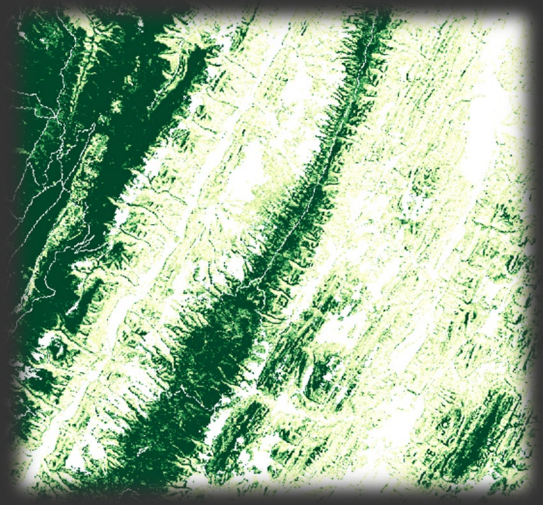
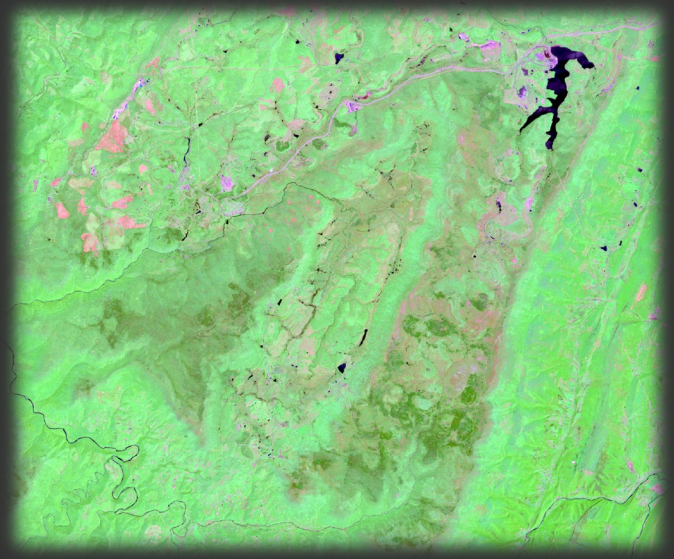
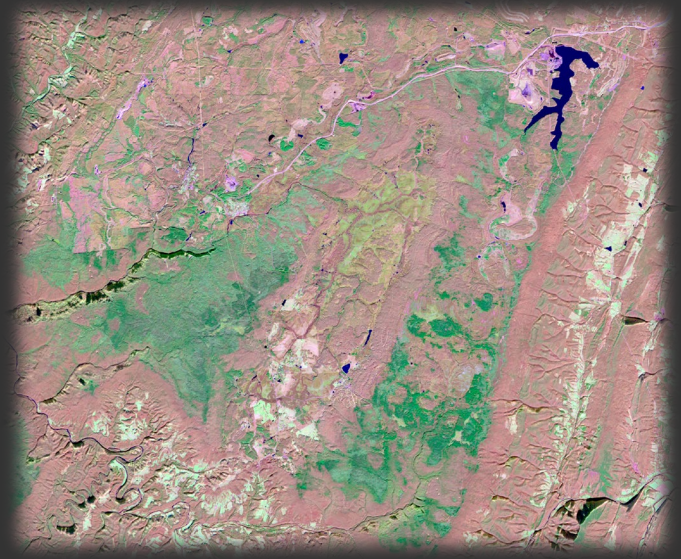
Forest Type Differentiation using Machine Learning, Phenology Metrics, and Land Surface Parameters







PECORA 2022
Technical Session 6-6: Forest Inventory and Condition Mapping
September 27th

Aaron E. Maxwell
West Virginia University
Department of Geology and Geography

Prof Maxwell

- Native of Preston County, West Virginia
- Undergrad at Alderson Broaddus College (Philippi, WV)
  - Biology, Chemistry, Environmental Science
- Graduate work at West Virginia University (Morgantown, WV)
  - Geology MS/PhD
- Assistant Professor WVU Department of Geology and Geography
- Faculty Director for WV GIS Tech Center
- PI for West Virginia View
- Interested in GIS, Remote Sensing, and Spatial Modeling

http://www.wvview.org/Prof_Maxwell.html

Thanks

AmericaView and United State Geological Survey

WV View is supported by AmericaView and the U.S. Geological Survey under Grant/Cooperative Agreement No. G18AP00077. 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.

