within a given region, as forest diversity composition impacts habitat suitability, best management practices, ecosystem services, and hydrologic and biogeochemical processes (Liu et al. 2018; Pasquarella et al. 2018). Specifically, with the rise in the impacts of global climate change, forests are increasingly important due to their role in climate change abatement (Liu et al. 2018). Therefore, thematic mapping products that simply differentiate forests from other land cover, or that only differentiate deciduous, mixed, and evergreen forests (i.e., the National Land Cover Database (NLCD) in the United States (Jin et al. 2013)), are inadequate to meet modeling and management needs to support sustainable development and climate change adaption. Also desirable is the ability to update these maps with ease using repeat collection remotely sensed images and other data sources since distributions are likely to adjust in response to climate change and landscape disturbances. Forest community mapping is also important for many professions such as wildlife habitat modeling, policymaking and governance, and natural resource management (Immitzer et al. 2010; Liu et al. 2018; Pasquarella et al. 2018).

Forest tree species composition depends on local biotic and abiotic conditions while also varying over time and space (Evans et al. 2011). For example, each species has a different response to environmental condition variation (e.g., temperature, precipitation, seasonality, and humidity), which can result in increased distribution and abundance of environmentally sensitive plants (Evans et al. 2011). With climate change, changing precipitation and temperature patterns will lead to shifts in the distribution of species. Unfortunately, these relationships are not always well understood or easily modeled to estimate forest stand tree species composition and associated forest community types. Further, forest communities are generally not easily

differentiated into a limited number of discrete, well-defined types, resulting in “fuzzy” class definitions, and boundaries between classes are often gradational. Traditional field survey methods are expensive and time consuming. Different research studies have sought to overcome these challenges through remote sensing techniques, which provide a methodical, synoptic view of the earth at differing time intervals and spatial resolutions (Adams et al., 2020; Negendra 2001; Pasquarella et al. 2018). Remote sensing is beneficial for forest community type classification as it can provide a high level of detail for large areas in a relatively short amount of time (Negrenda, 2001).

This study investigates the use of random forest (RF), support vector machine (SVM), and k -Nearest neighbor (k NN) machine learning (ML) for classifying and estimating the probability of forest type occurrence over a broad spatial extent (i.e., every forested pixel in an entire state) at a moderate spatial resolution and at the pixel-level. Predictions were made using a combination of digital terrain variables derived from digital terrain models (DTMs) along with phenological and seasonal variability metrics derived from the Landsat multispectral time series and made available via the Global Land Analysis and Discovery (GLAD) Phenology dataset (Potapov et al. 2020). This will be accomplished by exploring the following objectives:

- Document the importance of a variety of digital terrain and spectral variables for predicting and differentiating forest community types in different landscapes.
- Produce classification and probability of occurrence maps for every forested pixel within the entire extent of West Virginia.
- Assess the accuracy of classification and probabilistic forest community type models using withheld validation data and a variety of assessment metrics.

Ultimately, this study explores the mapping of forest community types and associated probabilities in an operational context using available spectral and terrain predictor variables and an existing field plot-based forest inventory as training data.

Chapter 2. Background

2.1. Remote Sensing and Forest Type Mapping

With improved satellite imaging in regards to spatial, spectral, and temporal resolutions and wider availability of open archives, classification methods have increasingly incorporated remotely sensed data, such as surface reflectance measured over varying wavelengths in the visible, near infrared (NIR), and shortwave infrared (SWIR) spectral ranges. Different plant species may reflect specific ranges of electromagnetic (EM) radiation differently (Gates et al. 1965). Also, satellite-based multispectral sensors allow for wall-to-wall mapping of spectral reflectance, ignoring the influence of data gaps and clouds, which allows models to be applied over entire spatial extents to generate complete map output; therefore, spectral reflectance data are often incorporated in classification projects (Negrenda, 2001). Multi-temporal data

classifications relying on a time series of repeated measurements have been shown to improve forest type mapping accuracy as seasonal patterns can aid in differentiation of communities that have similar spectral signatures in one season but different signature in others (Liu et al. 2018). Difficulties in mapping species composition and community types include that spectral reflectance can change with the seasons and species can have overlapping ground distributions (Adams et al., 2020). Further, a variety of methods have been developed to summarize time series multispectral data, many of which have not been consistently or completely applied to global data archives (Potapov et al. 202). Thus, we argue that there is a need to investigate input predictor variables that represent consistent, global products, such as the GLAD Phenology metrics explored here.

Pasquerella et al. (2018) is an example study that highlights the value of time series data and ancillary data for forest type mapping. Within their study, which differentiated between eight forest classes including Hardwood Swamp, Softwood Swamp, Northern Hardwood, Central Hardwoods, White Pine, Hemlock/White Pine, Spruce-Fir, and Pitch Pine/Scrub Oak, they documented that single-date late autumn images significantly outperformed other single date classifications (accuracy of $74.4 \pm 1.3\%$), that multi-date imagery consistently outperformed single-date imagery (highest overall accuracy of $78.5 \pm 1.2\%$), and that spectral-temporal variables derived from a time series of available Landsat multispectral observations performed the best (mean accuracy of $80.50 \pm 2.23\%$) based on a pixel-based assessment. Classification improved further when ancillary datasets, including wetland probability estimates and topographic variables, were included alongside the spectral-temporal features ($83.39 \pm 2.31\%$ overall accuracy) (Pasquarella et al. 2018).

2.2. Machine Learning

Within remote sensing and land classification, ML has matured to become an important tool for extracting actionable information from large, complex datasets (Maxwell et al. 2018, 2020). ML serves as a framework for identifying important variables, building accurate predictions, and exploring complex relationships and spatial patterns within a model (Evans et al. 2011). These algorithms tend to produce higher overall classification accuracies than traditional parametric classification methods (e.g., Gaussian maximum likelihood) which is attributed to their non-parametric nature and ability to model complex patterns and relationships within a complex feature space (i.e., many variables that may be correlated, of different data types, measured on different scales, and not normally distributed) (Maxwell et al. 2016, 2018, 2020). As an operational example, the NLCD uses machine learning methods to map general land cover and land cover change for the entire United States on a regular basis (Jin et al. 2013; Maxwell et al. 2018).

This study specifically makes use of the SVM, k NN, and RF algorithms. The SVM algorithm attempt to define an optimal hyperplane to separate classes. This hyperplane is the plane that provides the maximum margin or separation between classes and is defined by a

subset of the available training samples, termed support vectors. In order to allow for improved separation of classes that are not linearly separable, the training samples can be projected into a higher dimensional feature spaces using the “kernel trick” where the optimal hyperplane may be more linear. Additionally, the cost parameter is used to control the complexity of the generated hyperplane by controlling the penalty associated with misclassifying a sample, practically allowing for adjusting the model to compensate for overfitting or underfitting. Methods have also allowed for the original SVM algorithm, which only allows for the differentiation of two classes, to be expanded to allow for the differentiation of three or more classes (James et al. 2013, Kuhn & Johnson 2013). SVM classifiers have been documented to be effective at classifying hyperspectral data (Li 2021, Sheykhmousa et al. 2020, Lu et al. 2017). Specifically, this algorithm is useful due to its high memory efficiency, ability to learn from small training sets, and its generally high accuracy rates for classification problems (Li 2021). Previous studies also indicate that this method is most effective when hyperparameters are optimized for generalization and to avoid overfitting (Li 2021). These benefits have made SVM popular as a classifier within the field of remote sensing. The main limitation found in previous studies is that this method can be sensitive to overfitting if hyperparameters are not properly tuned (Sheykhmousa et al. 2020).

k NN is a nonparametric method that has been successfully used for classification problems. This method is conceptually simple; a new sample is classified based on the nearest training samples within the multidimensional feature space defined by the input predictor variables. The nearest samples are determined based on some measure of distance in the feature space, such as simple Euclidean distance. The number of neighbors that are considered is controlled by the k parameter, which is commonly selected using a tuning process. The new sample is then assigned to the majority class of the selected neighbors (James et al. 2013, Kuhn & Johnson 2013).

RF is an ensemble decision tree method. Decision trees use recursive binary partitioning to split the data into more homogeneous subsets and generate rulesets to perform classification. In order to determine decision rules, the Gini Impurity metric is used, which is a measure of how often a random element from the dataset is incorrectly labeled. Each tree in the ensemble uses a subset of the training samples, which are selected using bootstrapping, or random sampling with replacement. Also, only a subset of the predictor variables is available for splitting at each decision node. The goal of using a subset of the training data and variables is to reduce the correlation between trees and minimize overfitting. Or, a set of weak classifiers are collectively strong and generalize well due to reduced overfitting (Brieman, 2001). RF has been applied to a variety of problems including classifying ecological zones (e.g., Immitzer et al. 2012), identifying landslide probability areas (e.g., Maxwell et al. 2020), and creating forest dead fuel load estimates (e.g., D’Este et al. 2021). Further, previous studies indicate that RF is robust to combining data from multiple sources, including incorporating data measured on varying scales and nominal data with varying numbers of categories (Liu et al. 2018). For example, studies have

combined topographic and spectral reflectance data to improve classification performance (e.g., Immitzer et al. 2012; Liu et al. 2018; Maxwell et al. 2020).

Although RF is not commonly used for mapping or modeling over large spatial extents, several studies have used it for such tasks in the context of slope failure prediction and forest classification (e.g., Liu et al 2018; Maxwell et al. 2016, 2020; Pasquarella et al. 2018). Liu et al. 2018 used an RF classification approach to differentiate eight forest types using a combination of Sentinel-1A synthetic aperture radar (SAR), Sentinel-2A MSI multispectral imagery, digital elevation models (DTMs) and associated derivatives, including aspect and slope, and multi-temporal, multispectral Landsat-8 images. They used a total of 43 input features including 39 spectral, three topographic, and one backscatter variable. They documented a very high overall accuracy rate (99.3%), and that topographic slope was the most important feature for forest type classification within their model (Liu et al. 2018). Additionally, they documented that mixed forest types (e.g., mixed broad-leaved forest) had greater confusion or error when compared against single-species forest types (e.g., Chinese red pine) (Liu et al. 2018). Alternatively, Lu et al. (2017) used an SVM algorithm with a combination of multispectral and hyperspectral data for forest classification in the Jiangxi province of China. They found that the highest accuracies were obtained when using a spatial-spectral-temporal fusion framework.

