Article

Forest Type Differentiation Using GLAD Phenology Metrics, Land Surface Parameters, and Machine Learning

Faith M. Hartley [1,†], Aaron E. Maxwell [1,*,†], Rick E. Landenberger [1] and Zachary J. Bortolot [2]

1   Department of Geology and Geography, West Virginia University, Morgantown, WV 26505, USA
2   Geography Program, James Madison University, Harrisonburg, VA 22807, USA
*   Correspondence: aaron.maxwell@mail.wvu.edu
†   These authors contributed equally to this work.

Abstract: This study investigates the mapping of forest community types for the entire state of West Virginia, United States, using Global Land Analysis and Discovery (GLAD) Phenology Metrics, Analysis Ready Data (ARD) derived from Landsat time series data, 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 was to explore the use of globally consistent ARD 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 188 GLAD Phenology Metrics used in the study resulted in an overall accuracy (OA) of 54.3% (map-level image classification efficacy (MICE) = 0.433). Accuracy increased to a mean OA of 64.8% (MICE = 0.496) when the Oak/Hickory and Oak/Pine classes were combined into an Oak Dominant class. Once selected terrain variables were added to the model, the mean OA for differentiating the seven forest types increased to 65.3% (MICE = 0.570), while the accuracy for differentiating six classes increased to 76.2% (MICE = 0.660). Our results highlight the benefits of combining spectral data and terrain variables and also the enhancement of the product's usefulness when probabilistic predictions 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 those trained using spring medians. We suggest further exploration of the GLAD Phenology Metrics as input for other spatial predictive mapping and modeling tasks.

Keywords: forest type mapping; forests; phenology; machine learning; digital terrain analysis; Landsat

1. Introduction

Forests are important ecosystems [1–3], as they provide wildlife habitats [4,5], play a major role in the carbon [6] and nitrogen [7–9] chemical cycles and climate and hydrologic systems [10–12], and provide a wealth of ecosystem services and resources to support society [4,13]. Decades of research have documented the ecological importance of temperate forests specifically, such as those in the eastern United States in general and the Appalachian region more specifically [14–16]. This importance is demonstrated by the abundance and variety of organisms they support (e.g., [17–20]) and their roles in local, regional, and global climate, chemical, and hydrologic systems (e.g., [6,21,22]). Appalachian forests occur in a humid, temperate continental climate with four distinct seasons and are dominated by a variety of tree species, including various deciduous genera and species (i.e., oaks (*Quercus*), hickories (*Carya*), maples (*Acer*), and American beech (*Fagus grandifolia*)) and conifers (i.e., eastern hemlock (*Tsuga canadensis*), white pine (*Pinus strobus*), and red spruce (*Picea rubens*)). Tree species composition and related forest community types vary across

environmental gradients and are also impacted by disturbance histories and management practices [14,16,23].

These ecosystems have been and continue to be impacted by land use practices and anthropogenic landscape change [22,24], global climate change [11,12], resource extraction [22,25,26], invasive species [27–29], and disease [15,30–32]. From the 1880s to 1920s, the Appalachian forests were harvested to provide raw materials needed to support societal expansion; as a result, old-growth forests are now rare in this region and, where present, tend to occur in small, isolated pockets or small patches [33–35]. Forest composition in the eastern United States and Appalachia has also been heavily impacted by pathogens, such as chestnut blight (*Cryphonectria parasitica*), which nearly eradicated the once abundant American chestnut (*Castanea dentata*). Forest management practices have evolved over centuries as humans and society adapted to live within, make use of, and manage the natural environment. Historically, this can be seen through indigenous community practices and more recently in state-centric management such as prescribed fires and forest thinning. Shifting from traditional exploitation that focused on increasing timber production to forest management (silviculture) and sustainability-based approaches has arisen to meet the challenges of protecting and sustaining forests and the communities they support. Specifically, sustainable forest management began in the 1960s and evolved to address deforestation and biodiversity loss resulting from widescale timber production [2,36,37].

Sustaining forests over time requires detailed and extensive information on a range of factors that affect forest health. Thus, mapping and monitoring forests in an accurate and efficient manner is of great importance in sustainability-based forest management, and multispectral satellite-based remote sensing is a key data source for undertaking this work [38,39]. As an example of a notable achievement, Hansen et al. [40] generated global datasets of forest cover and forest cover change between 2000 and 2012 from the Landsat archive using cloud computing methods at a 30 m moderate spatial resolution. Cloud computing platforms such as Google Earth Engine, along with hosted data archives and analysis-ready data (ARD), have greatly improved our ability to efficiently analyze moderate spatial resolution satellite data over large spatial extents [41]. However, key questions and limitations remain. Although forests can generally be accurately differentiated from other land cover classes using multispectral data, mapping of specific forest community types or individual species has proven more challenging [40,42,43] despite several early efforts to use Landsat data to make these differentiations (see, for example, Beaubien [44] and Bryant [45]). Derived products that differentiate between forest community types over broad spatial extents are generally not available. For example, the U.S. National Land Cover Database (NLCD) products [46] only differentiate three broad forest types: deciduous, evergreen, and mixed forests; although useful as a very broad classification, more detailed information is necessary if sustainability is the goal.

This complexity in differentiating broad community types can generally be attributed to (1) the difficulty in defining non-overlapping and distinct types, (2) gradational boundaries and/or fuzzy definitions between types, (3) co-occurrence of types within environmental and physical landscape regimes, (4) lack of spectral separability, and (5) complex disturbance histories and logging practices such as "selectively logging" one or two valuable species, which then result in temporary, artificial community types [47]. Different forest types with disparate tree species compositions offer unique environments with varying ecological, chemical, climatic, and hydrologic conditions; thus, despite the difficulties noted above, mapping and differentiating forest community types is important in understanding the ecological and environmental processes occurring at a location in the context of sustainable land management and climate change resiliency [1,2,39].

There is a need to explore globally consistent data products for mapping and differentiating broad forest community types at a moderate spatial resolution. Towards this goal, we evaluate the Global Land Analysis and Discovery (GLAD) Phenology Metrics [48], digital terrain variables derived from digital terrain models (DTMs) (also referred to as land surface models (LSMs)), and machine learning (ML) for differentiating broad forest commu-

nity types for the entire state of West Virginia, United States. Training data are derived from field plots collected as part of the West Virginia Natural Heritage Program (WVNHP). Both "hard" classifications and probabilistic outputs are generated, compared, and assessed, and the value of including terrain variables is explored. The results obtained using the GLAD Phenology Metrics are compared to other methods of summarizing multitemporal Landsat data and seasonal patterns.

2. Background

2.1. Remote Sensing for Forest Type Mapping

Table 1 below summarizes studies that have explored the mapping of forest community types. Note that this is not meant to be an exhaustive cataloging; instead, these studies are highlighted to make generalizations regarding this specific mapping task. Also, we intentionally selected studies that used moderate spatial resolution multispectral data from the Landsat sensors (e.g., Landsat 4 and 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and/or Landsat 8 Operational Land Imager (OLI)) or the Sentinel-2A/B Multispectral Instrument (MSI), potentially combined with additional ancillary data, such as DTM-derived digital terrain variables. Several consistent findings are notable.

First, using only digital terrain variables generally results in less accurate predictions in comparison to using only multispectral data (e.g., Adams et al. [49]); however, combining terrain variables with multispectral data generally improves classification accuracy in comparison to just using the spectral data (e.g., Liu et al. [50], Pasquarella et al. [51], and Adams et al. [49]). This is generally attributed to reduced confusion between classes that may not be well separated based on spectral reflectance within a limited set of bands in the visible, near-infrared (NIR), and shortwave infrared (SWIR) spectral ranges, but differ in regard to the topographic positions and conditions they occupy. For example, Adams et al. [49] note the value of the topographic aspect for differentiating community types that occupy sites with varying levels of moisture availability, which is correlated with this terrain variable (i.e., in the mid-latitudes of the northern hemisphere, southwestern-facing slopes tend to be drier, while northeastern-facing slopes tend to be wetter). Specifically, including the topographic aspect as a predictor variable improved the separability of the Dry Oak and Mixed Mesophytic forest types [49]. One complexity is that a variety of digital terrain variables can be derived from DTMs using variable settings (e.g., different moving window sizes and shapes); as a result, it is not always clear what variables and analysis scales may be of value for differentiating specific communities in different landscapes [52,53].

Prior studies have also generally documented that either including multispectral data from multiple seasons or summarizing the seasonality or phenology using a time series of data generally improves classification performance in comparison to using single-date imagery (e.g., Pasquarella et al. [51] and Adams et al. [49]). For differentiating eight forest types across two Landsat scenes comprising the western portion of the state of Massachusetts, United States, Pasquarella et al. [51] documented that single-date, late-autumn imagery outperformed early-spring and early-autumn imagery with an overall accuracy of around 74%, which they attributed to this seasonal timeframe offering the best phenological differentiation between the defined classes. Furthermore, combining data from multiple dates representing different seasons consistently outperformed all single-date results. The best performance was obtained using harmonic regression coefficients derived from a Landsat time series to more fully capture the phenology at the pixel location. Adams et al. [49] similarly suggest that summarizing the Landsat time series using harmonic regression coefficients outperforms models that use multi-date seasonal composites. Harmonic regression offers a means to summarize time and/or seasonal oscillations of a variable, such as pixel brightness, greenness, or wetness, by fitting a Fourier series, estimating coefficients, and using these coefficients as input predictor variables for predictive modeling [54]. More generally, Wilson et al. [54] argue for and demonstrate the use of harmonic regression

methods for deriving variables from a time series of multispectral data to estimate a variety of forest attributes.

Table 1. Example studies that have explored multitemporal, multispectral, and moderate-resolution datasets for forest type differentiation. Separate experiments included in the same publication are provided as separate records in the table.

Study	Data	Study Area	Methods	Classes	Overall Accuracy
Immitzer et al. [55]	MSI	Two small extents in Bavaria, Germany	Random Forest	7	64%
Immitzer et al. [55]	MSI	Two small extents in Bavaria, Germany	Random Forest; GEOBIA	7	66%
Liu et al. [50]	MSI	~226,000 ha	Random Forest; GEOBIA	8	54%
Liu et al. [50]	MSI; Terrain	~226,000 ha	Random Forest; GEOBIA	8	70%
Liu et al. [50]	MSI; OLI; SAR; Terrain	~226,000 ha	Random Forest; GEOBIA	8	83%
Pasquarella et al. [51]	TM; ETM+	Western Massachusetts, United States	Random Forest; Late-Autumn	8	74%
Pasquarella et al. [51]	TM; ETM+	Western Massachusetts, United States	Random Forest; Multi-Date;	8	79%
Pasquarella et al. [51]	TM; ETM+	Western Massachusetts, United States	Random Forest; Harmonic Regression	8	81%
Pasquarella et al. [51]	TM; ETM+; Terrain; Ancillary	Western Massachusetts, United States	Random Forest; Harmonic Regression	8	83%
Hościło and Lewandowska [56]	MSI	~380,000 ha	Random Forest	8	76%
Hościło and Lewandowska [56]	MSI; Terrain	~380,000 ha	Random Forest	8	82%
Adams et al. [49]	Terrain	17 Counties in Ohio, United States	Random Forest	7	51%
Adams et al. [49]	OLI	17 Counties in Ohio, United States	Random Forest; Seasonal Composites; Spectral Indices	7	62%
Adams et al. [49]	OLI	17 Counties in Ohio, United States	Random Forest; Harmonic Regression	7	66%
Adams et al. [49]	OLI; Terrain	17 Counties in Ohio, United States	Random Forest; Seasonal Composites; Spectral Indices	7	70%
Adams et al. [49]	OLI; Terrain	17 Counties in Ohio, United States	Random Forest; Harmonic Regression	7	75%

GEOBIA = Geographic object-based image analysis; SAR = Synthetic aperture radar; TM = Thematic Mapper; ETM+ = Enhanced Thematic Mapper Plus; OLI = Operational Land Imager; MSI = Multispectral Instrument.