Thanks

- Faith Hartley
- Rick Landenberger
- Zachary Bortolot

Publication

geographies MDPI

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

check for updates

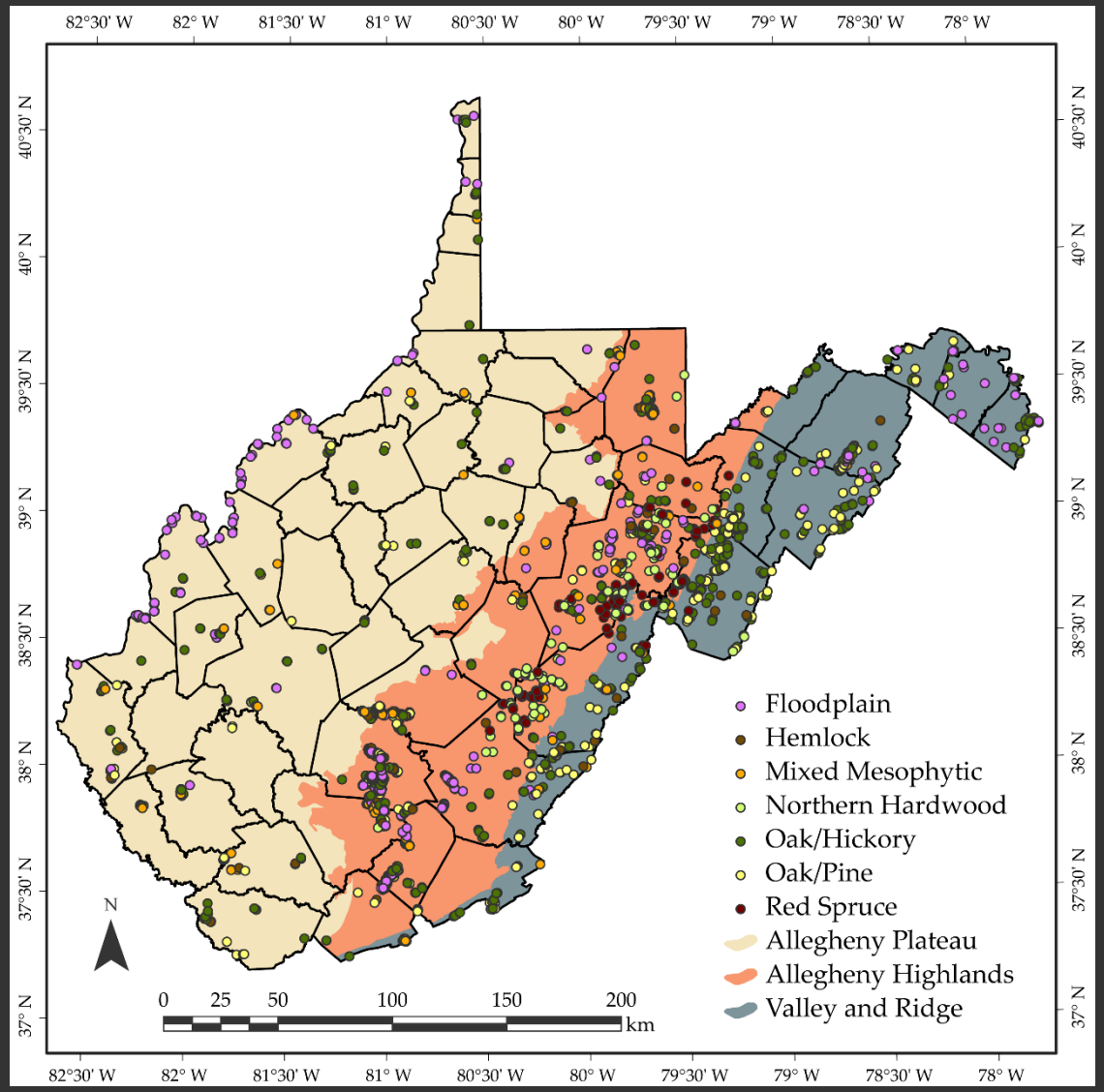
Citation: Hartley, F.M.; Maxwell, A.E.; Landenberger, R.E.; Bortolot, Z.J. Forest Type Differentiation Using GLAD Phenology Metrics, Land Surface Parameters, and Machine Learning. *Geographies* **2022**, *2*, 491–515. https://doi.org/10.3390/geographies2030030