2.3. Input Data Considerations

Studies indicate that Landsat imagery is useful for monitoring landcover and forest change due to its ability to capture the spectral and temporal variability in ground reflectance (Pasquarella et al., 2018). However, Landsat imagery availability and timing can make large-area forest type mapping challenging, especially during heavy cloud cover, as clouds and cloud shadows impact analyses (Pasquarella et al. 2018). Landsat ETM+ and OLI images are collected once every 8 to 16 days and are available globally as Level 1, geometrically corrected data, and as Level 2, surface reflectance products (Potapov et al., 2020). Forest and resource mapping have been a goal of the Landsat program since its development, with the opening of the Landsat archive in 2008 making this goal more attainable for researchers (Pasquarella et al. 2018).

As forest community spatial distribution can be closely linked with topographic characteristics, including topographic derivatives with multispectral imagery has been shown to improve forest classification accuracies by 10-27% (Liu et al. 2018). Despite Liu et al.'s 2018 high overall accuracy, their Landsat-only classification accuracy rate was only 50%, significantly lower than the combined Landsat 8, Sentinel-2A, and DTM-derivatives classification result (82.78%). Previous studies have provided evidence that multi-source imagery is beneficial for mapping forest community types (e.g., Gao et al. 2015; Liu et al. 2018; Lu et al. 2017; Melville et al. 2018; Nink et al. 2019). For example, Lu et al. (2017) quantified a Landsat 8-based forest type classification overall accuracy rate using single date imagery at 69.95% but obtained an almost 14% improvement in accuracy by using a spatial-spectral-temporal integrated data fusion method (Table 1). Despite the utility of multispectral data fusion for forest type classification,

mapping projects, such as the NLCD in the United States, have had success using one multispectral imagery source when differentiating a small number of community types (i.e., deciduous, mixed, and evergreen) (Pasquerella et al. 2018).

2.4. Results from Prior Studies

Table 1 provides an overview of some forest community type mapping projects conducted since 2009. The table illustrates how variable overall accuracy rates for forest type mapping are with values as low as 50% and as high as 96%. All of the included studies tested several feature spaces and/or methods, which resulted in a range of reported accuracies. For example, Nink et al. (2019) tested spectral vegetation indices between two different states in Germany in a cross-border forest type classification of broadleaf or coniferous forests. Alternatively, Pasquerella et al. (2018) examined differences in spectral-temporal Landsat forest type mapping versus conventional single and multi-date classification methods. Similarly, Adams et al. (2020) used RF classification with Landsat imagery; however, this study used Landsat 8 Operational Land Imager (OLI) images with topographic variables including elevation, slope, aspect, topographic wetness index, and topographic position index for forest community type mapping. Immitzer et al. (2012) differed from other studies in the table in their use of high-resolution commercial imagery, specifically Maxar's WorldView-2 satellite data, and documented very high user accuracies for the four main tree species (95.9%) but saw lower accuracy for the classification of all ten tree species (82.4 %). Additionally, their study showed higher accuracy rates for object-based RF classification as opposed to pixel-based classification (Immitzer et al. 2012). Kim et al. (2009) similarly used an object-based image analysis approach; however, their study utilized a single Ikonos image.

I argue that the disparity in reported accuracies can be attributed to variability in (1) the number of and definition of classes differentiated, (2) landscapes characteristics and abiotic gradients, (3) spatial extent, (4) quantity and quality of training data, (5) input predictor variables include, and (6) classification methods used. Thus, there is a need to specifically investigate the mapping of forest community types across large spatial extents using consistently available predictor variables and training data in the context of operational forest type mapping. This study differs from those in Table 1 in that the goal is to produce probabilistic map products alongside “hard” classifications. Given the commonly reported low classification accuracies when a large number of forest community types are differentiated, or more than just deciduous, evergreen, and mixed forests are mapped, I argue that there is a need to investigate probabilistic mapping as a means to generate products that may be more accurate and useful.

ML algorithms have proven to be beneficial for obtaining probabilistic map outputs for other mapping tasks (e.g., Evans & Cushman 2009; Maxwell et al. 2020). For example, Wright and Gallant (2007) noted the value of RF-based probabilistic output as opposed to hard classifications for wetland mapping in Yellowstone National Park in the western United States. This is important as it has been a challenge for the remote sensing community, in the face of

rapid climate change, to provide frequent and complete observations for forest and vegetation monitoring across large areas (Gao et al. 2015), and seeking alternatives to “hard” classifications that may be too inaccurate or unreliable is necessary to meet this goal. By producing probability surfaces rather than discrete patches of community types, vegetation representation errors including community type definitions, stand delineations, and omission/commission errors are reduced (Evans & Cushman 2009). This study hopes to expand upon previous work by incorporating forest community type field observations, topographic derivatives, Landsat-derived phenology and seasonal metrics, and ML to investigate variable importance and community probability of occurrence values for every forested pixel within the state of West Virginia. A community occurrence probability map will be helpful for planning fieldwork, monitoring and modeling the effects of climate change, managing and protecting forests, and understanding ecosystem relationships and processes.

Table 1. Subset of forest community type mapping projects.

Study	Multi-Spectral Data	Study Area(s)	Classification Method(s)	Number of Classes	Reported Accuracy
Kim et al. (2009)	Ikonos	Guilford Courthouse National Military Park, Greensboro, North Carolina	Segmentation or Object Based Image Analysis	3	60-83%
Immitzer et al. (2012)	WorldView-2	Burgenland, Austria	Random Forest Linear Discrimination Analysis	10	65.7-95.9%
Lu et al. (2017)	HJ-1A CCD HJ-1A HSI MOD09GA Landsat-8	Gan River nature reserve, Jiangxi, China	SVM	7	69.95-83.6%
Liu et al. (2018)	Sentinel-2A Sentinel-1A Landsat-8	Dabie Mountains, Wuhan, China	Random Forest	8	50-82.78%
Pasquarella et al. (2018)	Landsat-4 Landsat-5 Landsat-7	Massachusetts, USA	Random Forest	8	62.4-85.7%

Nink et al. (2019)	Landsat-8 Sentinel-2A Sentinel-2B	Saarland and Luxembourg, Germany	Automated bootstrap	2	79-96%
Adams et al. 2020)	Landsat-8 OLI	Southern Unglaciaded Allegheny Plateau, Ohio, USA	Random Forest	7	70.3-74.9%

Chapter 3. Study Area

The United States Department of Agriculture (USDA) classified 78% of West Virginia’s land area as forest, or over 12 million acres (Widmann 2007). Forests are a critical part of the state’s natural resources; however, forest community types are especially difficult to distinguish due to local topographic variability and the mountain forest’s diverse nature (Widmann et al. 2007). According to Strausbaugh and Core in the *Flora of West Virginia* text, the dominant types of forest vegetation vary between the three physiographic provinces within the state (Widmann et al. 2007). Furthest west, the Western Hill Section, which is the largest province in the state, is characterized by Oak-Pine, Oak-Chestnut, Floodplain, and Cove Hardwoods or Mixed Mesophytic forests. This area hosts the Allegheny Plateau, which is situated between the Ohio River to the west and the more mountainous region to the east and is dissected by a dendritic stream network. The second division is the Allegheny Mountain section which includes a northeast to southwest oriented mountain range dominated by Northern Evergreen, including Red Spruce stands, and Northern Hardwood community types. This area contains the highest elevations in the state and is bound to the east by the Allegheny Front. The final, lowest elevation, and most eastern division is the Ridge and Valley province which is generally the driest area in the state and hosts Oak-Hickory-Pine community types. The landscape is dominated by structurally controlled linear ridges and valleys and a trellis drainage network. This area originally had a large chestnut population that declined in the 1920s and 30s with the chestnut blight caused by a fungus (Elkins 2012).

Chapter 4. Data

The Landsat-based spectral data for this study comes from the University of Maryland’s Global Land Analysis and Discovery (GLAD) Laboratory. Specifically, I used GLAD’s Phenological Metrics, which are based on Landsat surface reflectance measurements and represent a set of globally consistent metrics to characterize land surface phenology (Potapov et al. 2020). These variables are provided to the public as Landsat analysis-ready data (ARD) (Potapov et al. 2020). GLAD selects Landsat images from 1997 to the present and then categorizes them into tiers of excellence, with Tier 1 data meeting “the highest geometric and radiometric standards.” This tier is the only one used for ARD processing. Images including seasonal snow cover are also excluded (Potapov et al. 2020). Some limitations of this dataset

include that it is unsuitable for wintertime image processing, surface water extent mapping, and precise analysis of land surface reflectance. These limitations did not impact the intent of this study.

Specifically, the Landsat ARD phenological metrics Type A dataset was used as the spectral predictor variables in this study. The variables are summarized in Table 2 below. To generate these metrics, reflectance data are normalized by taking every possible ratio of bands (e.g., $BR = (\text{Blue} - \text{Red}) / (\text{Blue} + \text{Red})$) and then multiplying by 10,000 and adding 10,000.5 to ensure positive values and the ability to store the data in a 16-bit integer format. Reflectance normalizing is important because it improves spectral consistency and addresses factors impacting surface reflectance measurements, such as local relief and topography (Potapov et al. 2020). The processing chain includes generating 16-day composites using best-available data, filling gaps resulting from cloud cover with best available data from the prior three years, and finally generating the metrics listed in Table 2 from the timeseries of 16-day composites. A total of 348 phenological variables from this dataset were used in this study.