In summary, prior studies have documented the value of incorporating digital terrain variables derived from DTMs and summarizing multispectral seasonal variability, as represented by a multispectral time series, for forest type mapping, as opposed to relying on only single-date imagery. This study attempts to expand upon this prior work by investigating the GLAD Phenology Metrics as another means to summarize phenology at

the pixel level from a time series of images. This is of specific value because choosing images, creating seasonal composites, and/or deriving harmonic regression coefficients can be time-intensive and computationally intensive, even when cloud computing platforms are used. As a result, there is value in assessing the utility of other means to characterize phenology, especially products that already provide globally consistent data that are analysis-ready.

2.2. GLAD Phenology

The Landsat-based spectral data used in this study come from the University of Maryland's Global Land Analysis and Discovery (GLAD) laboratory. Specifically, we used the GLAD Phenological Metrics Type C product, which is based on Landsat surface reflectance estimates and represents a set of globally consistent metrics to characterize land surface phenology [48]. This metric set was used alongside Global Ecosystem Dynamics Investigation (GEDI) light detection and ranging (LiDAR) data to estimate canopy heights for forests across the entire globe [57]. A full discussion of the methodology used to generate these metrics is provided by Potapov et al. [48] and is summarized here. Full documentation is available at the GLAD website (https://glad.umd.edu/ard/glad-landsat-ard-tools, accessed on 26 June 2022). These variables are provided to the public as Landsat Analysis Ready Data (ARD).

The process of creating ARD begins with the collection of Landsat 5 TM, Landsat 7 ETM+, and Landsat 8 OLI images from 1997 to the present. These images are then categorized into quality tiers, 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; thus, this dataset is unsuitable for wintertime image processing and surface water extent mapping. The processing chain includes estimation of surface reflectance, reflectance normalization, and observation quality assessment at the pixel level. It results in 16-day composites using the best available data. Before generating summary metrics, gaps in the time series are filled using observations from prior years or linear regression when quality observations from prior years are not available [48].

The GLAD Phenology Type C variables are summarized in Table 2. These variables consist of a set of statistics derived from the blue, green, red, NIR, SWIR1, and SWIR2 bands, along with derived indices calculated using normalized difference ratios between pairs of bands. The spectral variability vegetation index (SVVI) of Coulter et al. [58] is also summarized. The SVVI is calculated by subtracting the sum of the standard deviations of the NIR, SWIR1, and SWIR2 bands from the sum of the standard deviations of all included bands. Statistics are calculated from the 16-day time series observations using ranks derived from the band or index of interest or by calculating ranks using a corresponding variable. Additional phenological measures are calculated from a normalized difference vegetation index (NDVI) time series relating to the start of the growing season (SOS) and end of the growing season (EOS), including the value, slope, and amplitude. The Tasseled Cap Transformation Greenness (TCG) [59,60] metric is also summarized [48].

In this study, it is hypothesized that GLAD metrics will be useful for differentiating broad forest community types, as they quantify spectral signatures associated with seasonality and phenology. For example, it is anticipated that Northern Hardwoods will have a shorter growing season than Mixed Mesophytic and Oak/Hickory types. Red Spruce forests and other forests that have an evergreen component, such as Hemlock and Oak/Pine, may show some degree of spectral separation from forests predominantly composed of deciduous species. Given the large number of variables provided by GLAD (i.e., 188), it is not clear which may be useful for differentiating forest types and how metrics may interact within the feature space. As a result, we explore variable selection, as described below. Given that spectral confusion between community types is anticipated, we hypothesize that the incorporation of terrain variables will improve separability. For example, terrain variables may allow for differentiation based on moisture levels (e.g., heat load index (HLI) and topographic aspect) and temperature ranges and variability (e.g., elevation) that impact forest community compositions.

Table 2. Phenological variables used in analysis and included in the GLAD Type C metric set. Q1 indicates the 1st quartile, Q2 indicates the 2nd quartile (i.e., median), and Q3 indicates the 3rd quartile. Abbreviations defined in this table are used throughout this paper.

Metrics Based on Ranking of 16-Day Observation Time Series		
Spectral	**Indices**	**Statistics**
Blue Green Red NIR SWIR1 SWIR2	(NIR-Green)/(NIR + Green) (GN) (NIR-Red)/(NIR + Red) (RN) (NIR-SWIR1)/(NIR + SWIR1) (NS1) (NIR-SWIR2)/(NIR + SWIR2) (NS2) (SWIR1-SWIR2)/(SWIR1 + SWIR2) (SWSW) SVVI Tasseled Cap Greenness (TCG)	Minimum (min) Maximum (max) Median (median) Average between min and Q1 (avgminQ1) Average between Q3 and max (avQ3max) Average between Q1 and Q3 (avQ1Q3) Average of all values (avg) Standard deviation (sd) Total absolute difference (tad) Amplitude min to max (avgminmax) Amplitude Q1 to Q3 (ampQ1Q3) Amplitude Q2 to max (amp Q2max)
Metrics Based on Ranking of 16-Day Observation Time Series by Value of Corresponding Variable		
Bands	**Corresponding Variables**	**Statistics**
Blue Green Red NIR SWIR1 SWIR2	(NIR-Red)/(NIR + Red) (RN) (NIR-SWIR2)/(NIR + SWIR2) (NS2) Brightness Temperature (LST)	Minimum (min) Maximum (max) Average between min and Q1 (avgminQ1) Average between Q3 and max (avQ3max) Amplitude min to max (avgminmax) Amplitude Q1 to Q3 (ampQ1Q3)
NDVI-Based Phenology Metrics		
Index		**Phenology Metrics**
(NIR-Red)/(NIR + Red) (RN)		Start of season value (RNph_sos) End of season value (RNph_eos) Start of season slope (RNph_sos_slope) End of season slope (RNph_eos_slope) Start of season amplitude (RNph_sos_amp) End of season amplitude (RNph_eos_amp) Growing season average (RNph_ave) Growing season total (RNph_sum)

2.3. Machine Learning

Within remote sensing and land classification, ML has matured to become an operational tool for extracting actionable information from large, complex datasets [61–63]. ML serves as a framework for identifying important variables, building accurate predictions, and exploring complex relationships and spatial patterns within a model. These algorithms tend to produce higher overall classification accuracies than traditional parametric classification methods (e.g., Gaussian maximum likelihood), which is attributed to their nonparametric nature and ability to model complex patterns and relationships within a complex feature space (i.e., many variables that may be correlated, of different measurement levels, measured on different scales, and/or not normally distributed) [62,63].

We initially experimented with several machine learning methods as a means to select an algorithm for use in this study. Ultimately, random forest (RF) was selected due to its classification performance in comparison to the other algorithms tested (support vector machines, *k*-nearest neighbors, and single decision trees). RF is an ensemble decision tree method proposed by Breiman [64]. 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 (i.e., 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. In other words, a set of weak classifiers is collectively strong and generalizes well due to reduced overfitting. The final classification is determined based on the majority vote amongst the trees in the ensemble, and probabilities for each mapped class can be obtained based on the relative proportion of class votes [64]. As highlighted in Table 1 above, RF has been applied to forest type mapping in several prior studies.

3. Materials and Methods

3.1. Study Area and Field Plots

The entire state of West Virginia, United States, was investigated in this study (Figure 1). Forests are the dominant land cover type in the state; specifically, an analysis of the NLCD 2019 land cover product [65] suggests that roughly 80% of the state is forested, which is equivalent to over 5 million hectares of land area. Dominant forest community types vary among the three physiographic provinces within the state. Furthest west, the Allegheny Plateau, which is the largest province in the state and is situated between the Ohio River to the west and the mountainous Allegheny Highlands to the east, is characterized by oak–pine, oak–chestnut, floodplain, and cove hardwoods or mixed mesophytic forests. This landscape is underlain by flat to gently dipping sedimentary rock units and has a high degree of local relief due to dissection by a dendritic stream network. The second division is the Allegheny Highlands section, which includes a northeast-to-southwest-oriented mountain range dominated by the red spruce and northern hardwood community types at the highest elevations. This area contains the highest elevations in the state, generally receives the highest annual rainfall, and has the shortest growing season. The Ridge and Valley province, immediately east of the Allegheny Mountains, is generally the driest and lowest-elevation section of the state and is dominated by oak–hickory–pine community types. The landscape is characterized by structurally controlled linear ridges and valleys (i.e., a highly eroded folded mountain belt) and a trellis drainage network [23].

Species composition data came from the West Virginia Division of Natural Resources' (WVDNR) Natural Heritage Program [66]. The WVDNR's Natural Heritage Program is responsible for collecting and maintaining field-based ecological community data. Specifically, the database includes information associated with vegetation structure, plant species composition, environmental characteristics, and location. The primary uses of these data are to support classification and characterization of vegetation in the state as needed for conservation purposes. Forest community types are determined based on field observations. Plots range in size from 25 to 400 m^2 (i.e., smaller than a Landsat pixel). The dates of collections are variable, ranging from 1963 to 2012 [66]. As described below, we excluded any plots that were not mapped as forest or woody wetland in multiple dates of the NLCD land cover products. So, it was assumed that if a plot was mapped as forested in the NLCD products, its forest type had not changed since the time of the field collection.

To reduce the number of categories of different community types and to limit the analysis to classes with a sufficient number of samples, the Hill Country Deciduous Forest and Successional classes were omitted. Additionally, Calcareous Forests and Woodlands, Oak/Hickory, and Dry/Mesic Oak Forests were merged into an Oak/Hickory class. Dry Rocky Pine/Oak Forests and Woodlands and Oak/Pine Forests were merged into an Oak/Pine class. Furthermore, based on an intersection of land cover datasets with the point coordinates associated with the plots, any plots that were not mapped as forested or woody wetland in the 2011, 2013, 2016, and 2019 NLCD land cover products [46,65] were excluded from the analysis. The number of available samples by class is summarized in Table 3. A total of 2216 samples, grouped into seven classes, were used in the study. Due to confusion between the Oak/Hickory and Oak/Pine classes, these classes were also merged into an Oak Dominant class to explore the differentiation of six classes as a separate component of the study.

Table 3. Forest community types mapped and the associated number of available field plots.