Academic Editor: Luca Salvati

Received: 27 June 2022
Accepted: 9 August 2022
Published: 15 August 2022

Hartley, F.M.; Maxwell, A.E.; Landenberger, R.E.; Bortolot, Z.J. Forest Type Differentiation Using GLAD Phenology Metrics, Land Surface Parameters, and Machine Learning. *Geographies* **2022**, *2*, 491-515. https://doi.org/10.3390/geographies2030030

Goal and Inputs

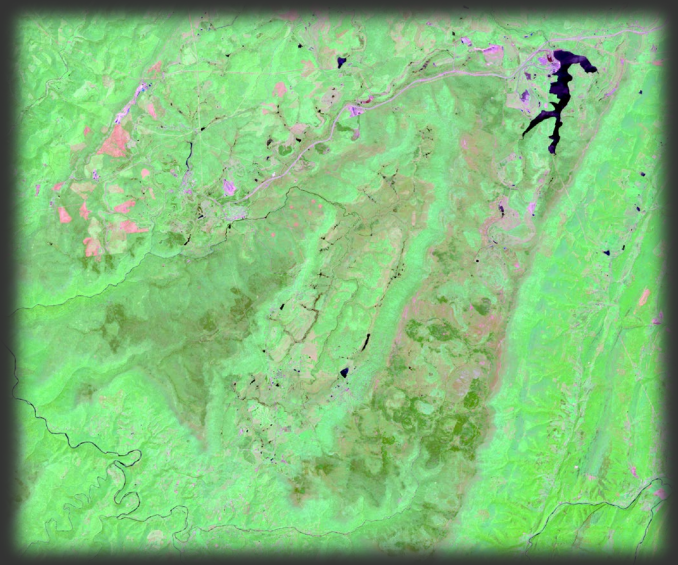


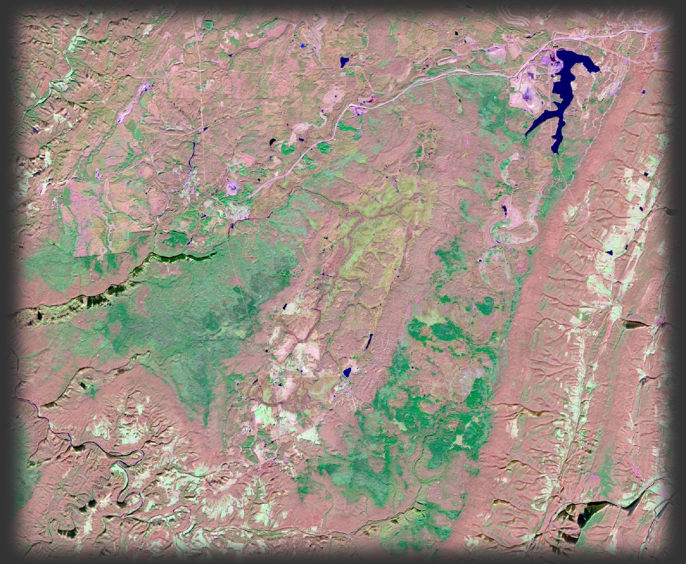
- Assess forest type mapping and probabilistic prediction at the Landsat scale using machine learning, phenology metrics, and land surface parameters
- Samples = WV Natural Heritage Program field plots
- Spectral = GLAD Phenology Metrics
- DEM = 10 m NED

Context

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.