Statewide DTM raster data were obtained from the National Elevation Dataset (NED) at a 1/3-arc second (approximately 10 m) spatial resolution. From these elevation data, a variety of metrics were generated to characterize the local terrain, as summarized in Table 3. Other than elevation (Elev) and in order to characterize local relief and rugosity, calculated variables include slope (Slp), mean slope within a moving window (SlpMn), topographic position index (TPI), topographic roughness index (TRI), surface area ratio (SAR), and surface relief ratio (SRR). Variables relating to moisture content and local incoming radiation that were used include linear aspect (LnAsp), topographic radiation aspect index (TRASP), heat load index (HLI), and site exposure index (SEI). A total of 11 terrain variables were used in the analysis. These metrics were selected due to potential correlation with forest community type distributions, since they characterize different aspects of the terrain surface, and because they can all be derived using only a DTM. For variables that rely on defining a local moving window, a circular window with a radius of 3 cells, or 30 meters, was used. All terrain derivatives were rescaled to a 30 m spatial resolution using pixel aggregation and the mean value from the original cells within the new, larger cell to match the scale of the Landsat-derived metrics.

Species composition data came from the West Virginia Division of Natural Resources' (WVDNR) Natural Heritage Program. The WVDNR's Natural Heritage Program is responsible for collecting and maintaining locally based ecological community characteristics. Natural Heritage Program scientists determine the species composition found in a plot using in-field observations. A spreadsheet of information gathered directly from West Virginia's forests was provided for this project. To spatially visualize the data, X, Y data were converted to spatial point features in ArcGIS and displayed by community type.

Table 2. Phenological variables used in analysis from GLAD. Q1 indicates the 1st quartile while Q3 indicates the 3rd quartile.

Spectral Bands	Blue Green Red NIR SWIR1 SWIR2
Derived Indices	(NIR-Red)/(NIR+Red) [RN] (NIR-SWIR1)/(NIR+SWIR1) [NS1] (Blue-Green)/(Blue+Green) [BG] (Blue-Red)/(Blue+Red) [BR] (Blue-NIR)/(Blue+NIR) [BN] (Green-Red)/(Green+Red) [GR] (Green-NIR)/(Green+NIR) [GN] (SWIR1-SWIR2)/(SWIR1+SWIR2) [SWSW] Spectral variability index (SVVI)
Statistics	Minimum [min] Maximum [max] Second lowest value [smin] Second highest value [smax] Median [median] Average between smin and Q2 [av50smin] Average between Q2 and smax [av50smax] Average between min and Q1 [avmin25] Average between Q3 and max [av75max] Average between Q1 and Q3 [av2575] Average of all values [avminmax] Average of all values except min and max [avminsmax]
Amplitude	max - min smax- smin av50smax - av50smin av75max - av25min

Table 3. Topographic variables used in analysis.

Variable	Abbreviation	Description/Equation
Linear Aspect	Asp	$270 - \frac{360}{2\pi} \times \arctan^2\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$
Elevation	Elev	Bare-ground surface height
Topographic Dissection Index	TDI	$\frac{z - z_{min}}{z_{max} - z_{min}}$
Heat Load Index	HLI	Index for annual direct incoming solar radiation based on latitude, slope, and aspect
Topographic Roughness Index	Rph3	$\sigma^2(z)$
Surface Area Ratio	SAR	$\frac{\text{Cell Size}^2}{\text{Cos}(\text{Slope in Degrees})}$
Site Exposure Index	SEI	$\text{Slp} \times \cos\left(\pi \frac{\text{Asp} - 180}{180}\right)$
Slope (Degrees)	Slp	$\arctan\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right) \left(\frac{180}{\pi}\right)$
Mean Slope	Slpmn	Calculates slope within a moving window
Slope Position	TPI	$z - z_{mean}$
Surface Relief Ratio	SRR	$\frac{z_{mean} - z_{min}}{z_{max} - z_{min}}$
Topographic Radiation Aspect Index	TRASP	$\frac{1 - \cos\left(\left(\frac{\pi}{180}\right) \times (\text{Asp} - 30)\right)}{2}$

5. Methods

To reduce the number of categories of different community types and to limit the analysis to classes with a sufficient number of samples, Hill County deciduous forest and successional classes were omitted. Additionally, Calcareous Forests and Woodlands were combined with Oak/Hickory and Dry/Mesic Oak Forests while Dry Rocky Pine/Oak Forests and Woodlands were combined with Oak and Pine Forests. The number of available samples by class are summarized in Table 4. A total of 2,249 samples were used in the study.

Table 4. Final West Virginia forest type count from the WVDNR.

Community Type	Count
Oak/Hickory and Dry/Mesic Oak Forests	684
Floodplain Forests and Swamp	454
Oak and Pine Forests	417
Mixed Mesophytic Forests	261
Hemlock Forests	166
Northern Hardwoods Forests	159
Red Spruce Forests	108
Total Count	2,249

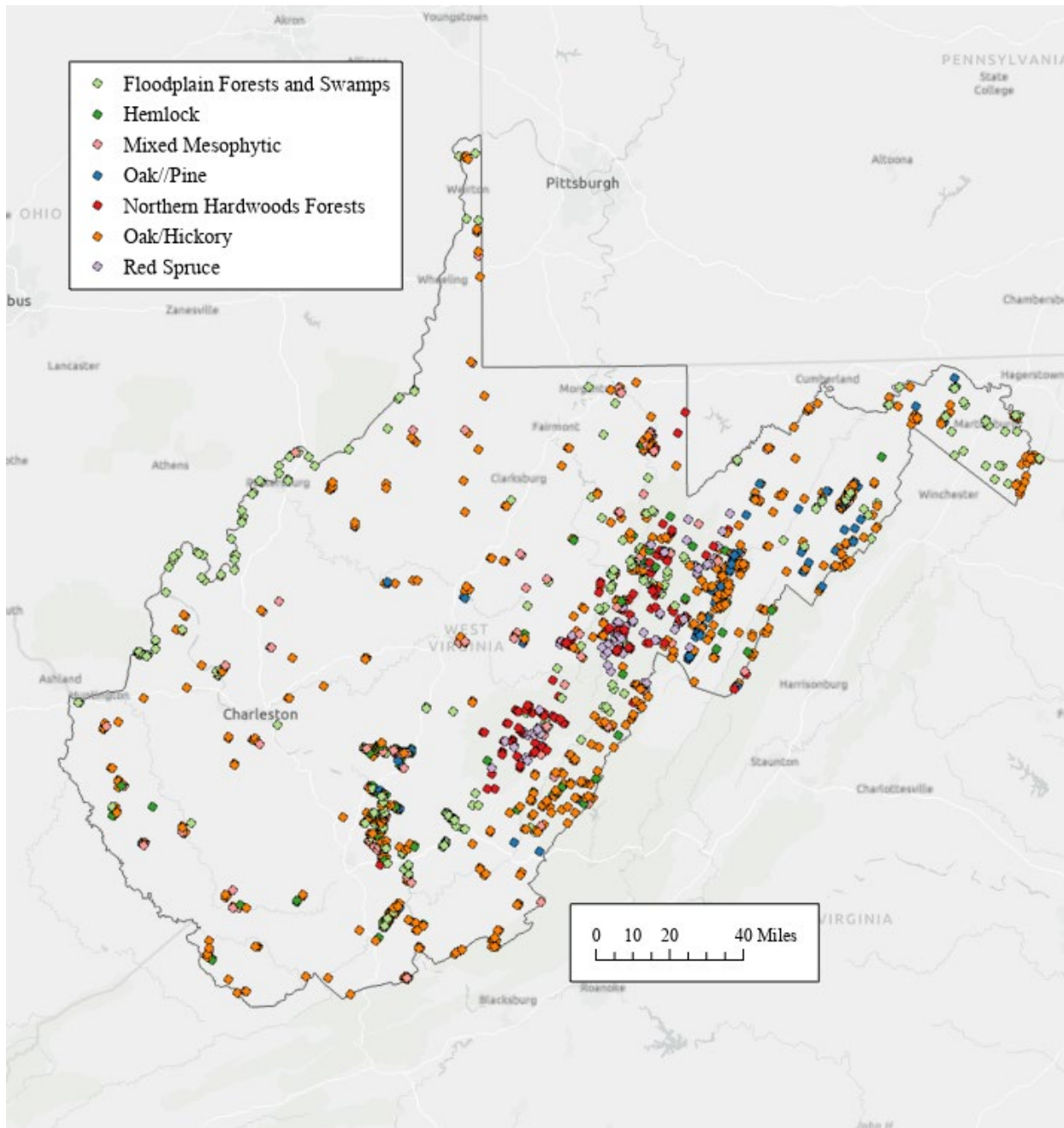


Figure 1. Map of point locations of forest type observations from WVDNR.

Data analysis and modeling were conducted using the R data science environment and language and the RStudio Integrated Development Environment (IDE). R was chosen because of its ease of use for exploratory work and due to the large number of available data analysis packages. The caret package was used to train and implement the three different algorithms. This package calls to randomForest for RF implementation, the class package for k NN, and the kernlab package for SVM, which are summarized in Table 5. RF requires the user to define the number of trees in the ensemble and the number of predictor variables that can be sampled for a

node candidate (mtry). The number of trees used in the RF model was set at 100, which was large enough to produce stable results. It has been shown that it is best to use a large number of trees and that increasing the number of trees in the ensemble will not result in overfitting (Brieman 2001). The mtry hyperparameter was optimized using five-fold cross-validation and a grid search over a set of candidate values. For k NN, the k parameter, or the number of nearest neighbors in the model, was tuned. For SVM, a radial basis kernel was used, and the cost parameter, or decision boundary, was tuned. In order to reduce the number of input predictor variables, a recursive feature elimination process was employed which also takes into account model parsimony. This algorithm was implemented using the rfUtilities package and the rf.modelSel() function, which uses the RF algorithm (Murphy et al. 2010). Within the model, the dependent variable is the forest type classification, and the predictor variables are all spectral and topographic variables, as described above.