Community Type	Plot Count
Floodplain	495
Hemlock	167
Mixed Mesophytic	249
Northern Hardwoods	153
Oak/Hickory	648
Oak/Pine	396
Red Spruce	108
Total	2216
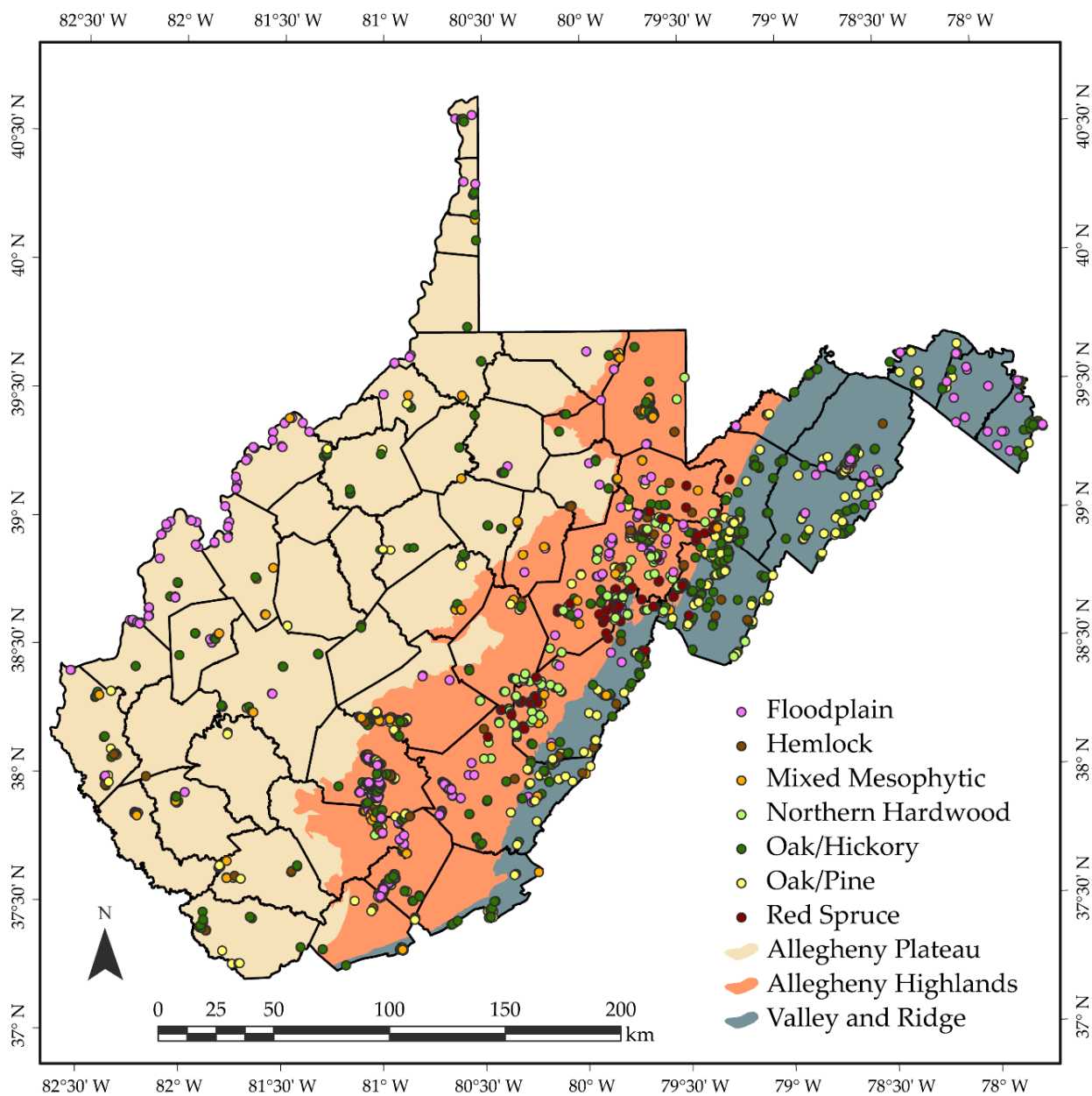
Figure 1. West Virginia, United States, study area map including WVNHP plot locations, physiographic regions, and county boundaries. Physiographic boundaries are represented using Major Land Resource Areas (MLRAs) as provided by the United States Department of Agriculture (USDA) Natural Resource Conservation Service (NRCS) [67].
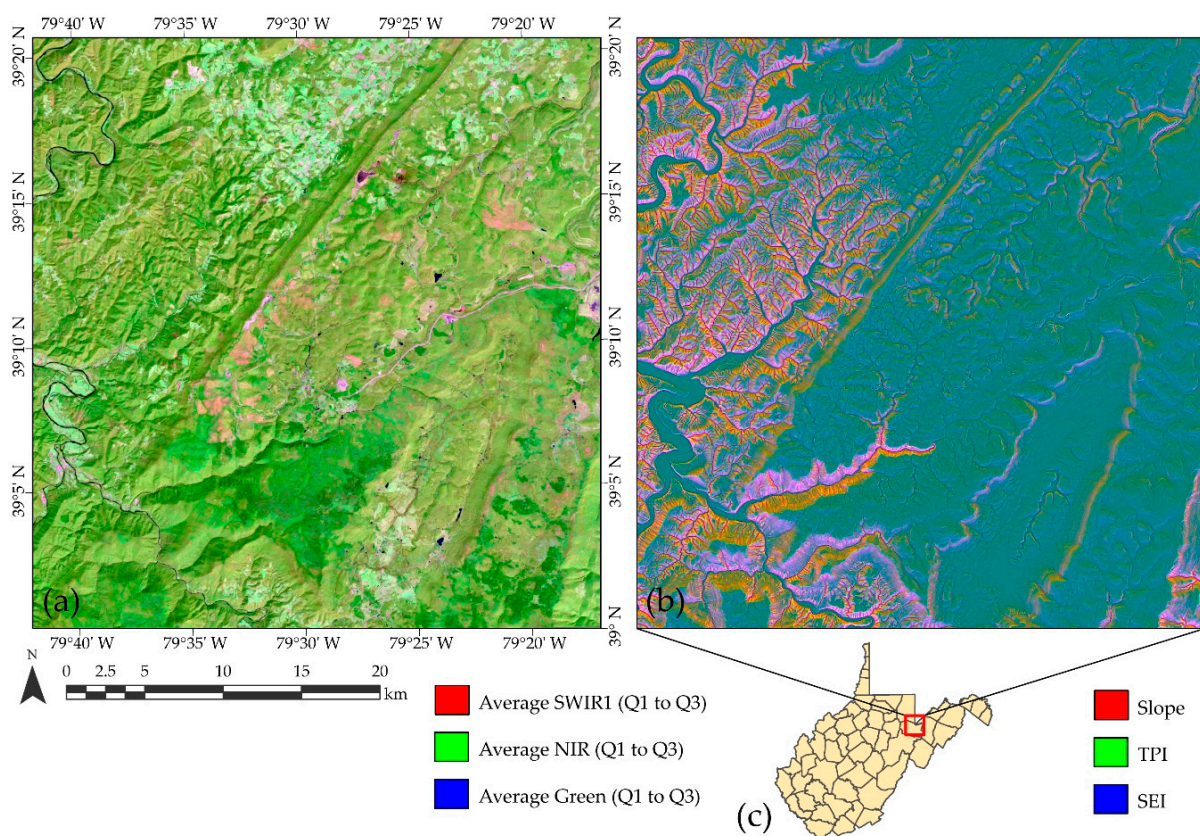
3.2. Predictor Variables

Table 4 lists the number of predictor variables calculated for each set of features explored and the associated abbreviations used in this study. GLAD 16-day composites were downloaded using Perl [68] scripts provided by the GLAD laboratory. The metric sets were also calculated using provided Perl scripts relative to the year 2019, which required downloading all 16-day composites for that year plus all composites from the prior four years in order to fill data gaps [48]. GLAD makes 16-day composites available, and metric sets must be generated using the provided scripts relative to a certain year, in this case 2019, using composites from that year plus data from prior years to fill data gaps. A total of 188 GLAD variables relative to 2019 were included. Figure 2 provides examples of GLAD spectral and digital terrain variables used in this study.

Table 4. Metric sets and associated abbreviations and the number of variables.

Feature Set	Abbreviation	Number of Variables
GLAD Phenology Type C	G	188
Digital Terrain Variables	T	17
Harmonic Regression Coefficients	H	32
Summer	Sm	10
Fall	Fall	10
Spring	Spr	10

Figure 2. Example spectral and digital terrain predictor variables. (**a**) False color composite showing GLAD Phenology Type C average spectral values between the 1st and 3rd quartiles for SWIR1, NIR, and green. (**b**) Slope, TPI, and SEI digital terrain variables. (**c**) Extent of (**a**,**b**) within West Virginia, United States.

To compare results obtained using the GLAD Phenology Metrics to other means by which to summarize spectral data, we also used Google Earth Engine [41] and the

gee_subset Python library [69] to extract all Landsat 8 OLI observations from the Level 2, Collection 2, Tier 1 archive that were collected between January 2013 and December 2022 at each sample location. The quality flags were then used to extract only clear observations not predicted to be contaminated by clouds, cloud shadows, or cirrus clouds. From the original image bands for each observation, we then calculated the normalized difference vegetation index (NDVI) [69] and the brightness (TCB), greenness (TCG), and wetness (TCW) Tasseled Cap Transformations [59]. For the blue, green, red, NIR, SWIR1, and SWIR2 bands and the NDVI, TCB, TCG, and TCW derivatives, we calculated the median values over seasonal time periods including summer (June, July, and August) (Sm), fall (September, October, and November) (Fall), and spring (March, April, and May) (Spr). We did not include winter median values due to the likelihood of snow cover. This process resulted in a total of 10 variables for each season (Table 4). Lastly, we calculated harmonic regression coefficients using a third-order Fourier series fitted using ordinary least squares and the rHarmonics [70] package within the R data science environment [71]. Specifically, harmonic regression coefficients were calculated from the NDVI, TCB, TCG, and TCW measures, resulting in a total of 32 variables (Table 4).

Statewide DTM raster data were obtained from the National Elevation Dataset (NED) [72] at a 1/3-arc second (approximately 10 m) spatial resolution. From these elevation data, 17 digital terrain variables were generated to characterize the local terrain, as summarized in Table 5. 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); mean (CrvMn), profile (CrvPro), and tangential (Crv-Tan) surface curvatures; topographic position index (TPI); topographic roughness index (TRI); topographic dissection index (TDI); surface area ratio (SAR); and surface relief ratio (SRR) [52,53,73–76]. Variables relating to moisture content and local incoming radiation that were used include linear aspect (LnAsp), cosine aspect transformation (AspCos), sine aspect transformation (AspSin), topographic radiation aspect index (TRASP), heat load index (HLI), and site exposure index (SEI) [52,53,74,76]. These 17 metrics were used because they may correlate with forest community type distributions, characterize different aspects of the terrain surface, and can all be derived using only a DTM.

Table 5. Digital terrain variables calculated from a digital terrain model (DTM) and used in this study.