GLAD Phenology

remote sensing MDPI

Article

Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping

Peter Potapov *, Matthew C. Hansen, Indrani Kommareddy, Anil Kommareddy, Svetlana Turubanova, Amy Pickens, Bernard Adusei, Alexandra Tyukavina and Qing Ying

Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA; mhansen@umd.edu (M.C.H.); indrani@umd.edu (I.K.); anilk@umd.edu (A.K.); sveta@umd.edu (S.T.); ahudson2@umd.edu (A.P.); badusei@umd.edu (B.A.); atyukav@umd.edu (A.T.); quing@umd.edu (Q.Y.)
* Correspondence: potapov@umd.edu; Tel.: +1-301-405-3083

Received: 30 December 2019; Accepted: 26 January 2020; Published: 29 January 2020

check for updates

Abstract: The multi-decadal Landsat data record is a unique tool for global land cover and land use change analysis. However, the large volume of the Landsat image archive and inconsistent coverage of clear-sky observations hamper land cover monitoring at large geographic extent. Here, we present a consistently processed and temporally aggregated Landsat Analysis Ready Data produced by the Global Land Analysis and Discovery team at the University of Maryland (GLAD ARD) suitable for national to global empirical land cover mapping and change detection. The GLAD ARD represent a 16-day time-series of tiled Landsat normalized surface reflectance from 1997 to present, updated annually, and designed for land cover monitoring at global to local scales. A set of tools for multi-temporal data processing and characterization using machine learning provided with GLAD ARD serves as an end-to-end solution for Landsat-based natural resource assessment and monitoring. The GLAD ARD data and tools have been implemented at the national, regional, and global extent for water, forest, and crop mapping. The GLAD ARD data and tools are available at the GLAD website for free access.

Keywords: Landsat; analysis ready data; surface reflectance; land surface phenology; image compositing; multi-temporal metrics; land cover; land cover change; time-series analysis; global analysis

Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A. and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, *12*(3), p.426.

https://glad.umd.edu/

GLAD Global Land Analysis & Discovery

GLAD Phenology

ELSEVIER

Remote Sensing of Environment
Volume 253, February 2021, 112165

Mapping global forest canopy height through integration of GEDI and Landsat data

Peter Potapov [a], Xinyuan Li [a], Andres Hernandez-Serna [a], Alexandra Tyukavina [a], Matthew C. Hansen [a], Anil Kommareddy [a], Amy Pickens [a], Svetlana Turubanova [a], Hao Tang [a], Carlos Edibaldo Silva [a], John Armston [a], Ralph Dubayah [a], J. Bryan Blair [b], Michelle Hofton [a]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.rse.2020.112165