Based on these results, a subset of variables was selected to use as a final feature space. A final model was trained using just these variables, all associated raster grids were stacked into a multiband dataset, and hard classifications and probabilistic predictions were made at all cells within the extent of West Virginia. This will result in a 30 m spatial resolution hard classification and probabilistic predictions of each mapped forest type for the entire state. Areas that were not forested were masked out using the NLCD 2019 land cover product. A summary of the main packages used in this study is provided in Table 5.

Table 5. Key packages used for analysis in RStudio.

Package	Citation	Description
randomForest	Liaw and M. Wiener (2002)	Used for classification and prediction of data
caret	Kuhn (2020)	Used for training and plotting classification and regression models
terra/raster	Hijmans (2020)	Used for display, analysis, and manipulation of raster data
dplyr	Hadley et al. (2021)	Used for data manipulation
pROC	Robin et al. (2011)	Used to display, analyze, smooth, and compare ROC curves
class	Venables and Ripley (2002)	Used for implementing k-NN
kernlab	Karatzoglou et al. (2004)	Used for implementing SVM
multiROC	Wei and Wang (2018)	Used to calculate the multi-class AUC ROC and multi-class AUC PR
rfUtilities	Murphy et al. (2010)	Used for feature selection

The RF algorithm has been used for variable importance estimation. Traditionally, this was undertaken by randomly permutating the predictor variable of interest while maintaining the other predictor variables, effectively removing any correlation between the predictor variable and the dependent variable. Greater decreases in predicting the out-of-bag (OOB) data with this

random permutation would indicate that this variable is of importance (Breiman, 2001). This method yields a more marginal assessment of importance in which importance within the entire feature set is not considered. It has been shown that this measure tends to underestimate the importance of a predictor variable when it is correlated with other predictor variables in the model (Debeer and Strobl, 2020; Strobl et al., 2009, 2008, 2007). In order to alleviate this bias, Strobl et al. (2007; 2008; 2009) suggested an augmented RF-based variable importance assessment method based on conditional inference trees. Using this method, it is possible to alleviate the impact of correlated variables and assess both marginal and conditional/partial variable importance, in which the correlation between variables is considered in the variable importance estimation to assess the added value of including the variable within the model (Debeer and Strobl, 2020; Strobl et al., 2009, 2008, 2007). In this study, I used an implementation of this method provided by the permimp R package (Debeer and Strobl, 2020). I specifically assessed marginal importance, or importance not considering the other predictor variables in the model.

To assess the model's accuracy, 1/3 of the data were omitted from the training process and reserved for testing (Maxwell et al. 2016). The omitted data were selected via a stratified random sample, grouped by community type, and extracted without replacement. Trained models were used to predict the withheld validation data and generate confusion matrices. From these confusion matrices, I calculated overall accuracy (OA) (i.e., the percent of the validation samples that were correctly classified) and class-level user's (UA) (1 – commission error) and producer's (1 – omission error) accuracy (PA) (Congalton and Green, 2019; Stehman, 1997; Stehman et al., 2009; Stehman and Czaplewski, 1998). It should be noted that the relative proportion of classes in the confusion matrix represent those from the Natural Heritage Database, which may not align with the true proportions on the landscape. Thus, confusion matrices represent a population confusion matrix relative to the Natural Heritage Database but may not represent a true population confusion matrix for the mapped landscape (Stehman, 2014, 2013, 2012).

Since the commonly used Kappa statistic has come under scrutiny and its use in remote sensing is now discouraged (Foody, 2020; Pontius and Millones, 2011), we did not calculate this metric. Instead, we calculated map-level image classification efficacy (MICE), which only uses the reference class margin totals as opposed to both the reference and classification margin totals to correct OA for chance agreement. This method has been suggested to be robust to accuracy inflation due to chance agreement and meaningful when the number of samples per class are imbalanced while not suffering from the logical flaws of Kappa (Shao et al., 2021). Lastly, we also calculated the percent of the validation samples in which the correct classification was within the top three predicted class probabilities as a means to assess how often the correct class is included amongst the top three classes in terms of predicted probability.

All of the metrics discussed above are based on classification results in which the class with the highest predicted probability is compared to the reference class. However, since I also wanted to assess the probabilistic products, measures that take into account probabilities at varying decision thresholds were also calculated. These metrics include the area under the receiver operating characteristic curve (AUC ROC) and area under the precision-recall curve

(AUC PR). The ROC curve takes into account only class producers' accuracies (1 – omission error) at varying decision thresholds and can be overly optimistic when classes are imbalanced. In contrast, the precision-recall curve considers the producer's and user's (1 – commission error) accuracy relative to the positive case and can be especially informative when classes are imbalanced since user's accuracy is impacted by class imbalance. Also, since this classification task was not a binary classification problem, multi-class versions of the metrics were used. Micro-average AUCs were calculated by stacking all groups together. (Cortes and Mohri, n.d.; Fan et al., 2006; Saito and Rehmsmeier, 2015; Wandishin and Mullen, 2009).

Table 6. Assessment metrics used for assessment with their description.

Assessment Metrics	Explanation
Confusion (Error) matrix	Spatial overlay of the classification results and testing data, source of other assessment metrics such as the overall accuracy and UA/PA
User's accuracy (UA)	1 - Commission Error
Producer's accuracy (PA)	1 - Omission Error
Area Under the Receiver Operating Characteristic Curve (AUC ROC)	Area under ROC curve, which takes into account multi-class sensitivity and specificity and varying decision thresholds
Area Under the Precision-Recall Curve (AUC PR)	Area under the P-R curve, which takes into account multi-class recall and precision at varying decision thresholds
Map-level Image Classification Efficacy (MICE)	Chance adjusted accuracy using reference class margin totals for accuracy correction

Chapter 6. Results

Table 7 lists the full set of 36 variables that were used in the Spectral + Terrain models and were selected using the RF-based model selection routine that incorporates parsimony after Murphy et al. (2010) whereas Figure 2 highlights the 25 most important predictor variables as estimated using the conditional permutation importance measure after Debeer and Strobl (2020). All the terrain variables were included along with several blue and green band measurements (e.g., BG_anminmax). Additionally, a few near-infrared and short-wave infrared variables were of high relative importance within the model (e.g., NS1_max). There were also a high number of included variables that combined the two different shortwave-infrared bands, including multiple measures of yearly central tendency (e.g., SWSW_smin).

Table 8 presents the overall accuracy (OA) rates for the different models. The RF model using the combined spectral and terrain predictor variables performed the best with an accuracy of 61.58%. The *k*NN and SVM models showed very similar accuracy rates for the three different feature spaces; however, the combined predictor variables generally showed an improvement in

model performance across all three algorithms. The k NN and RF models performed better with just the terrain predictors; however, the SVM model performed better using only the spectral predictors.

Further accuracy measures are noted in Table 9. The MICE values fluctuated with the highest measure being for the RF All (i.e., Spectral + Terrain) model at 0.52, which, although low due to the complexity of this task, does indicate that the result contains some useful content. A model that is not better than random would yield a MICE value close to 0 (Shao et al., 2021). The SVM Ter (i.e., only Terrain variables) and the SVM Spec (i.e., only Spectral variables) had the lowest values overall. These metrics were converted from percentages to proportions for table consistency. The aFS is the average F1-Score for the model which measures the harmonic mean of the class-level precision and recall metrics and indicates that the RF algorithm consistently outperformed the other algorithms. The estimated confidence intervals for OA are provided in the final two columns of Table 9 (OAU and OAL) and generally follow the same overall accuracy patterns with the RF All model performing the highest overall. Generally, non-overlapping confidence intervals for OA suggest that the models are truly different or that differences did not arise from sampling variability or random chance.

The micro-AUC ROC and AUC PR generally suggest stronger performance than the threshold-based metrics, which supports the value of the probabilistic outputs. The Top3 metric is the percent of samples in which the correct class was one of the three highest predicted class probabilities out of all seven classes. These values were relatively high, suggesting that even if the correct class was not predicted, it was commonly in the top set of the predicted probabilities. This further highlights the utility of the various probabilistic outputs.

Table 7. Full set of variables used in the models.