Variable	Abbreviation	Description/Equation
Linear Aspect	AspLn	$270 - \frac{360}{2\pi} \times \arctan^2\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$
Cosine Aspect Transformation	AspCos	Cos(Aspect); measure of eastwardness
Sine Aspect Transformation	AspSin	Sin(Aspect); measure of northwardness
Topographic Radiation Aspect Index	TRASP	$\frac{1 - \cos\left(\left(\frac{\pi}{180}\right) \times (\text{Asp} - 30)\right)}{2}$
Elevation	Elev	Bare-ground surface height
Slope (Degrees)	Slp	$\text{Arctan}\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right)\left(\frac{180}{\pi}\right)$
Mean Slope	SlpMn	Calculates slope within a moving window
Mean Curvature	CrvMn	Average of minimum and maximum curvatures
Profile Curvature	CrvPro	Curvature in direction of maximum slope
Tangential Curvature	CrvTan	Curvature in direction tangent to contour line
Topographic Position Index	TPI	$z - z_{mean}$
Topographic Dissection Index	TDI	$\frac{z - z_{min}}{z_{max} - z_{min}}$
Topographic Roughness Index	TRI	$\sigma^2(z)$

Table 5. *Cont.*

Variable	Abbreviation	Description/Equation
Surface Area Ratio	SAR	$\dfrac{\text{Cell Size}^2}{\text{Cos(Slope in Degrees)}}$
Surface Relief Ratio	SRR	$\dfrac{z_{mean}-z_{min}}{z_{max}-z_{min}}$
Heat Load Index	HLI	Index for annual direct incoming solar radiation based on latitude, slope, and aspect
Site Exposure Index	SEI	$\text{Slope} \times \cos\left(\pi\dfrac{\text{Aspect }-180}{180}\right)$

For variables that require defining a local moving window, a circular window with a radius of 5 cells, or 50 m, was used. All terrain derivatives were rescaled from a 10 m to a 30 m spatial resolution using pixel aggregation and the mean value from the original 9 cells within the new, larger cell to match the spatial resolution of the Landsat-derived metrics. This process was undertaken using ArcPy [77] and the ArcGIS Pro [78] software.

Once all the predictor variables were calculated, cell values at the plot locations were extracted using either the Extract Multi-Values to Points Tool [79] in ArcGIS Pro [77,78] or the raster package [80] in R. The resulting data table was used as input for all subsequent modeling and assessment.

3.3. Feature Selection, Hyperparameter Optimization, and Model Training

The randomForest [81] and caret [82] packages in the R data science environment [71] were used to optimize and train the RF algorithm using the different feature spaces investigated. RF requires the user to define the number of trees in the ensemble (ntree) and the number of random predictor variables from which to sample at each node (mtry). The number of trees used in the RF model was set at 1000, 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 [64]. The mtry hyperparameter was optimized using 10-fold cross-validation and a grid search over a set of candidate values. Once the optimal mtry parameter was selected, 50 training replicates were run using different training and testing partitions in order to quantify the variability in the classification results. Training and testing partitions were selected using a bootstrap method in which 70% of the data were used to train the model while the remaining 30% were used to assess the model performance, with each model run being provided with a different data partition. To combat class imbalance, classes were weighted relative to the inverse of their abundance in the WVNHP plots used in the study.

Due to the larger number of variables included in the GLAD Phenology dataset and in order to combat potentially reduced classification performance due to the Hughes phenomenon (i.e., "curse of dimensionality") [83], a recursive feature elimination process was employed to select a subset of the available features in any models in which the GLAD metrics were included in the feature space. This algorithm was implemented in R [71] using the rfUtilities package [84], which uses the RF algorithm. We tested the top sets of variables from 100% to 10% with a step size of 10%. The parsimony parameter, which allows for a smaller and simpler feature set to be selected if only a slight decrease in accuracy is recorded, was set to 0.05. In other words, a more parsimonious model was selected if the accuracy was not substantially reduced. Variable selection was not implemented when the feature space being assessed did not include the GLAD Phenology Metrics, as the remaining sets did not include a large number of variables.

3.4. Model Assessment and Comparison

For each of the 50 model replicates using each feature space, trained models were used to predict the withheld validation data and generate confusion matrices. From these confusion matrices, we 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 (PA) (1—omission error) accuracies [85–87]. In order to aggregate the class-level metrics, we calculated the average class user's (aUA) and producer's (aPA) accuracies and also combined the metrics into an average F1-score (aF1). F1-score is the harmonic mean of the user's and producer's accuracies [88]. It should be noted that the relative proportion of classes in the confusion matrix represents those from the WVNHP plots used in this study, which may not align with the true proportions of the landscape. Thus, confusion matrices represent a population confusion matrix relative to the used plots but may not represent a true population confusion matrix for the mapped landscape [89–92].

Because the commonly used Kappa statistic has come under scrutiny and its use in remote sensing is now discouraged [93,94], we did not calculate this metric. Instead, we calculated map-level image classification efficacy (MICE) [95], which only uses the reference class margin totals as opposed to both the reference and classification margin totals to adjust 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 is imbalanced while not suffering from the logical flaws of Kappa [95]. Lastly, we also calculated the percent of the validation samples in which the correct classification was within the top three predicted class probabilities (top 3) as a means to assess how often the correct class is included amongst the classes predicted with the highest likelihood.

All 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, because we also wanted to assess the probabilistic products, measures that consider 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 the class producer's accuracy (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 because the user's accuracy is impacted by class proportions [88,96–100]. Also, because this classification task was not a binary classification problem, multiclass versions of the metrics were used as implemented in the multiROC R package [101]. Specifically, micro-average AUCs were calculated by stacking all groups together, which is more sensitive to class imbalance than the alternative macro-averaging method not used in this study [101].

4. Results and Discussion

4.1. Results Using GLAD Phenology Metrics and Terrain Variables

Table 6 provides the accuracy assessment results for the models that made use of the GLAD Phenology Metrics. The first two rows correspond to the prediction of seven classes, whereas the last two rows relate to the models in which the Oak/Hickory and Oak/Pine classes were combined into an Oak Dominant class, resulting in a total of six classes. The mean overall accuracy calculated for 50 model replicates for differentiating the seven forest community types using only variables selected from the 188 GLAD Phenology Metrics was 54.3% (MICE = 0.433). The accuracy increased to 64.8% (MICE = 0.496) when the Oak/Hickory and Oak/Pine classes were combined.

Once selected terrain variables were added to the model, the mean overall accuracy for differentiating the seven classes increased by 11.0% to 65.3% (MICE = 0.570), while the mean accuracy for differentiating the six classes increased by 11.4% to 76.2% (MICE = 0.660). Furthermore, variable importance estimates provided by the RF models generally suggested that the 10 most important variables in the G + T models were all terrain variables. Thus, our results highlight the value of including terrain variables as predictor variables for differentiating forest community types. This is in support of prior studies that have documented the importance of terrain variables for forest community differentiation (e.g., Liu et al. [50], Pasquarella et al. [51], Hościło and Lewandowska [56], and Adams et al. [49]).

Table 6. Accuracy assessment results for models with selected GLAD Phenology Metrics as predictor variables. Average values are based on 50 model replicates. G = GLAD Phenology Metrics, and T = Digital Terrain Variables.

Set	Number of Classes	OA	MICE	Top 3	aUA	aPA	aFS	AUC ROC	AUC PR
G	7	0.543	0.433	0.886	0.484	0.547	0.497	0.875	0.579
G + T	7	0.653	0.570	0.938	0.587	0.637	0.601	0.930	0.730
G	6	0.648	0.496	0.933	0.501	0.601	0.527	0.906	0.698
G + T	6	0.762	0.660	0.966	0.615	0.673	0.631	0.953	0.837

Generally, classes were mapped with a higher PA than UA, which suggests that commission errors were generally more prominent than omission errors. This is also reflected in the average multiclass micro-AUC ROC and micro-AUC PR results, in which AUC ROC was generally higher than AUC PR. As noted above, the precision-recall curve and associated AUC metric consider both UA and PA and are generally less positively biased than AUC ROC when classes are imbalanced [97,98], as is the case in this study. Thus, these results further highlight the impact of commission errors.

As highlighted by the top 3 metric included in Table 6, even if the correct class was not selected, it was commonly one of the top three classes predicted to have the highest probability of occurrence. For example, the correct class was in the top three highest predicted probabilities for 93.8% of the withheld samples on average for the seven-class G+T model. Furthermore, the AUC ROC and AUC PR results generally suggest stronger performance than the associated "hard" classification metrics (i.e., OA and MICE). This highlights the value of the probabilistic predictions. We argue that probabilistic output can be especially informative for classes that may not be well differentiated or have fuzzy definitions or gradational boundaries on the landscape.

Figure 3 shows an example hard classification for a model that used GLAD Phenology Metrics and terrain predictor variables selected from the larger set using the feature selection process of Murphy et al. [84] and implemented in the rfUtilities R package [84]. This output, and the probability surfaces shown in Figure 4, were generated by predicting to each pixel within each GLAD processing tile then merging the results into a state-wide product. Within the state extent, Northern Hardwoods and Red Spruce forests are predicted predominantly at high elevations in the Allegheny Highlands, whereas Oak/Hickory and Oak/Pine types dominate in the drier Valley and Ridge physiographic region. Forest types in the Allegheny Plateau are generally differentiated at a finer, hillslope spatial scale, and the impact of the topographic aspect is evident. The Mixed Mesophytic forest type was more likely to be predicted for northeast-facing slopes, whereas Oak/Hickory or Oak/Pine were more commonly predicted on southeast-facing slopes, similar to the findings of Adams et al. [49] in the eastern hill region of Ohio, United States, a region that has similar topographic characteristics and forest communities to those occurring in the Allegheny Plateau section of West Virginia. Floodplain communities were generally predicted in lower topographic positions and valley bottoms, as expected. Classes that had a more variable presentation on the landscape in regard to the topographic positions and elevations at which they were predicted are the Oak/Pine and Hemlock classes.

The probabilistic results shown in Figure 4 correspond to the hard classification presented in Figure 3. Again, the Northern Hardwood and Red Spruce classes had the highest predicted probabilities at higher elevations in the Allegheny Highlands and low predicted probabilities at other locations in the state. The Floodplain class was prominent at lower topographic positions and in valley bottoms, as anticipated. The Oak/Hickory, Oak/Pine, Mixed Mesophytic, and Hemlock classes were less aligned with topographic and elevation gradients and tended to have a wide range of predicted probabilities across the state. As noted above, we argue that these results are of added value, as they provide additional and complementary information alongside the classification output.