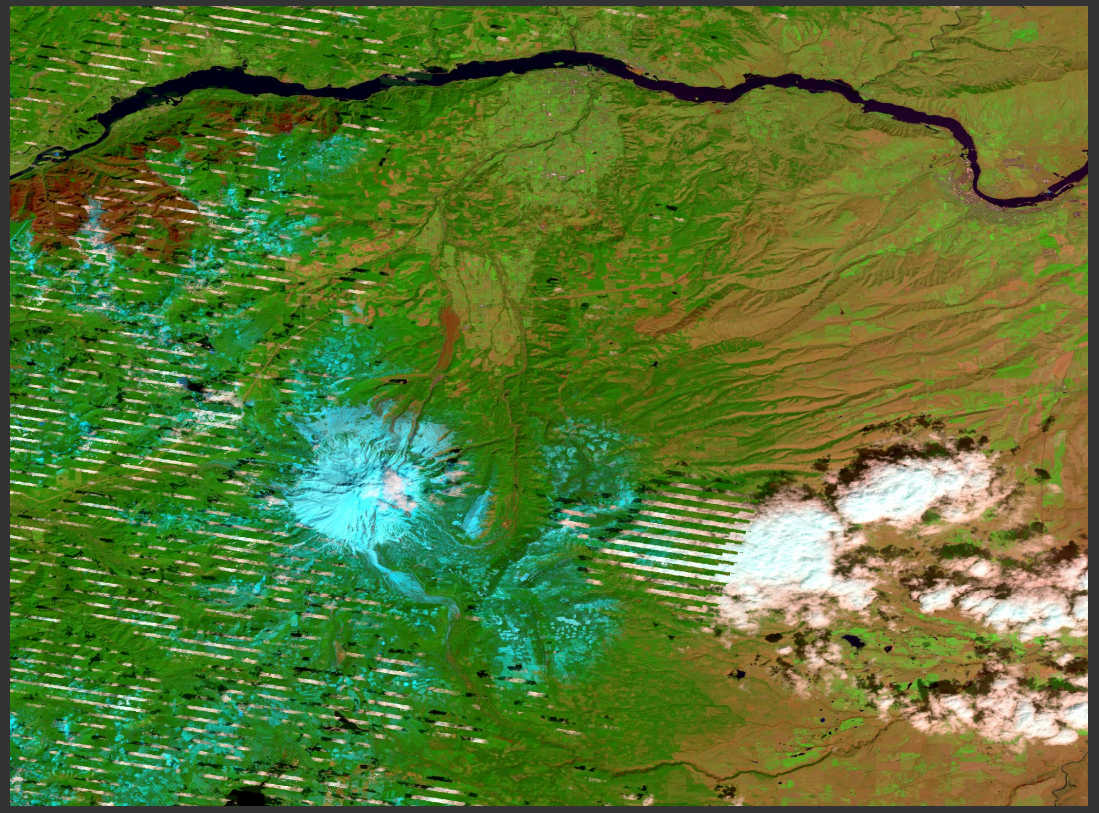
Get rights and content

Abstract

Consistent, large-scale operational monitoring of forest height is essential for estimating forest-related carbon emissions, analyzing forest degradation, and quantifying the effectiveness of forest restoration initiatives. The Global Ecosystem Dynamics Investigation (GEDI) lidar instrument onboard the International Space Station has been collecting unique data on vegetation structure since April 2019. Here, we employed global Landsat analysis-ready data to extrapolate GEDI footprint-level forest canopy height measurements, creating a 30 m spatial resolution global forest canopy height map for the year 2019. The global forest height map was compared to the GEDI validation data (RMSE = 6.6 m; MAE = 4.45 m, R^2 = 0.62) and available airborne lidar data (RMSE = 9.07 m; MAE = 6.36 m, R^2 = 0.61). The demonstrated integration of GEDI data with time-series optical imagery is expected to enable multidecadal historic analysis and operational forward monitoring of forest height and its dynamics. Such capability is important to support global climate and sustainable development initiatives.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E. and Armston, J., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, p.112165.