BG_av2575	GN_max	SVVI_min	SWSW_smin
BG_avminmax	GR_max	SVVI_smin	dem30m
BG_avsminsmax	NS1_av50smax	swir2_smin_SVVI	diss3
BG_median	NS1_av75smax	SWSW_avminmax	hli2
BR_av50smax	NS1_max	SWSW_avsmin25	rph3
BR_av75smax	NS1_smax	SWSW_avsmin50	sar
BR_smax	SVVI_avsmin25	SWSW_avsminsmax	sei
GN_av75smax	SVVI_avsmin50	SWSW_median	slp
slpmn3	sp3	ssr3	trasp

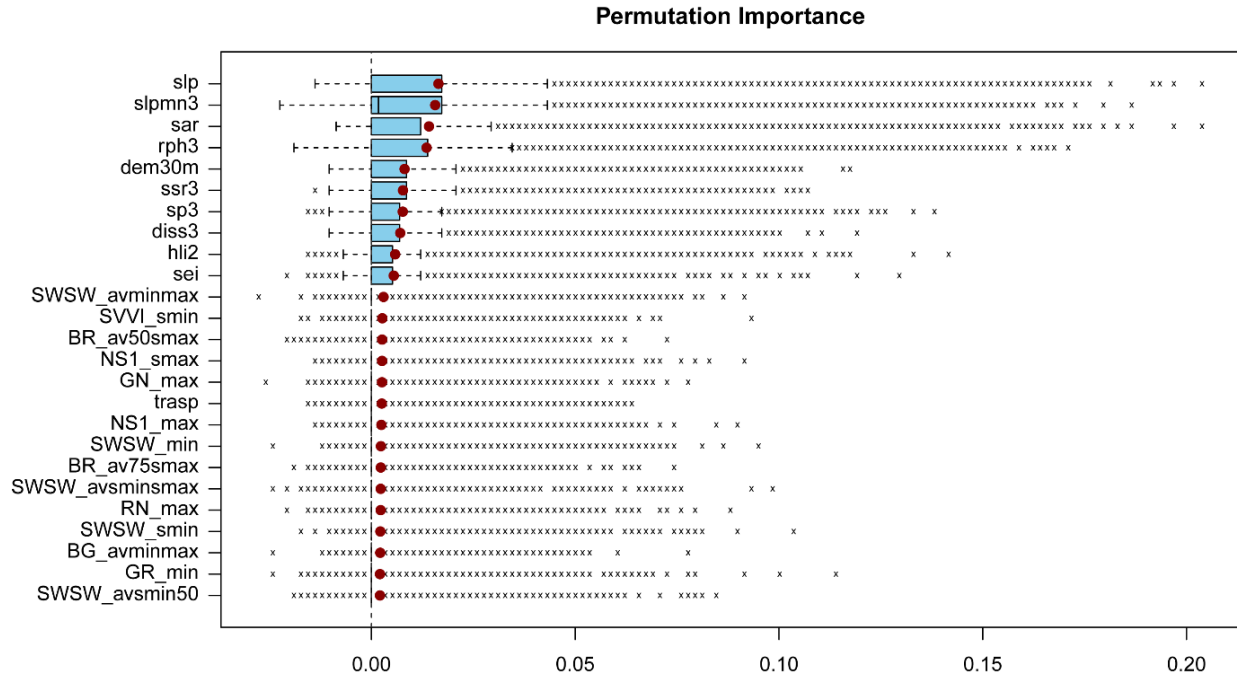


Figure 2. List of top 25 important variables as estimated using conditional permutation importance.

Table 8. Overall model accuracy (OA) rates.

<i>k</i> NN	<i>k</i> NN	<i>k</i> NN	RF All	RF	RF	SVM	SVM	SVM
All	Terrain	Spectral		Terrain	Spectral	All	Terrain	Spectral
53.81	48.88	45.59	61.58	56.35	53.36	52.17	41.55	42.75

Table 9. Additional model accuracy rates.

Run	MICE	aUA	aPA	aFS	Micro AUC ROC	Micro AUC PR	Top3	OAU	OAL
RF All	0.52	0.61	0.58	59.0	0.92	0.70	0.92	0.57	0.65
SVM All	0.40	0.45	0.40	41.4	0.84	0.57	0.83	0.48	0.56
kNN All	0.42	0.56	0.50	51.0	0.86	0.58	0.87	0.49	0.57
RF Spec	0.42	0.51	0.52	50.6	0.86	0.54	0.87	0.49	0.57
SVM Spec	0.28	0.37	0.37	33.4	0.80	0.44	0.79	0.38	0.46
kNN Spec	0.32	0.49	0.43	44.2	0.81	0.46	0.80	0.41	0.49
RF Ter	0.45	0.52	0.49	50.0	0.89	0.62	0.91	0.52	0.60
SVM Ter	0.27	0.41	0.36	33.8	0.80	0.49	0.77	0.38	0.45
kNN Ter	0.36	0.52	0.44	45.5	0.85	0.53	0.87	0.45	0.53

The resulting probabilities for the best performing model, RF All, for each of the community types are provided in Figure 3 while Figure 4 shows the resulting hard classification from the same model (i.e., each forested pixel was assigned to the class with the highest predicted probability). Darker green areas indicate higher predicted likelihood of occurrence. Higher probabilities of Floodplain Forests and Swamps are predicted in lower topographic positions and valley bottoms, as expected. Northern Hardwoods and Red Spruce types show higher probabilities at higher elevations and within the Allegheny Highlands physiographic region and are subsequently differentiated spectrally at a finer spatial scale (see bottom inset map in Figure 4). Hemlock, Mixed Mesophytic, and both Oak Forest community types are spread across the state and are differentiated at the hillslope-scale. Specifically, topographic aspect and its association with incoming solar radiation and moisture levels appears to be important for differentiating the Mixed Mesophytic and Oak/Hickory types, with Mixed Mesophytic dominating in wetter aspects (e.g., northeast facing slopes) and Oak/Hickory dominating in drier aspects (e.g., southwest facing slopes) (see top inset map in Figure 4). More generally, Oak/Hickory and Oak/Pine forests were predicted to be more abundant than Mixed Mesophytic in the drier Valley and Ridge physiographic section in the eastern portion of the state. Generally, Floodplain and Swamp, Red Spruce, and Northern Hardwoods were predicted with higher probabilities than the other types, suggesting that they are more separable based on spectral and/or terrain characteristics.

The UA and PA rates were highest overall for Floodplain and Swamps with the highest UA being 92.0% for Floodplain and Swamps in the RF All model and the highest PA being 94.9% for the SVM All model (Table 10, Table 11). Hemlock had lower accuracy rates as measured with UA and PA, and Red Spruce was also notably low with 0.0% in both tables for the SVM All and SVM Terrain models. The *k*NN model using all predictor variables showed confusion between the Hemlock and Mixed Mesophytic Forest classes as well as the Oak/Hickory and Oak/Pine types. Within the only terrain predictors model, the Hemlock class was confused with each of the other community types, suggesting that this class may occupy a range of topographic positions on the landscape. The Oak/Pine and Oak/Hickory classes showed higher confusion in all models, but were especially confused in the SVM models. This pattern continued in the *k*NN model using only spectral predictor variables. Notably, both classes including Oak trees showed higher confusion with Floodplain and Swamp forests.

The confusion matrices, from which the summary metrics presented above were derived, compare the pixels in a class with the number of pixels in the same class in the withheld 1/3rd of the WVDNR data. Tables 12 through 14 display accuracy for the *k*NN model with Table 12 (model including all variables) showing generally lower confusion at the class level. Within these tables there was notable confusion within the Oak/Pine and Oak/Hickory community types, which I attribute to these types occurring in similar topographic conditions while also having similar spectral signatures. Or, they are not well separated using the provided predictor variables. Also, these classes are somewhat similar and may share gradational boundaries. Tables 15 through 17 show class level-error for the RF models including the final classification output (i.e., Table 15 provides an assessment of the classification presented in Figure 4). These tables show a higher level of agreement, aligning with the higher accuracy measures from Table 10. Specifically, these models seemed better at distinguishing between the Hemlock and Mixed Mesophytic classes. Tables 18 through 20 are for the SVM models. This model had the greatest issue with classifying Red Spruce with the All variables and terrain only models not properly classifying any Red Spruce and with the spectral only model only properly identifying two. Overall, the Red Spruce class had the least misclassification by class for the *k*NN and RF models. It appears from the combination of these tables that the mixed classes, such as Mixed Mesophytic or Oak/Hickory, had greater confusion when compared with Hemlock or Northern Hardwoods.

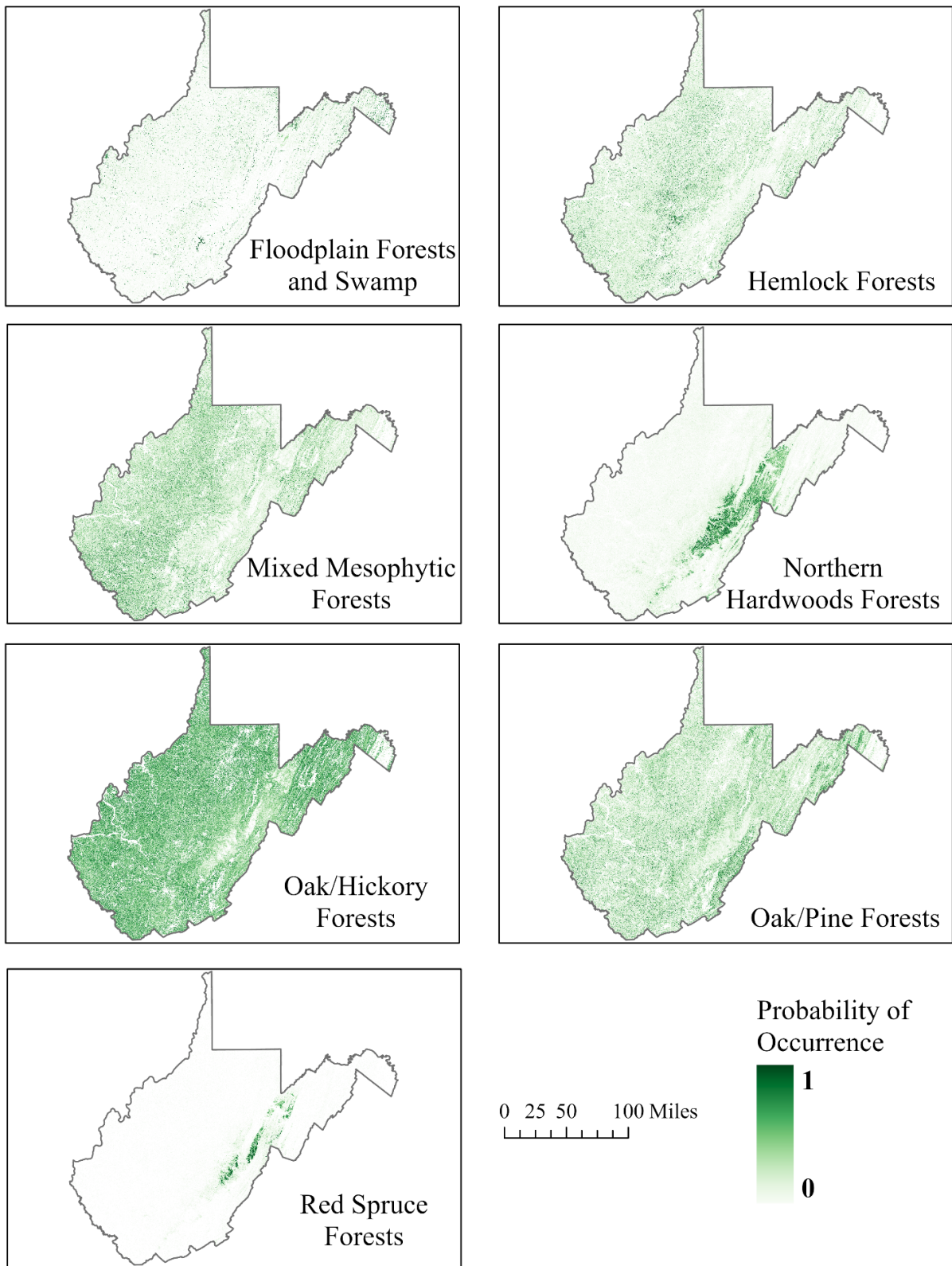


Figure 3. Probability raster grids for each of the forest classes in West Virginia.