For wetland mapping in Yellowstone National Park, Wright and Gallant [102] also highlight the value of probabilistic predictions as a means to complement classification output. Because many machine learning methods generate probabilistic predictions alongside hard classifications [103,104], we argue that this additional information should be provided to end users, especially when classes are difficult to separate with high accuracy or certainty, or when there are gradational boundaries between classes in the landscape. For example, as moisture content is associated with topographic aspect, it is expected that forest community types that occupy drier conditions, such as Oak–Hickory–Pine communities, will have gradational boundaries with community types that occupy wetter conditions, such as Mixed Mesophytic communities. Such gradational patterns are not captured in the classification output.
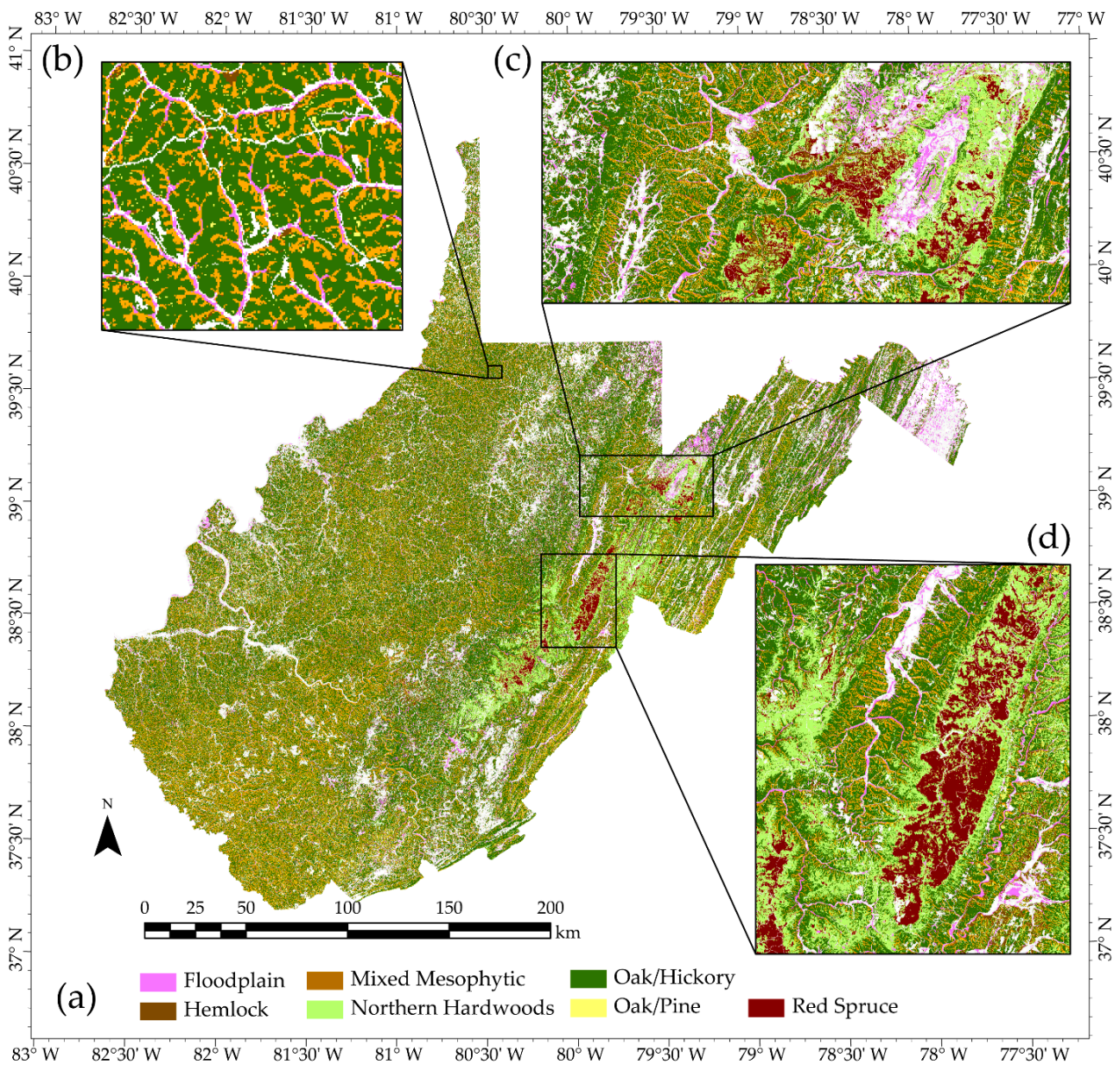
Figure 3. Example classification output created using the selected GLAD Phenology Metrics and digital terrain variables for differentiating seven classes. Non-forest areas, such as development, pastureland, and water, were masked from the results using the 2019 NLCD land cover product [65]. (**a**) shows the results for the entire state while (**b**–**d**) show the results at a higher spatial resolution.
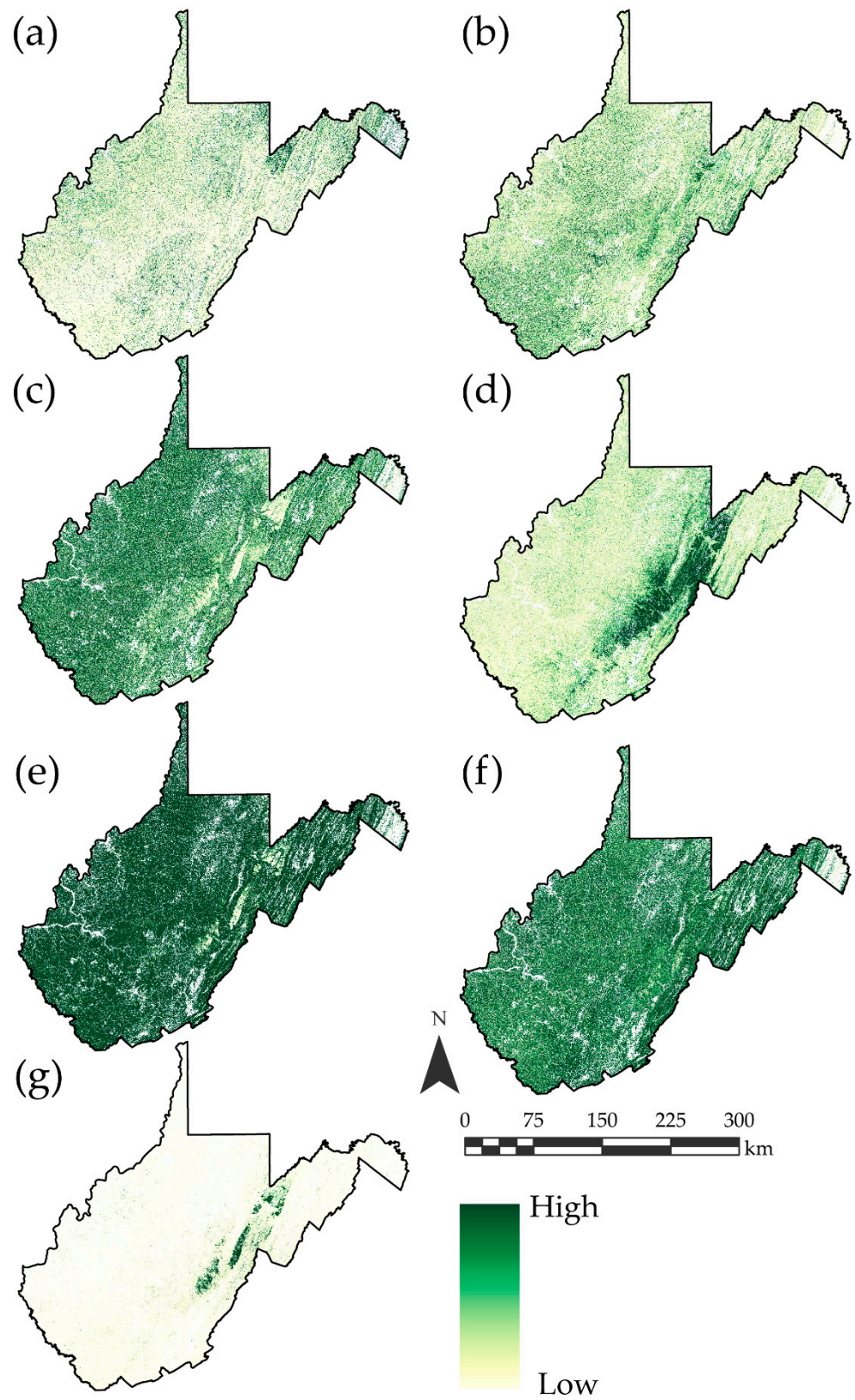
Figure 4. Example probabilistic output created using the selected GLAD Phenology Metrics and terrain variables. These data were created using the same model as that used to generate the classification presented in Figure 3. (**a**) Floodplain, (**b**) Hemlock, (**c**) Mixed Mesophytic, (**d**) Northern Hardwood, (**e**) Oak/Hickory, (**f**) Oak/Pine, (**g**) Red Spruce.

Table 7 provides the confusion matrix associated with the data presented in Figures 3 and 4. Using a combination of spectral and terrain variables, the Floodplain and Red Spruce classes were generally well differentiated from the other communities. In contrast, the Hemlock class showed confusion with the Mixed Mesophytic, Oak/Hickory, and Oak/Pine classes and was difficult to accurately map. A large degree of confusion was generally noted between the Oak/Hickory and Oak/Pine classes. For comparison, Table 8 provides an example result generated using only the spectral data. Without the inclusion of terrain variables, the Floodplain class was not as well differentiated, and there was generally more confusion between the Mixed Mesophytic and Oak/Hickory types. Notably, all UAs decreased for all classes except Red Spruce, which remained the same, and PAs decreased for four of the seven classes when the terrain variables were not included; thus, this generally suggests that the terrain variables were important variables for mapping all the classes and not just for differentiating certain community types.

Table 7. Example confusion matrix for seven-class classification using the selected GLAD Phenology Metrics and terrain variables. Overall accuracy = 0.648, MICE = 0.563.

		Reference								
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce	Totals	UA
Prediction	Floodplain	137	3	1	1	4	7	1	154	0.890
	Hemlock	2	15	3	1	2	5	1	29	0.517
	Mix. Meso.	0	17	40	5	18	2	0	82	0.488
	North. Hard.	0	1	2	16	4	0	6	29	0.552
	Oak/Hick.	1	7	28	15	140	51	0	242	0.579
	Oak/Pine	0	7	0	2	26	51	0	86	0.593
	Red Spruce	0	0	0	4	0	3	24	31	0.774
	Totals	140	50	74	44	194	119	32		
	PA	0.979	0.300	0.541	0.364	0.722	0.429	0.750		

Table 8. Example confusion matrix for seven-class classification using just the selected GLAD Phenology Metrics. Overall accuracy = 0.528, MICE = 0.415.

		Reference								
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce	Totals	UA
Prediction	Floodplain	108	3	8	3	30	11	2	165	0.655
	Hemlock	4	16	5	0	4	7	2	38	0.421
	Mix. Meso.	0	9	12	4	17	4	0	46	0.261
	North. Hard.	0	2	1	14	7	0	3	27	0.519
	Oak/Hick.	18	10	45	19	119	43	1	255	0.467
	Oak/Pine	10	8	3	1	17	52	0	91	0.571
	Red Spruce	0	2	0	3	0	2	24	31	0.774
	Totals	140	50	74	44	194	119	32		
	PA	0.771	0.320	0.162	0.318	0.613	0.437	0.750		

Table 9 provides an example confusion matrix for differentiating the six classes as opposed to seven using selected GLAD Phenology Metrics and terrain variables. Again, confusion between the Oak/Hickory and Oak/Pine classes is a dominant source of error in

the seven-class result (Table 7), and combining these classes greatly improved the model performance. This highlights the importance of the number of classes and specific class definitions for the final model accuracy. Unfortunately, some community types which may be important to differentiate, such as Oak Dominant and Mixed Mesophytic, may not be easily separated. Once more, this highlights the value of the probabilistic output.

Table 9. Example confusion matrix for six-class classification using the selected GLAD Phenology Metrics and terrain variables. Overall accuracy = 0.761, MICE = 0.658.