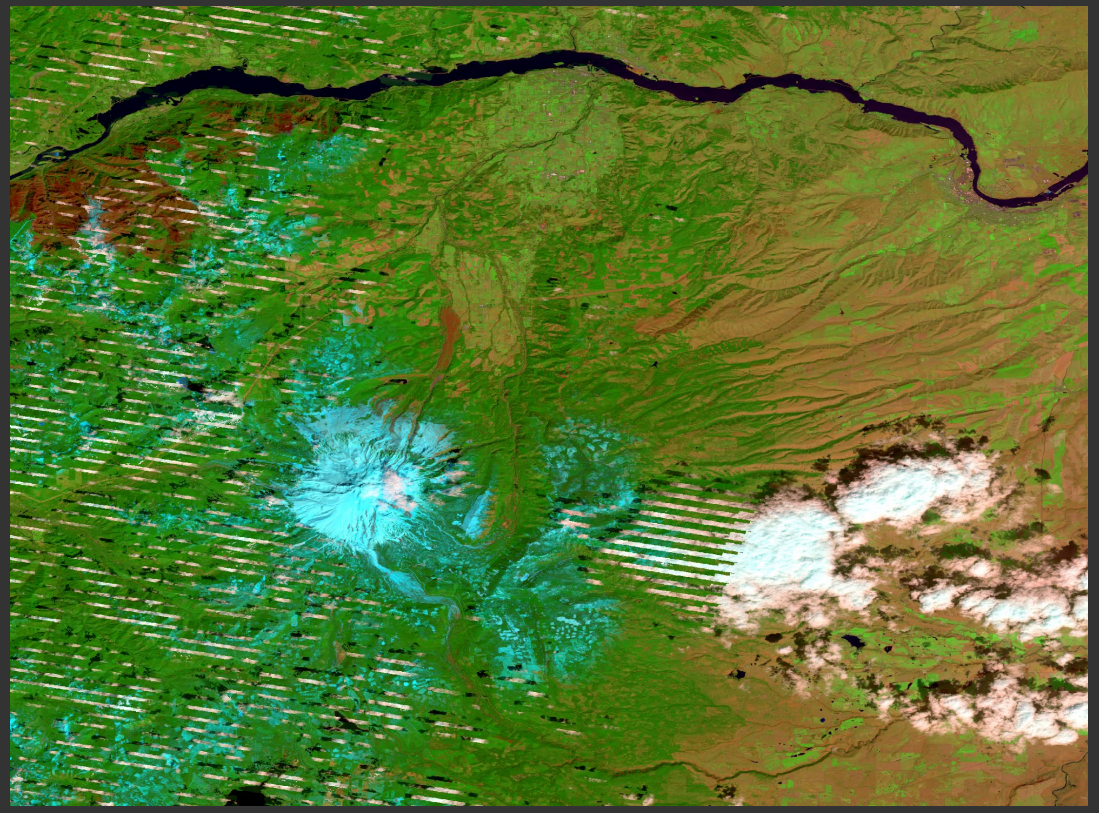
GLAD Phenology



- Downloaded all 16-day composites from beginning of 2015 to end of 2019 using Perl script
- Generated 2019 product, which is gap filled using prior years or adjacent pixels
- Used Perl script to derive full phenology metric set

Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A. and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, *12*(3), p.426.

GLAD Phenology



Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A. and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, *12*(3), p.426.

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)

Land Surface Parameters

Earth-Science Reviews 226 (2022) 103944

Contents lists available at ScienceDirect

Earth-Science Reviews

journal homepage: www.elsevier.com/locate/earscirev

Land-surface parameters for spatial predictive mapping and modeling

Aaron E. Maxwell *, Charles M. Shobe

Department of Geology and Geography, West Virginia University, Morgantown, WV USA

ARTICLE INFO

Keywords:
Geomorphometry
Land-surface parameters
Digital land surface model
Digital elevation model
Landforms
Spatial predictive modeling
machine learning