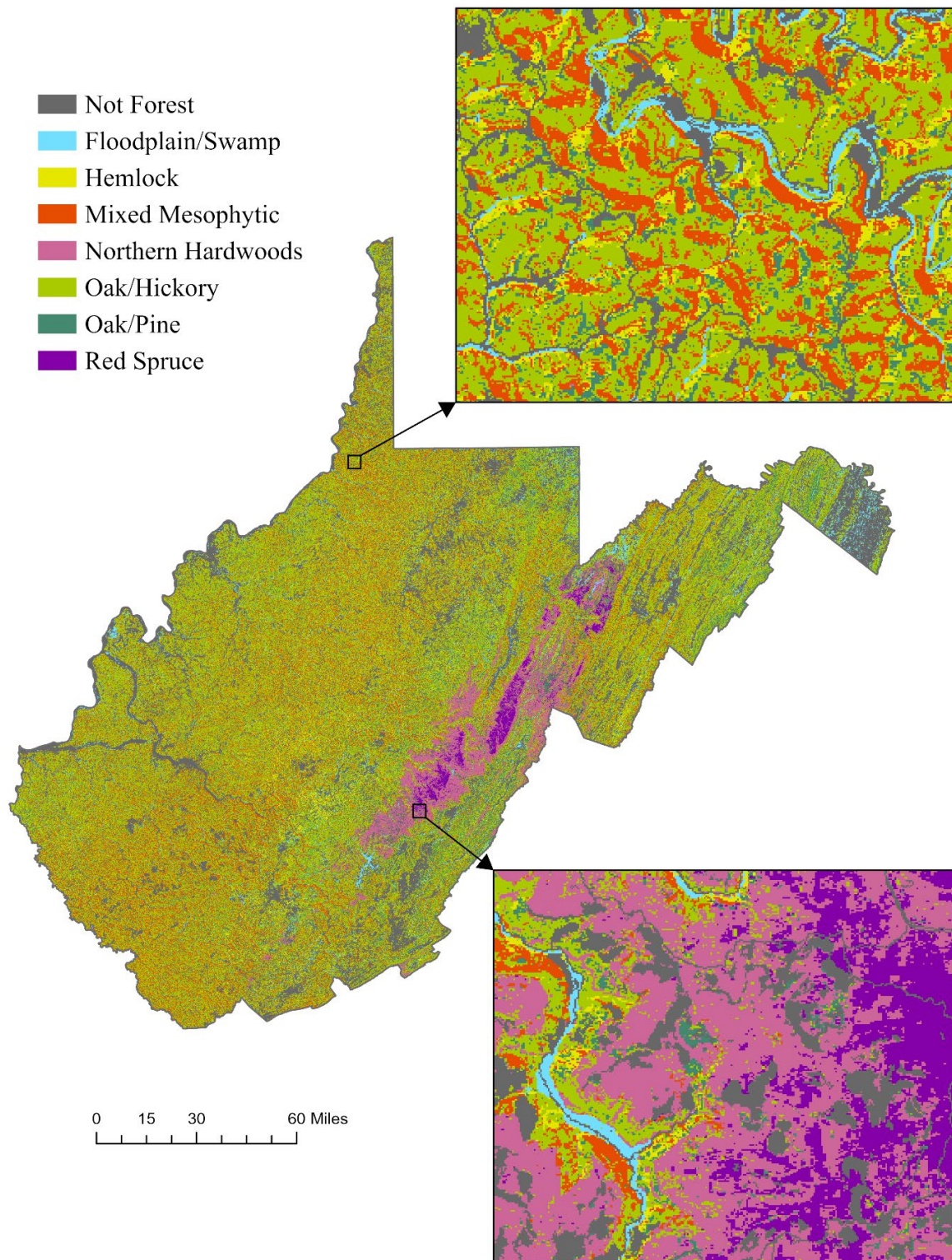


Figure 4. Final map of classification results based on RF using spectral and terrain variables with each color representing an individual community type.

Table 10. Confusion Matrix from *k*NN model using all predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	124	3	0	4	0	1	4	136	85.5
	Hemlock	4	20	9	3	3	6	4	49	24.1
	Mix. Meso.	3	17	38	3	13	3	0	77	43.2
	North. Hard.	1	3	5	23	3	3	8	46	37.7
	Oak/Hick.	9	25	28	12	74	55	2	205	66.1
	Oak/Pine	2	15	8	12	19	55	13	124	44.7
	Red Spruce	2	0	0	4	0	0	26	32	45.6
	Totals	145	83	88	61	112	123	57		
PA	91.2	40.8	49.4	50.0	36.1	44.4	81.2			

Table 11. Confusion Matrix from *k*NN model using only terrain predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	124	7	2	2	0	0	1	136	86.7
	Hemlock	5	11	12	3	8	6	4	49	14.3
	Mix. Meso.	3	14	42	4	12	1	1	77	44.2
	North. Hard.	2	2	2	19	2	4	15	46	33.9
	Oak/Hick.	5	28	32	10	54	62	14	205	54.0
	Oak/Pine	3	15	5	14	24	50	13	124	40.7
	Red Spruce	1	0	0	4	0	0	27	32	36.0
	Totals	143	77	95	56	100	123	75		
PA	91.2	22.4	54.5	41.3	26.3	40.3	84.4			

Table 12. Confusion Matrix from *k*NN model using only spectral predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	100	5	11	2	7	7	4	136	60.2
	Hemlock	4	26	5	2	3	5	4	49	26.3
	Mix. Meso.	9	15	24	7	14	8	0	77	32.0
	North. Hard.	3	2	4	20	4	6	7	46	36.4
	Oak/Hick.	31	31	25	14	66	38	0	205	58.9
	Oak/Pine	17	19	6	6	18	45	13	124	40.9
	Red Spruce	2	1	0	4	0	1	24	32	46.2
	Totals	166	99	75	55	112	110	52		
PA	73.5	53.1	31.2	43.5	32.2	36.3	75.0			

Table 13. Confusion Matrix from RF model using all predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	126	1	2	1	4	1	1	136	92.0
	Hemlock	2	17	8	3	10	8	1	49	36.2
	Mix. Meso.	2	9	39	2	24	1	0	77	52.0
	North. Hard.	0	2	3	27	7	1	6	46	48.2
	Oak/Hick.	4	13	22	11	123	32	0	205	56.4
	Oak/Pine	2	5	1	9	49	53	5	124	55.2
	Red Spruce	1	0	0	3	1	0	27	32	67.5
	Totals	137	47	75	56	218	96	40		
PA	92.6	34.7	50.6	58.7	60.0	42.7	84.4			

Table 14. Confusion Matrix from RF model using only terrain predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	127	3	0	2	4	0	0	136	90.1
	Hemlock	5	5	11	3	17	7	1	49	17.9
	Mix. Meso.	3	7	42	3	20	2	0	77	51.2
	North. Hard.	1	0	4	22	4	4	11	46	38.6
	Oak/Hick.	3	8	20	13	111	48	2	205	53.9
	Oak/Pine	1	5	5	6	49	50	8	124	44.2
	Red Spruce	1	0	0	8	1	2	20	32	47.6
	Totals	141	28	82	57	206	113	42		
PA	93.4	10.2	54.5	47.8	54.1	40.3	62.5			

Table 15. Confusion Matrix from RF model using only spectral predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	100	3	4	0	23	4	2	136	60.6
	Hemlock	2	15	7	2	9	13	1	49	31.9
	Mix. Meso.	12	7	28	3	22	5	0	77	45.9
	North. Hard.	4	2	3	20	7	3	7	46	60.6
	Oak/Hick.	30	6	16	5	124	24	0	205	54.1
	Oak/Pine	13	13	3	1	42	47	5	124	49.0
	Red Spruce	4	1	0	2	2	0	23	32	60.5
	Totals	165	47	61	33	229	96	38		
PA	73.5	30.6	36.4	43.5	60.5	37.9	71.9			

Table 16. Confusion Matrix from SVM model using all predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	129	2	1	0	2	2	0	136	89.6
	Hemlock	2	17	21	3	4	1	1	49	20.0
	Mix. Meso.	3	6	54	3	11	0	0	77	38.3
	North. Hard.	1	3	3	21	7	11	0	46	39.6
	Oak/Hick.	5	25	46	10	102	17	0	205	61.8
	Oak/Pine	2	31	16	10	39	26	0	124	32.5
	Red Spruce	2	1	0	6	0	23	0	32	0.0
	Totals	144	85	141	53	165	80	1		
PA	94.9	34.7	70.1	45.7	47.9	21.0	0.0			

Table 17. Confusion Matrix from SVM model using only terrain predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	126	6	1	1	2	0	0	136	88.7
	Hemlock	4	12	20	3	5	5	0	49	11.1
	Mix. Meso.	3	7	61	3	3	0	0	77	27.5
	North. Hard.	1	3	3	29	2	8	0	46	36.2
	Oak/Hick.	5	48	97	15	30	10	0	205	52.6
	Oak/Pine	2	32	40	16	14	20	0	124	33.3
	Red Spruce	1	0	0	13	1	17	0	32	0.0
	Totals	142	108	222	80	57	60	0		
PA	92.6	24.5	79.2	63.0	14.6	16.1	0.0			

Table 18. Confusion Matrix from SVM model using only spectral predictor variables.