		Reference							
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak Dominant	Red Spruce	Totals	UA
Prediction	Floodplain	135	4	2	2	8	1	152	0.888
	Hemlock	1	10	1	1	5	1	19	0.526
	Mix. Meso.	1	11	36	2	13	0	63	0.571
	North. Hard.	0	1	2	13	3	7	26	0.500
	Oak/Hick.	3	24	33	22	280	0	362	0.773
	Red Spruce	0	0	0	4	4	23	31	0.742
	Totals	140	50	74	44	313	32		
	PA	0.964	0.200	0.486	0.295	0.895	0.719		

4.2. Comparison of GLAD Phenology Metric Results with Other Methods

Table 10 below provides a comparison of the mean assessment metrics for the model replicates with different feature spaces. The top portion of the table includes all models that made use of the digital terrain variables, while the bottom portion consists of all models that only used spectral data. Figure 5 shows the distribution of these metrics amongst the 50 model replicates for each feature space. Generally, the models that incorporated the terrain variables outperformed the associated models using just the spectral characteristics. Similar to the findings of Adams et al. [49], even though the terrain variables were of importance in the model, models that only used these terrain variables performed poorly in comparison to models that combined them with spectral predictors, which highlights the value of combining these disparate data types.

Table 10. Average assessment metrics for different feature spaces calculated from 50 model replicates for predicting seven classes. G = GLAD Phenology Metrics, T = Digital Terrain Variables, H = Harmonic Regression Coefficients, Sm = Summer, Fall = Fall, Spr = Spring.

Set	OA	MICE	Top 3	aUA	aPA	aFS	AUR ROC	AUC PR
T	0.574	0.471	0.891	0.463	0.495	0.462	0.899	0.649
Sm + T	0.636	0.549	0.928	0.564	0.607	0.571	0.925	0.715
Fall + T	0.632	0.544	0.918	0.563	0.597	0.564	0.921	0.705
Spr + T	0.665	0.584	0.933	0.606	0.639	0.615	0.933	0.742
H + T	0.702	0.631	0.950	0.641	0.691	0.657	0.946	0.779
G + T	0.653	0.570	0.938	0.587	0.637	0.601	0.930	0.730
G + H + T	0.705	0.634	0.952	0.647	0.697	0.663	0.947	0.780
All	0.709	0.640	0.953	0.652	0.701	0.669	0.946	0.778
Sm	0.451	0.320	0.808	0.387	0.462	0.405	0.815	0.450
Fall	0.463	0.334	0.820	0.421	0.463	0.431	0.828	0.478
Spr	0.487	0.364	0.809	0.419	0.443	0.417	0.833	0.513
H	0.626	0.537	0.910	0.579	0.630	0.595	0.907	0.654
G	0.543	0.433	0.886	0.484	0.547	0.497	0.875	0.579
G + H	0.637	0.550	0.924	0.588	0.643	0.606	0.918	0.687
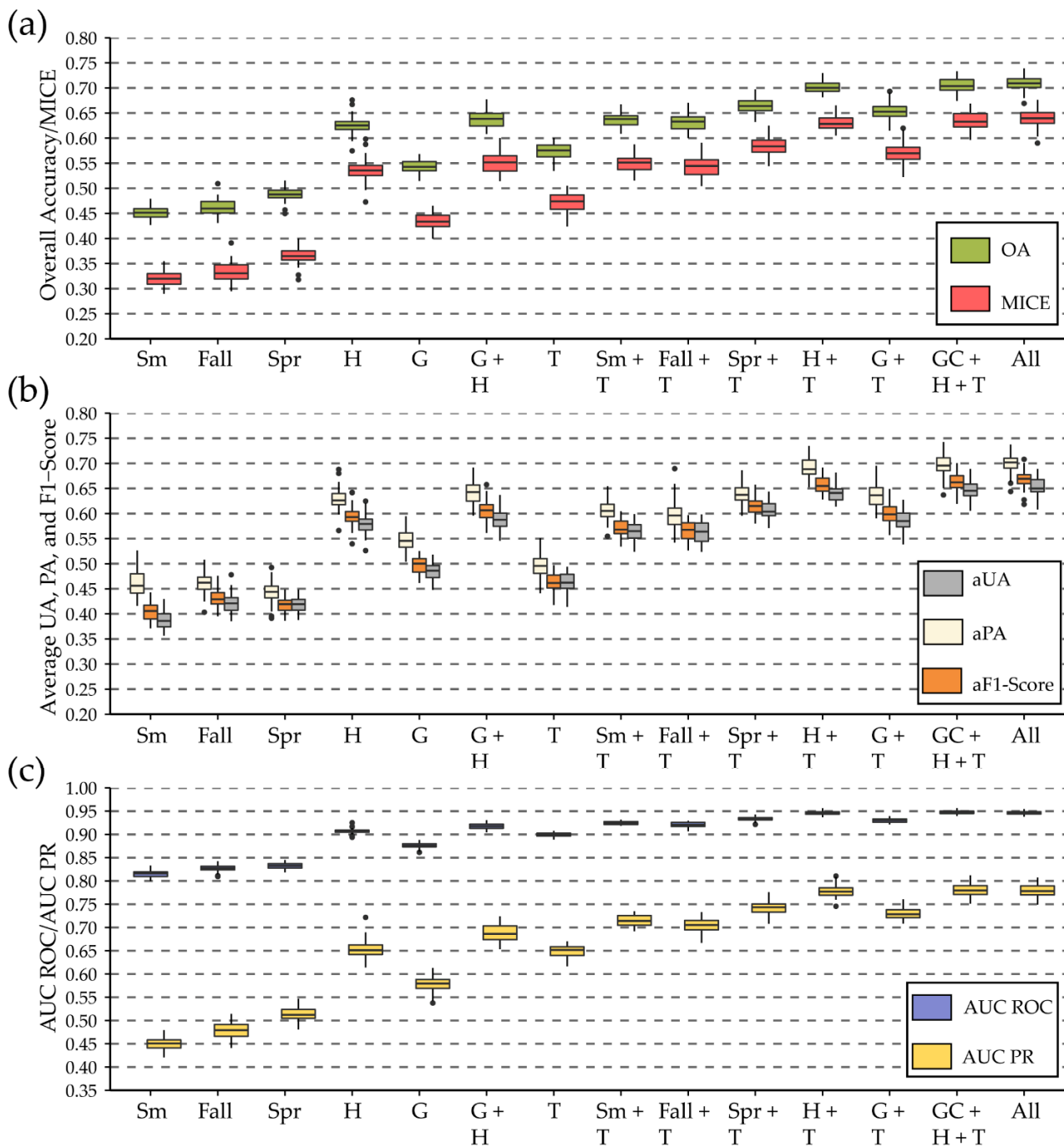
Figure 5. Boxplots showing distribution of accuracy assessment results for 50 replicates of each feature space. (**a**) OA and MICE, (**b**) aUA, aPA, and AF1-Score, (**c**) multiclass AUC ROC and AUC PR.

In the models that used only median values from a single season or combined one of these sets with the terrain variables, the spring imagery generally performed better than the summer and fall imagery. The GLAD Phenology Metrics generally performed similarly to the spring imagery or better if the terrain variables were not included; however, the best results based on the average assessment metrics generally resulted from using a combination of the harmonic regression coefficients and the terrain variables. The models using harmonic regression coefficients and terrain variables yielded a mean overall accuracy of 70.2%, whereas the models using the GLAD Phenology Metrics and the terrain variables yielded an overall accuracy of 65.3%. Including both the GLAD Phenology Metrics and

harmonic regression coefficients alongside the digital terrain variables yielded an OA of 70.5% (MICE = 0.634); thus, average accuracy only improved by 0.3% for the G + H + T model in comparison to the G + H model. When the terrain variables were not included, the G + H model yielded a 1.1% improvement in accuracy in comparison to the G model. In summary, including the GLAD Phenology Metrics did not greatly improve models trained using just the harmonic regression coefficients.

Allowing the model to choose from all of the available predictor variables (i.e., summer, fall, and spring medians; harmonic regression coefficients; GLAD Phenology Metrics; and terrain variables) yielded the highest accuracy of 70.9% (MICE = 0.640). In summary, the GLAD Phenology Metrics generally outperformed the summer and fall spectral medians and performed similarly to the spring medians or better if the terrain variables were not included; however, they did not provide the level of performance achievable when using the harmonic regression coefficients. Thus, this study supports findings from prior studies that highlight the value of harmonic regression methods for summarizing a time series of imagery and for representing multitemporal, multispectral data as predictor variables in spatial predictive models (e.g., Adams et al. [49], Pasquarella et al. [51], and Wilson et al. [54]).

4.3. Summary of Key Findings and Recommendations

The key results of this study can be summarized as follows:

1. Including digital terrain variables generally improved the accuracy of forest type differentiation compared to only using spectral data.
2. The number of classes and class definitions can have a large impact on the accuracy of the resulting map products.
3. Even if the correct class was not always predicted with the highest probability, it was generally in the top set of the highest predicted probabilities. We attribute this result to specific classes being difficult to separate, pixels not mapping well to a specific class, and/or class boundaries being gradational within the landscape. This highlights the value of supplementing "hard" classification products with associated probabilistic predictions.
4. GLAD Phenology Metrics were generally of value for mapping and differentiating forest community types. However, they did not provide the level of accuracy obtained using harmonic regression coefficients.

Forest type classification over large spatial extents and at 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 challenging to acquire cloud-free imagery for a large extent within a specific timeframe, a problem that is potentially compounded because humid and cloudy regions typically support diverse forest community types. Specific time periods during fall senescence or spring leaf-out may be vital for differentiating communities; however, cloud-free data may not be available. Furthermore, key dates would likely 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 multitemporal imagery, such as the GLAD Phenology Metrics used in this study or harmonic regression coefficients used in prior studies (e.g., Adams et al. [49] and Pasquarella et al. [51]), 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. We 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. Furthermore, we argue that there is a need for the adoption of a consistent set of seasonal metrics from the Landsat time series to complement other commonly used metrics, such as vegetation indices or the Tasseled Cap Transformation. As this study suggests that the GLAD Phenology Metrics did not perform as well as harmonic regression coefficients, generating ARD products representing

harmonic regression coefficients should be pursued by data providers so that users do not have to generate these metrics from scratch using the available time series and so that consistent methods are adopted across studies. Also, using a different processing year for generating the GLAD metrics, or a different range of years for performing harmonic regression analysis, may impact the model accuracy.

Although the GLAD Phenology Metrics did not offer the level of performance obtained with the harmonic regression coefficients, they do offer some key additional advantages. These data are generated using a consistent methodology and are available globally. Gap-filling methods allow for the estimation of metrics at every pixel location, which supports wall-to-wall mapping efforts. Lastly, these data can be easily downloaded and processed using the scripts provided by the data originators [48]. Thus, we argue that these data should be explored as input for additional mapping and modeling tasks.

Previous studies have used RF and spectral data for forest type mapping and obtained accuracy rates above 80%, such as Liu et al. [50]; however, these studies had much smaller spatial extents. We argue that the lower classification accuracies obtained in this study in comparison to some other forest type mapping experiments 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 hillslope characteristics, and over 1400 m 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. This study was conducted at the scale of an entire state with inconsistent physiographic characteristics and multiple Landsat scenes in order to specifically explore operational, large-area mapping using spectral and terrain variables.