		Reference							Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce		
Prediction	Floodplain	100	10	2	1	22	1	0	136	49.5
	Hemlock	3	30	5	2	7	1	1	49	27.3
	Mix. Meso.	16	14	22	2	23	0	0	77	38.6
	North. Hard.	5	1	2	11	19	8	0	46	37.9
	Oak/Hick.	45	19	21	8	110	2	0	205	49.8
	Oak/Pine	28	35	5	2	39	11	4	124	25.6
	Red Spruce	5	1	0	3	1	20	2	32	28.6
	Totals	202	110	57	29	221	43	7		
PA	73.5	61.2	28.6	23.9	53.7	8.9	6.2			

Chapter 7. Discussion

Forest type classification over large spatial extents and a moderate spatial resolution (i.e., the Landsat scale) using remotely sensed data is a difficult task that has not yet been operationalized. One component of the issue is that it can be difficult to acquire cloud-free imagery for a large extent within a specific timeframe. For example, specific time periods during fall senescence or spring leaf-out may be key for differentiating communities; however, cloud-free data may not be available. Further, key dates would vary over large spatial extents as a result of latitudinal and elevational gradients, further complicating the selection of key imagery. This study generally supports the use of products that summarize large sets of multi-temporal imagery, such as the GLAD Phenology spectral data used in this study or harmonic regression coefficients used in prior studies (i.e., Adams et al. 2020 and Pasquarella et al. 2018), as such methods offer a means to generally characterize yearly central tendency and seasonal variability consistently across large spatial extents and for each individual cell. I argue that globally consistent datasets that offer aggregated spectral measurements over broad spatial extents and characterize seasonal variability, such as the GLAD Phenology metrics, are generally underused in spatial predictive mapping and modeling. Further, I argue that there is a need for the adoption of a consistent set of seasonal metrics from the Landsat timeseries to compliment other commonly used metrics, such as vegetation indices or the tasseled cap transformation.

Previous studies have used RF and spectral data for forest type mapping and obtained accuracy rates above 80%, such as Liu et al. (2018); however, these studies had much smaller spatial extents. I argue that the lower classification accuracies obtained in this study in comparison to other forest type mapping studies can be partially attributed to the large spatial extent that was mapped (i.e., the entire state of West Virginia), which spans multiple Landsat scenes, multiple physiographies with varying hill-slope characteristics, and over 1,400 meters of elevation change across approximately 3 degrees of latitude (37°-40° North). Studies that have attempted to map smaller spatial extents or areas that are covered by a single Landsat scene may

not capture the complexity of mapping over larger extents and, as a result, may offer optimistic assessments that may not scale to operational mapping of large spatial extents.

Another issue with this task is the complexity of mapping fuzzy or gradational classes as well as the natural mixed nature of forests. Specifically, this can be seen in the confusion between the Oak/Hickory and the Oak/Pine classes in the confusion matrices. This further highlights the value of probabilistic output, which I argue provides a more meaningful characterization or mapping when class definitions are fuzzy and class boundaries are gradational. Probabilistic outputs, such as those generated in this study and by Pasquarella et al. (2018), help highlight complex landscape patterns and should be used alongside hard classifications to characterize the landscape and mapping uncertainty more fully.

It should also be noted that the class definitions used and the number of differentiated classes mapped can have a large impact on the reported accuracy of the resulting classification. For example, in Table 19 below I have combined the Oak/Hickory and Oak/Pine classes into an Oak/Hickory/Pine class. This table was generated from the best obtained classification: RF using the spectral and terrain variables. With this change, overall accuracy increased from 61.6% to 73.0%, and MICE increased from 0.52 to 0.62.

Table 19. Confusion Matrix from RF model using only spectral and terrain variables with the Oak/Hickory and Oak/Pine classes combined to an Oak/Hickory/Pine class.

		Reference						Totals	UA
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick./Pine	Red Spruce		
Prediction	Floodplain	126	1	2	1	5	1	136	92.6
	Hemlock	2	17	8	3	18	1	49	34.7
	Mix. Meso.	2	9	39	2	25	0	77	50.6
	North. Hard.	0	2	3	27	8	6	46	58.7
	Oak/Hick./Pine	6	18	23	20	257	5	329	78.1
	Red Spruce	1	5	1	3	1	27	38	71.1
	Totals	137	52	76	56	314	40		
PA	92.0	32.7	51.3	48.2	81.0	67.5			

My results support prior findings, such as those of Liu et al. (2018) and Adams et al. (2020), that document the value of including digital terrain variables. Terrain variables are specifically useful for this task as they help characterize biophysical conditions that impact forest community composition. For example, areas where oak species are dominant likely experience dryer conditions, which could topographically be linked to higher sun exposure as experienced by southwest facing slopes in the Northern hemisphere. This work suggests that terrain derivatives should be included alongside spectral data to potentially improve separation of classes that are spectrally similar but occupy locations with varying abiotic conditions. Further,

digital elevation data are readily available at a moderate spatial resolution (e.g., the NED dataset in the United States). Thus, incorporating such data is generally possible. However, creating derivatives from DTMs can be time, memory, and computationally intensive. Making DTM derivatives more readily available could aid in their adoption for large-area mapping and modeling tasks. Being able to generate such output on-the-fly, such as using the recently introduced Raster Functions in the ArcGIS software environment, could speed up and simplify the use of such variables in predictive models.

Of the three tested ML algorithms, RF generally performed best. The highest OA, MICE, AUC ROC, and AUC PR metrics were obtained using the RF algorithm and a combination of spectral and terrain predictor variables. As noted by Maxwell et al. (2018), a single algorithm has not been shown to always perform better than others. Instead, the optimal algorithm is often case specific, and multiple algorithms should be assessed in order to gauge and compare performance for specific tasks. As a result, I suggest that multiple algorithms should be assessed. Other than just generally strong predictive performance, RF has other positive attributes, such as the ability to use predictor variables measured on different scales or a mix of continuous and categorical predictors, provide an assessment of predictor variable importance, and assess model performance using the out-of-bag samples. It is also generally computationally efficient and not highly sensitive to hyperparameter values, such as the number of trees in the ensemble and the number of variables to randomly select from at each decision node (Brieman 2001). Further, RF has been integrated into a variety of computational environments (e.g., Python and R), commercial software packages (e.g., ArcGIS Pro and ITT ENVI), and open-source tools (e.g., QGIS).

Based on the output classification raster grid (see Figure 4) it seems that aspect is heavily linked to community type across the state, which I attribute to its impact on and association with incoming solar radiation and moisture levels. Some specific classes were especially well differentiated, such as Floodplains and Swamps, which I attribute the distinct topographic position occupied. Also, Northern Hardwoods and Red Spruce were generally well separated from the other classes based on elevation. A lot of the confusion within the models resulted from differentiating between the mixed community classes such as Oak/Pine and Oak/Hickory types, and merging these classes increased accuracy as demonstrated above, potentially suggesting that they may not be adequately separable given the spectral and terrain predictor variables used. Further, given the fuzzy nature of class definitions, pixels may not fit well into either of these groups and may represent a mix of both types. Again, this highlights the value of the probabilistic output.

One other complexity in this modeling process was class imbalance. Generally, ML methods have been shown to more accurately predict the more abundant classes in the training set (James et al. 2013, Kuhn & Johnson 2013). In this example, I oversampled the minority classes to provide a balanced training set. However, this may have resulted in overfitting. Future studies could investigate other methods to deal with class imbalance, such as applying class weightings in the training process. The issue of class imbalance is an area in need of additional research, as many datasets are imbalanced due to true disparities in abundance across real

landscapes. In extreme cases of class imbalance, stratified sampling may not be possible, or it may simply be too difficult to collect a large number of minority class samples (James et al. 2013, Kuhn & Johnson 2013).

Chapter 8. Conclusion

This project is unique in that it explores operational mapping at the state scale and assessed spectral metrics that have not previously been explored for forest type mapping (i.e., the GLAD Phenology metrics). This study incorporated both spectral and terrain predictor variables along with field plot observations to classify forest types for all forested pixels in the state of West Virginia. An overall accuracy of 61.58% was obtained using the RF algorithm and a combination of spectral and terrain variables selected from a larger set using recursive feature elimination. Both a traditional hard classification output along with probability maps by community type were generated. My results highlight the complexity of forest type mapping and the value of probabilistic outputs and the inclusion of digital terrain variables. Increased knowledge about forest community types and how to classify them is important for modern forest management and ecosystem analysis, especially at large spatial extents such as entire states or countries. Future research projects might focus on how to better differentiate between mixed forest classes and species with overlapping ground distributions within machine learning-based models.