Another issue with this task is the complexity of mapping fuzzy or gradational classes as well as the naturally heterogenous nature of forest community types. Specifically, this is seen in the confusion between the Oak/Hickory and Oak/Pine classes in the confusion matrices and the improvement in accuracy when these classes were combined. Furthermore, inherently mixed classes consisting of a variety of tree species, such as Mixed Mesophytic, were particularly difficult to map due to confusion with multiple other classes. This further highlights the value of probabilistic output, which we 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. [51], highlight complex landscape patterns and should be used alongside hard classifications to characterize the landscape and mapping uncertainty more fully.

Our findings support those of Liu et al. [50], Pasquarella et al. [51], Hościło and Lewandowska [56], and Adams et al. [49], in that digital terrain variables were found to be important for mapping and differentiating forest community types. Furthermore, 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-intensive, memory-intensive, and computationally intensive [52,53]. Making DTM derivatives more readily available could aid in their adoption for large-area and operational mapping and modeling tasks. Being able to generate such output on the fly, such as by using the recently introduced Raster Functions in the ArcGIS software environment [78], could speed up and simplify the use of such variables in spatial predictive mapping and modeling. As another option, derivatives could be generated in a cloud computing environment, such as Google Earth Engine [41], which has DTM data available within its data catalog.

There were a few notable limitations to this study. First, we had to make use of available ground plot data that may not represent the full variability and correct class proportions for the entire mapped extent (i.e., the entire state of West Virginia). We used the NLCD data to remove plots that were not mapped as forested and assumed that the remaining plots had the same forest type as when they were field-surveyed. It is possible

that some plots may have been disturbed, introducing some uncertainty into the analysis. Unfortunately, the lack of a large number of randomized reference samples is a common issue for empirical modeling tasks generally and forest type differentiation more specifically, as also highlighted by Pasquarella et al. [51]. Future studies should explore other datasets for input into forest type predictive modeling tasks, such as those made available by the Forest Inventory and Analysis (FIA) dataset curated by the United States Department of Agriculture (USDA) Forest Service in the United States. There is also a need to investigate ARD datasets, such as the GLAD Phenology Metrics, for estimating other forest attributes. For example, a continuous variable, such as percent oak or percent "dry-adapted species", could be predicted as opposed to differentiating a set of predefined classes.

5. Conclusions

This study highlights the difficulty of differentiating complex forest community types over a large spatial extent using moderate spatial resolution data while highlighting the value of digital terrain variables and ARD data products. We also document the value of generating both hard classifications and probabilistic predictions, especially when classes are difficult to separate, have fuzzy definitions, and/or have gradational boundaries in the landscape. The number of classes and defined classes being mapped and differentiated can also have a large impact on model performance. The GLAD Phenology Metrics that were specifically explored in this study generally outperformed summer and fall median values and performed similarly to spring median values. However, they did not provide the level of accuracy obtained using harmonic regression coefficients, which highlights the need to generate ARD data that consistently represent such variables and ease their use in spatial predictive mapping and modeling tasks. Although the GLAD metrics did not perform to the level of the harmonic regression coefficients in this specific study and application, we argue that they should be considered as a spectral input for mapping tasks because they are well documented, globally consistent, can be generated without data gaps, and are available as ARD products.

Author Contributions: Conceptualization, F.M.H. and A.E.M.; formal analysis, F.M.H. and A.E.M.; investigation, F.M.H. and A.E.M.; data curation, A.E.M.; writing—original draft preparation, F.M.H. and A.E.M.; writing—review and editing, F.M.H., A.E.M., R.E.L. and Z.J.B. All authors have read and agreed to the published version of the manuscript.

References

1. Binkley, D. *Forest Ecology: An Evidence-Based Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
2. Franklin, J.F.; Johnson, K.N.; Johnson, D.L. *Ecological Forest Management*; Waveland Press: Long Grove, IL, USA, 2018.
3. Smail, R.A. *Forest Land Conversion, Ecosystem Services, and Economic Issues for Policy: A Review*; General Technical Report PNW-GTR-797; U.S. Department of Agriculture, Forest Service: Washington, DC, USA, 2009.
4. Hasan, S.S.; Zhen, L.; Miah, M.G.; Ahamed, T.; Samie, A. Impact of Land Use Change on Ecosystem Services: A Review. *Environ. Dev.* **2020**, *34*, 100527. [CrossRef]

5. Foley, J.A. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [CrossRef] [PubMed]
6. Houghton, R.A.; Hackler, J.L.; Lawrence, K.T. The US Carbon Budget: Contributions from Land-Use Change. *Science* **1999**, *285*, 574–578. [CrossRef]
7. Galloway, J.N. The Global Nitrogen Cycle: Changes and Consequences. *Environ. Pollut.* **1998**, *102*, 15–24. [CrossRef]
8. Gruber, N.; Galloway, J.N. An Earth-System Perspective of the Global Nitrogen Cycle. *Nature* **2008**, *451*, 293–296. [CrossRef] [PubMed]
9. Vitousek, P.M.; Aber, J.D.; Howarth, R.W.; Likens, G.E.; Matson, P.A.; Schindler, D.W.; Schlesinger, W.H.; Tilman, D.G. Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. *Ecol. Appl.* **1997**, *7*, 737–750. [CrossRef]
10. Bonan, G.B. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **2008**, *320*, 1444–1449. [CrossRef] [PubMed]
11. Jackson, R.B.; Randerson, J.T.; Canadell, J.G.; Anderson, R.G.; Avissar, R.; Baldocchi, D.D.; Bonan, G.B.; Caldeira, K.; Diffenbaugh, N.S.; Field, C.B.; et al. Protecting Climate with Forests. *Environ. Res. Lett.* **2008**, *3*, 044006. [CrossRef]
12. Kirilenko, A.P.; Sedjo, R.A. Climate Change Impacts on Forestry. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19697–19702. [CrossRef] [PubMed]
13. Chow, J.; Kopp, R.J.; Portney, P.R. Energy Resources and Global Development. *Science* **2003**, *302*, 1528–1531. [CrossRef] [PubMed]
14. Delcourt, H.R. Eastern Deciduous Forests. 2000. Available online: https://ci.nii.ac.jp/naid/10020542173/ (accessed on 11 June 2022).
15. Dyer, J.M. Revisiting the Deciduous Forests of Eastern North America. *BioScience* **2006**, *56*, 341–352. [CrossRef]
16. Eastern Deciduous Forest (U.S. National Park Service). Available online: https://www.nps.gov/im/ncrn/eastern-deciduous-forest.htm (accessed on 30 May 2022).
17. Coddington, J.A.; Young, L.H.; Coyle, F.A. Estimating Spider Species Richness in a Southern Appalachian Cove Hardwood Forest. *J. Arachnol.* **1996**, *24*, 111–128.
18. Keddy, P.A.; Drummond, C.G. Ecological Properties for the Evaluation, Management, and Restoration of Temperate Deciduous Forest Ecosystems. *Ecol. Appl.* **1996**, *6*, 748–762. [CrossRef]
19. Green, N.B.; Pauley, T.K. *Amphibians and Reptiles in West Virginia*; University of Pittsburgh Press: Pittsburgh, PA, USA, 1987.
20. Wickham, J.; Wood, P.B.; Nicholson, M.C.; Jenkins, W.; Druckenbrod, D.; Suter, G.W.; Strager, M.P.; Mazzarella, C.; Galloway, W.; Amos, J. The Overlooked Terrestrial Impacts of Mountaintop Mining. *BioScience* **2013**, *63*, 335–348. [CrossRef]
21. Wickham, J.D.; Riitters, K.H.; Wade, T.; Coan, M.; Homer, C. The Effect of Appalachian Mountaintop Mining on Interior Forest. *Landsc. Ecol.* **2007**, *22*, 179–187. [CrossRef]
22. Drummond, M.A.; Loveland, T.R. Land-Use Pressure and a Transition to Forest-Cover Loss in the Eastern United States. *BioScience* **2010**, *60*, 286–298. [CrossRef]
23. Strausbaugh, P.; Core, E. *Flora of West Virginia Ed. 2. Part 1*; West Virginia University: Morgantown, VA, USA, 1970.
24. Turner, M.G.; Pearson, S.M.; Bolstad, P.; Wear, D.N. Effects of Land-Cover Change on Spatial Pattern of Forest Communities in the Southern Appalachian Mountains (USA). *Landsc. Ecol.* **2003**, *18*, 449–464. [CrossRef]
25. Bernhardt, E.S.; Palmer, M.A. The Environmental Costs of Mountaintop Mining Valley Fill Operations for Aquatic Ecosystems of the Central Appalachians. *Ann. N. Y. Acad. Sci.* **2011**, *1223*, 39–57. [CrossRef] [PubMed]
26. Brown, M.L.; Canham, C.D.; Murphy, L.; Donovan, T.M. Timber Harvest as the Predominant Disturbance Regime in Northeastern US Forests: Effects of Harvest Intensification. *Ecosphere* **2018**, *9*, e02062. [CrossRef]
27. Cole, P.G.; Weltzin, J.F. Light Limitation Creates Patchy Distribution of an Invasive Grass in Eastern Deciduous Forests. *Biol. Invasions* **2005**, *7*, 477–488. [CrossRef]
28. Flory, S.L.; Clay, K. Invasive Shrub Distribution Varies with Distance to Roads and Stand Age in Eastern Deciduous Forests in Indiana, USA. *Plant Ecol.* **2006**, *184*, 131–141. [CrossRef]
29. Jo, I.; Fridley, J.D.; Frank, D.A. Linking Above-and Belowground Resource Use Strategies for Native and Invasive Species of Temperate Deciduous Forests. *Biol. Invasions* **2015**, *17*, 1545–1554. [CrossRef]
30. Hubbart, J.A.; Guyette, R.; Muzika, R.-M. More than Drought: Precipitation Variance, Excessive Wetness, Pathogens and the Future of the Western Edge of the Eastern Deciduous Forest. *Sci. Total Environ.* **2016**, *566*, 463–467. [CrossRef] [PubMed]
31. Havill, N.P.; Vieira, L.C.; Salom, S.M. *Biology and Control of Hemlock Woolly Adelgid*; FHTET-2014-05; Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team: Morgantown, WV, USA, 2014; pp. 1–21.
32. Vose, J.M.; Wear, D.N.; Mayfield III, A.E.; Nelson, C.D. Hemlock Woolly Adelgid in the Southern Appalachians: Control Strategies, Ecological Impacts, and Potential Management Responses. *For. Ecol. Manag.* **2013**, *291*, 209–219. [CrossRef]
33. Mountaineers and Rangers: A History of Federal Forest Management in the Southern Appalachians, 1900–1981 (Chapter 1). Available online: http://npshistory.com/publications/usfs/region/8/history/chap1.htm (accessed on 30 May 2022).
34. Schuler, T.M.; Gillespie, A.R. Temporal Patterns of Woody Species Diversity in a Central Appalachian Forest from 1856 to 1997. *J. Torrey Bot. Soc.* **2000**, *127*, 149–161. [CrossRef]
35. Thomas-Van Gundy, M.A.; Strager, M.P. *European Settlement-Era Vegetation of the Monongahela National Forest, West Virginia*; General Technical Report NRS-GTR-101; US Department of Agriculture, Forest Service, Northern Research Station: Newtown Square, PA, USA, 2012; Volume 101, pp. 1–39.
36. Mery, G.; Alfaro, R.; Kanninen, M.; Lobovikov, M. Changing Paradigms in Forestry: Repercussions for People and Nature. In *Forests in the Global Balance-Changing Paradigms*; International Union of Forest Research Organizations: Helsinki, Finland, 2005; pp. 13–20.

37. Chiras, D.D.; Reganold, J.P.; Owen, O.S. *Natural Resource Conservation: Management for a Sustainable Future*; Benjamin Cummings: San Francisco, CA, USA, 2010.

38. Lechner, A.M.; Foody, G.M.; Boyd, D.S. Applications in Remote Sensing to Forest Ecology and Management. *One Earth* **2020**, *2*, 405–412. [CrossRef]

39. Wulder, M.A.; Franklin, S.E. *Remote Sensing of Forest Environments: Concepts and Case Studies*; Springer Science & Business Media: Berlin, Germany, 2012.

40. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef]

41. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

42. Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of Studies on Tree Species Classification from Remotely Sensed Data. *Remote Sens. Environ.* **2016**, *186*, 64–87. [CrossRef]

43. Ruefenacht, B.; Finco, M.V.; Nelson, M.D.; Czaplewski, R.; Helmer, E.H.; Blackard, J.A.; Holden, G.R.; Lister, A.J.; Salajanu, D.; Weyermann, D.; et al. Conterminous U.S. and Alaska Forest Type Mapping Using Forest Inventory and Analysis Data. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 1379–1388. [CrossRef]

44. Beaubien, J. Forest Type Mapping from Landsat Digital Data. *Photogramm. Eng. Remote Sens.* **1979**, *45*, 1135–1144.

45. Bryant, E.; Dodge, A.G.; Warren, S.D. Landsat for Practical Forest Type Mapping—A Test Case. *Photogramm. Eng. Remote Sens.* **1980**, *46*, 1575–1584.

46. Homer, C.; Dewitz, J.; Yang, L.; Jin, S.; Danielson, P.; Coulston, J.; Herold, N.; Wickham, J.; Megown, K. Completion of the 2011 National Land Cover Database for the Conterminous United States—Representing a Decade of Land Cover Change Information. *Photogramm. Eng.* **2015**, *81*, 345–354.

47. Fajvan, M.A.; Grushecky, S.T.; Hassler, C.C. The Effects of Harvesting Practices on West Virginia's Wood Supply. *J. For.* **1998**, *96*, 33–39. [CrossRef]

48. Potapov, P.; Hansen, M.C.; Kommareddy, I.; Kommareddy, A.; Turubanova, S.; Pickens, A.; Adusei, B.; Tyukavina, A.; Ying, Q. Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sens.* **2020**, *12*, 426. [CrossRef]

49. Adams, B.; Iverson, L.; Matthews, S.; Peters, M.; Prasad, A.; Hix, D.M. Mapping Forest Composition with Landsat Time Series: An Evaluation of Seasonal Composites and Harmonic Regression. *Remote Sens.* **2020**, *12*, 610. [CrossRef]

50. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest Type Identification with Random Forest Using Sentinel-1A, Sentinel-2A, Multi-Temporal Landsat-8 and DEM Data. *Remote Sens.* **2018**, *10*, 946. [CrossRef]

51. Pasquarella, V.J.; Holden, C.E.; Woodcock, C.E. Improved Mapping of Forest Type Using Spectral-Temporal Landsat Features. *Remote Sens. Environ.* **2018**, *210*, 193–207. [CrossRef]

52. Franklin, S.E. Interpretation and Use of Geomorphometry in Remote Sensing: A Guide and Review of Integrated Applications. *Int. J. Remote Sens.* **2020**, *41*, 7700–7733. [CrossRef]

53. Maxwell, A.E.; Shobe, C.M. Land-Surface Parameters for Spatial Predictive Mapping and Modeling. *Earth-Sci. Rev.* **2022**, *226*, 103944. [CrossRef]

54. Wilson, B.T.; Knight, J.F.; McRoberts, R.E. Harmonic Regression of Landsat Time Series for Modeling Attributes from National Forest Inventory Data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *137*, 29–46. [CrossRef]

55. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]

56. Hościło, A.; Lewandowska, A. Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. *Remote Sens.* **2019**, *11*, 929. [CrossRef]

57. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]

58. Coulter, L.L.; Stow, D.A.; Tsai, Y.-H.; Ibanez, N.; Shih, H.; Kerr, A.; Benza, M.; Weeks, J.R.; Mensah, F. Classification and Assessment of Land Cover and Land Use Change in Southern Ghana Using Dense Stacks of Landsat 7 ETM+ Imagery. *Remote Sens. Environ.* **2016**, *184*, 396–409. [CrossRef]

59. Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. Derivation of a Tasselled Cap Transformation Based on Landsat 8 At-Satellite Reflectance. *Remote Sens. Lett.* **2014**, *5*, 423–431. [CrossRef]

60. Christ, P.F.; Elshaer, M.E.A.; Ettlinger, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; Rempfler, M.; Armbruster, M.; Hofmann, F.; D'Anastasi, M.; et al. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. *arXiv* **2016**, arXiv:1610.02177. [CrossRef]

61. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

62. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]

63. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]

64. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

65. Yang, L.; Jin, S.; Danielson, P.; Homer, C.; Gass, L.; Bender, S.M.; Case, A.; Costello, C.; Dewitz, J.; Fry, J.; et al. A New Generation of the United States National Land Cover Database: Requirements, Research Priorities, Design, and Implementation Strategies. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 108–123. [CrossRef]

66. Vanderhorst, J.; Byers, E.; Streets, B. Natural Heritage Vegetation Database for West Virginia. *Biodivers. Ecol.* **2012**, *4*, 440. [CrossRef]

67. Major Land Resource Area (MLRA) | NRCS Soils. Available online: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053624 (accessed on 28 February 2021).

68. The Perl Programming Language—www.Perl.Org. Available online: https://www.perl.org/ (accessed on 25 June 2022).

69. Kriegler, F.; Malila, W.; Nalepka, R.; Richardson, W. Preprocessing Transformations and Their Effects on Multispectral Recognition. *Remote Sens. Environ.* **1969**, *VI*, 97.

70. Philipp, M. RHarmonics 2021. Available online: https://github.com/MBalthasar/rHarmonics (accessed on 11 June 2022).

71. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

72. Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. The National Elevation Dataset. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–32.

73. Hengl, T.; Reuter, H.I. *Institute for Environment and Sustainability (European Commission. Joint Research Centre) (Eds.) Geomorphometry: Concepts, Software, Applications*, 1st ed.; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands; Oxford, UK; Boston, MA, USA, 2009; ISBN 978-0-12-374345-9.

74. Wilson, J.P.; Gallant, J.C. *Terrain Analysis: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000; ISBN 978-0-471-32188-0.

75. MacMillan, R.A.; Shary, P.A. Chapter 9 Landforms and Landform Elements in Geomorphometry. In *Developments in Soil Science*; Hengl, T., Reuter, H.I., Eds.; Geomorphometry; Elsevier: Amsterdam, The Netherlands, 2009; Volume 33, pp. 227–254.

76. Evans, J.S. GradientMetrics 2022. Available online: https://github.com/jeffreyevans/GradientMetrics (accessed on 11 June 2022).

77. ArcGIS Pro Python Reference—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/arcpy/main/arcgis-pro-arcpy-reference.htm (accessed on 11 June 2022).

78. Download ArcGIS Pro—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/get-started/download-arcgis-pro.htm (accessed on 11 June 2022).

79. Extract Multi Values to Points (Spatial Analyst)—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/extract-multi-values-to-points.htm (accessed on 11 June 2022).

80. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. Available online: https://rdrr.io/cran/raster/ (accessed on 11 June 2022).

81. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.

82. Kuhn, M. Caret: Classification and Regression Training. Available online: https://cran.r-project.org/web/packages/caret/caret.pdf (accessed on 11 June 2022).

83. Hughes, G. On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]

84. Murphy, M.A.; Evans, J.S.; Storfer, A. Quantifying Bufo Boreas Connectivity in Yellowstone National Park with Landscape Genetics. *Ecology* **2010**, *91*, 252–261. [CrossRef]

85. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC press: Boca Raton, FL, USA, 2019.

86. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]

87. Stehman, S.V.; Foody, G.M. Accuracy Assessment. In *The SAGE Handbook of Remote Sensing*; Sage: London, UK, 2009; pp. 297–309.

88. Tharwat, A. Classification Assessment Methods. *Appl. Comput. Inform.* **2020**, *ahead-of-print*. [CrossRef]

89. Stehman, S.V. A Critical Evaluation of the Normalized Error Matrix in Map Accuracy Assessment. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 743–751. [CrossRef]

90. Stehman, S.V. Sampling Designs for Accuracy Assessment of Land Cover. *Int. J. Remote Sens.* **2009**, *30*, 5243–5272. [CrossRef]

91. Stehman, S.V. Impact of Sample Size Allocation When Using Stratified Random Sampling to Estimate Accuracy and Area of Land-Cover Change. *Remote Sens. Lett.* **2012**, *3*, 111–120. [CrossRef]

92. Stehman, S.V. Estimating Area from an Accuracy Assessment Error Matrix. *Remote Sens. Environ.* **2013**, *132*, 202–211. [CrossRef]

93. Foody, G.M. Explaining the Unsuitability of the Kappa Coefficient in the Assessment and Comparison of the Accuracy of Thematic Maps Obtained by Image Classification. *Remote Sens. Environ.* **2020**, *239*, 111630. [CrossRef]

94. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]

95. Shao, G.; Tang, L.; Zhang, H. Introducing Image Classification Efficacies. *IEEE Access* **2021**, *9*, 134809–134816. [CrossRef]

96. Fan, J.; Upadhye, S.; Worster, A. Understanding Receiver Operating Characteristic (ROC) Curves. *CJEM* **2006**, *8*, 19–20. [CrossRef]

97. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef]

98. Wandishin, M.S.; Mullen, S.J. Multiclass ROC Analysis. *Weather. Forecast.* **2009**, *24*, 530–547. [CrossRef]

99. Maxwell, A.E.; Warner, T.A.; Guillén, L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sens.* **2021**, *13*, 2450. [CrossRef]

100. Maxwell, A.E.; Warner, T.A.; Guillén, L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sens.* **2021**, *13*, 2591. [CrossRef]

101. Wei, R.; Wang, J. MultiROC: Calculating and Visualizing ROC and PR Curves across Multi-Class Classifications. 2018. Available online: https://github.com/WandeRum/multiROC (accessed on 11 June 2022).

102. Wright, C.; Gallant, A. Improved Wetland Remote Sensing in Yellowstone National Park Using Classification Trees to Combine TM Imagery and Ancillary Environmental Data. *Remote Sens. Environ.* **2007**, *107*, 582–605. [CrossRef]

103. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 112.

104. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 26.