Chapter 9. References

- Adams, B., Iverson, L., Matthews, S., Peters, M., Prasad, A. and Hix, D.M. (2020). Mapping forest composition with Landsat time series: an evaluation of seasonal composites and harmonic regression. *Remote Sensing*, 12(4), p.610.
- Brieman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- Congalton, R.G., Green, K., 2019. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.
- Core, E. L., & Strausbaugh, P. D. (1964). Flora of West Virginia (3-1 ed., Ser. 65). Morgantown, WV: West Virginia University. Retrieved October 7, 2020, from <https://babel.hathitrust.org/cgi/pt?id=uva.x002700247&view=1up&seq=6>
- Cortes, C., Mohri, M., n.d. Confidence Intervals for the Area Under the ROC Curve 8.
- Costanza, J. K., Faber-Langendoen, D., Coulston, J. W., & Wear, D. N. (2018). Classifying forest inventory data into species-based forest community types at broad extents: Exploring tradeoffs among supervised and unsupervised approaches. *Forest Ecosystems*, 5(1). doi:10.1186/s40663-017-0123-x

- D'Este, M., Elia, M., Giannico, V., Spano, G., Laforteza, R., & Sanesi, G. (2021). Machine learning techniques for fine dead fuel load estimation using Multi-Source Remote Sensing Data. *Remote Sensing*, 13(9), 1658. doi:10.3390/rs13091658
- Debeer, D., Strobl, C., 2020. Conditional permutation importance revisited. *BMC bioinformatics* 21, 1–30.
- Dostálová, A., Wagner, W., Milenković, M., & Hollaus, M. (2018). Annual seasonality IN Sentinel-1 signal for Forest mapping and forest type classification. *International Journal of Remote Sensing*, 39(21), 7738–7760. <https://doi.org/10.1080/01431161.2018.1479788>
- Elkins, J. R. (2012). Chestnut Blight. Retrieved February 16, 2021, from <https://www.wvencyclopedia.org/articles/1159#:~:text=Loss%20of%20the%20American%20chestnut,rare%20because%20of%20the%20blight.>
- Evans, J. S., & Cushman, S. A. (2009). Gradient modeling of conifer species using random forests. *Landscape Ecology*, 24(5), 673–683. doi:10.1007/s10980-009-9341-0
- Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling Species Distribution and Change Using Random Forest. In *Predictive Species and Habitat Modeling in Landscape Ecology: 139 Concepts and Applications* (1st ed., pp. 139–159). New York City, New York: Springer-Verlag. doi:10.1007/978-1-4419-7390-0
- Fan, J., Upadhye, S., Worster, A., 2006. Understanding receiver operating characteristic (ROC) curves. *CJEM* 8, 19–20. <https://doi.org/10.1017/S1481803500013336>
- Foody, G.M., 2020. Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment* 239, 111630. <https://doi.org/10.1016/j.rse.2019.111630>
- Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., & Yang, Y. (2015). Fusing Landsat and MODIS data for Vegetation Monitoring. *IEEE Geoscience and Remote Sensing Magazine*, 3(3), 47–60. <https://doi.org/10.1109/mgrs.2015.2434351>
- Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied Optics*, 4(1), 11. doi:10.1364/ao.4.000011
- Ghosh, A., Sharma, R., & Joshi, P. (2014). Random forest classification of urban landscape using Landsat archive and Ancillary data: Combining seasonal maps with decision Level fusion. *Applied Geography*, 48, 31–41. doi:10.1016/j.apgeog.2014.01.003

- Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote Sensing*, 4(9), 2661-2693. doi:10.3390/rs4092661
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer Texts in Statistics. doi:10.1007/978-1-4614-7138-7
- Jin, S., Yang, L., Danielson, P., Homer, C., Fry, J., & Xian, G. (2013). A comprehensive change detection method for updating the national land cover database to circa 2011. *Remote Sensing of Environment*, 132, 159–175. <https://doi.org/10.1016/j.rse.2013.01.012>
- Keith, H., Mackey, B. G., & Lindenmayer, D. B. (2009). Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *Proceedings of the National Academy of Sciences*, 106(28), 11635-11640. doi:10.1073/pnas.0901970106
- Kim, M., Madden, M. and Warner, T.A. (2009). Forest type mapping using object-specific texture measures from multispectral Ikonos imagery. *Photogrammetric Engineering & Remote Sensing*, 75(7), pp.819-829.
- Kuhn, M. and Johnson, K., 2013. Applied predictive modelling Springer. New York Heidelberg Dordrecht London.
- Li, H. (2021). An Overview on Remote Sensing Image Classification Methods with a Focus on Support Vector Machine. 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), Signal Processing and Machine Learning (CONF-SPML), 2021 International Conference on, CONF-SPML, 50–56.
- Li Ma, Crawford, M. M., & Jinwen Tian. (2010). Local Manifold Learning-Based k -Nearest-Neighbor for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, Geoscience and Remote Sensing, IEEE Transactions on, IEEE Trans. Geosci. Remote Sensing, 48(11), 4099–4109.
- Liu, Y., Gong, W., Hu, X., & Gong, J. (2018). Forest type identification with random forest USING Sentinel-1A, Sentinel-2a, multi-temporal Landsat-8, and DTM Data. *Remote Sensing*, 10(6), 946. <https://doi.org/10.3390/rs10060946>
- Lu, M., Chen, B., Liao, X., Yue, T., Yue, H., Ren, S., Li, X., Nie, Z., & Xu, B. (2017). Forest types classification based on multi-source data fusion. *Remote Sensing*, 9(11), 1153. <https://doi.org/10.3390/rs9111153>
- Maxwell, A. E., Warner, T. A., & Strager, M. P. (2016). Predicting Palustrine Wetland Probability Using Random Forest Machine Learning and Digital Elevation Data-Derived

- Terrain Variables. *Photogrammetric Engineering & Remote Sensing*, 82(6), 437-447. Retrieved February 28, 2021, from <https://www.ingentaconnect.com/contentone/asprs/pers/2016/00000082/00000006/art00016?crawler=true>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural NETWORK-BASED deep Learning remote Sensing Studies—Part 2: Recommendations and best practices. *Remote Sensing*, 13(13), 2591. doi:10.3390/rs13132591
- Maxwell, A., Sharma, M., Kite, J., Donaldson, K., Thompson, J., Bell, M., & Maynard, S. (2020). Slope Failure Prediction Using Random Forest Machine Learning and LiDAR in an Eroded Folded Mountain Belt. *Remote Sensing*, 12(486). Retrieved February 28, 2021, from <https://www.mdpi.com/2072-4292/12/3/486>
- Melville, B., Lucieer, A., & Aryal, J. (2018). Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland Native Grassland communities in Tasmania, Australia. *International Journal of Applied Earth Observation and Geoinformation*, 66, 46-55. doi:10.1016/j.jag.2017.11.006
- Mery, G., Alfaro, R., Kanninen, M., & Lobovikov, M. (2005). *Changing paradigms in forestry: Repercussions for people and nature*. Retrieved November 1, 2020, from <https://www.iufro.org/science/wfse/forests-global-balance/>.
- Murphy, M.A., Evans, J.S. & Storfer, A. (2010). Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology*, 91(1), .252-261.
- Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22(12), 2377-2400. doi:10.1080/01431160117096
- Nink, S., Hill, J., Stoffels, J., Buddenbaum, H., Frantz, D. and Langshausen, J. (2019). Using Landsat and Sentinel-2 data for the generation of continuously updated forest type information layers in a cross-border region. *Remote Sensing*, 11(20), p.2337.
- Palmer, M., Kuegler, O., & Christensen, G. (2018). Oregon's Forest Resources, 2006–2015: Ten-Year forest inventory and analysis report. *United States Department of Agriculture Pacific Northwest Research Station*. doi:10.2737/pnw-gtr-971

- Pasquarella, V. J., Holden, C. E., & Woodcock, C. E. (2018). Improved mapping of forest type using spectral-temporal Landsat features. *Remote Sensing of Environment*, 210, 193-207. doi:10.1016/j.rse.2018.02.064
- Pontius, R.G., & Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32, 4407–4429. <https://doi.org/10.1080/01431161.2011.552923>
- Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Ying, Q. (2020). Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sensing*, 12(3), 426. doi:10.3390/rs12030426
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS One* 10, e0118432.
- Shao, G., Tang, L., Zhang, H., 2021. Introducing Image Classification Efficacies. *IEEE Access* 9, 134809–134816. <https://doi.org/10.1109/ACCESS.2021.3116526>
- Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 13, 6308–6325.
- Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *International Journal of Remote Sensing* 35, 4923–4939. <https://doi.org/10.1080/01431161.2014.930207>
- Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. *Remote Sensing of Environment* 132, 202–211. <https://doi.org/10.1016/j.rse.2013.01.016>
- Stehman, S.V., 2012. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sensing Letters* 3, 111–120. <https://doi.org/10.1080/01431161.2010.541950>
- Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62, 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Stehman, S.V., Czaplewski, R.L., 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64, 331–344. [https://doi.org/10.1016/S0034-4257\(98\)00010-8](https://doi.org/10.1016/S0034-4257(98)00010-8)

- Stehman, S.V., Foody, G.M., others, 2009. Accuracy assessment, in: The SAGE Handbook of Remote Sensing. Sage London, pp. 297–309.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9, 1–11.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics* 8, 1–21.
- Strobl, C., Hothorn, T., Zeileis, A., 2009. Party on!
- Wandishin, M.S., Mullen, S.J., 2009. Multiclass ROC Analysis. *Weather and Forecasting* 24, 530–547. <https://doi.org/10.1175/2008WAF2222119.1>
- Widmann, R. H., Dye, C. R., & Cook, G. W. (2007). Forests of the Mountain State. *Northern Research Station Resource Bulletin*. Retrieved February 28, 2021, from https://www.ncrs.fs.fed.us/pubs/rb/rb_nrs17.pdf
- Wright, C., & Gallant, A. (2007). Improved wetland remote sensing in Yellowstone National park using classification trees to Combine TM imagery and ancillary environmental data. *Remote Sensing of Environment*, 107(4), 582-605. doi:10.1016/j.rse.2006.10.019