ABSTRACT

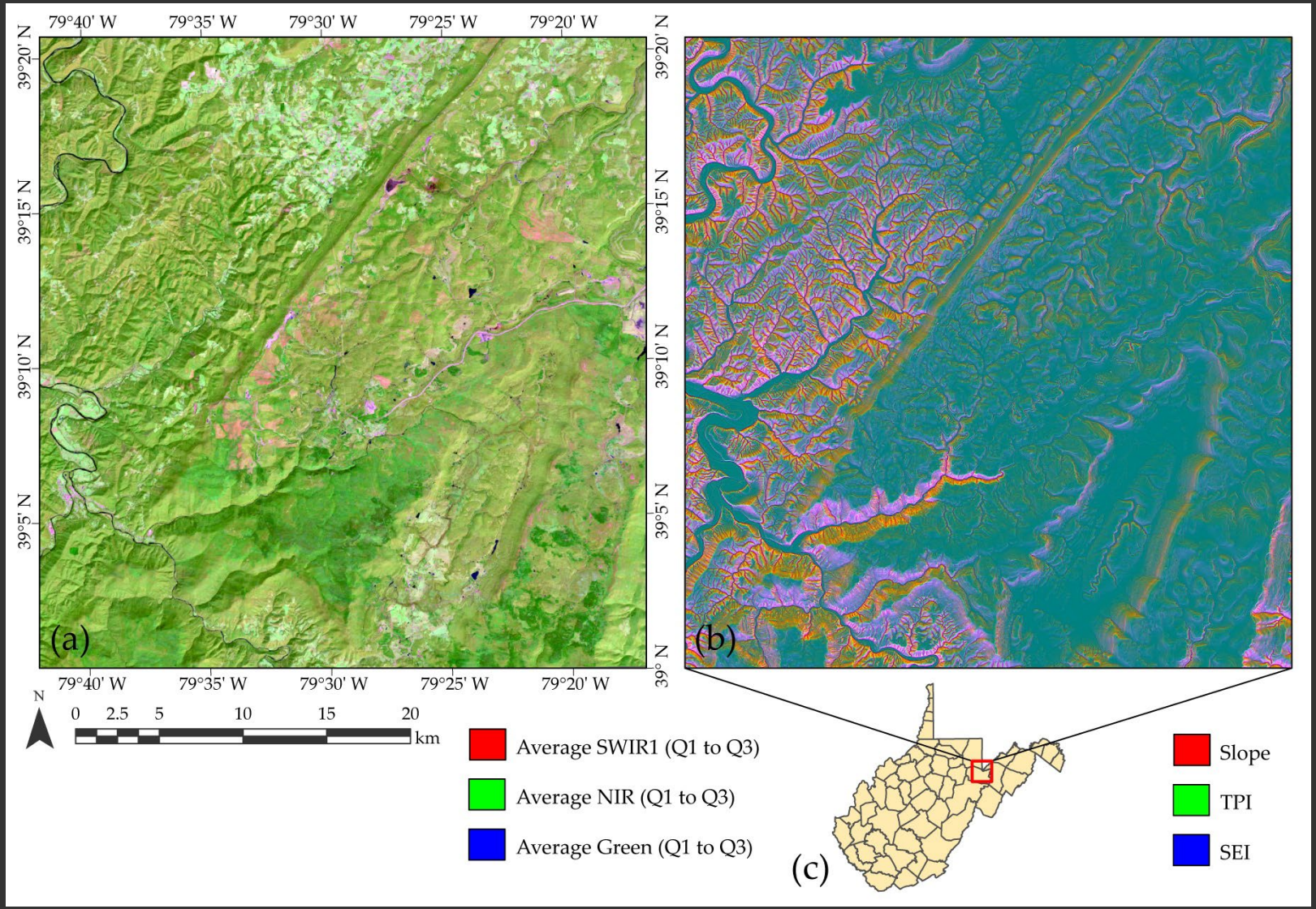
Land-surface parameters derived from digital land surface models (DLSMs) (for example, slope, surface curvature, topographic position, topographic roughness, aspect, heat load index, and topographic moisture index) can serve as key predictor variables in a wide variety of mapping and modeling tasks relating to geomorphic processes, landform delineation, ecological and habitat characterization, and geohazard, soil, wetland, and general thematic mapping and modeling. However, selecting features from the large number of potential derivatives that may be predictive for a specific feature or process can be complicated, and existing literature may offer contradictory or incomplete guidance. The availability of multiple data sources and the need to define moving window shapes, sizes, and cell weightings further complicate selecting and optimizing the feature space. This review focuses on the calculation and use of DLSM parameters for empirical spatial predictive modeling applications, which rely on training data and explanatory variables to make predictions of landscape features and processes over a defined geographic extent. The target audience for this review is researchers and analysts undertaking predictive modeling tasks that make use of the most widely used terrain variables.

To outline best practices and highlight future research needs, we review a range of land-surface parameters relating to steepness, local relief, rugosity, slope orientation, solar insolation, and moisture and characterize their relationship to geomorphic processes. We then discuss important considerations when selecting such parameters for predictive mapping and modeling tasks to assist analysts in answering two critical questions: What landscape conditions or processes does a given measure characterize? How might a particular metric relate to the phenomenon or features being mapped, modeled, or studied? We recommend the use of landscape- and problem-specific pilot studies to answer, to the extent possible, these questions for potential features of interest in a mapping or modeling task. We describe existing techniques to reduce the size of the feature space using feature selection and feature reduction methods, assess the importance or contribution of specific metrics, and parameterize moving windows or characterize the landscape at varying scales using alternative methods while highlighting strengths, drawbacks, and knowledge gaps for specific techniques. Recent developments, such as explainable machine learning and convolutional neural network (CNN)-based deep learning, may guide and/or minimize the need for feature space engineering and ease the use of DLSMs in predictive modeling tasks.

Maxwell, A.E. and Shobe, C.M., 2022. Land-surface parameters for spatial predictive mapping and modeling. *Earth-Science Reviews*, p.103944.

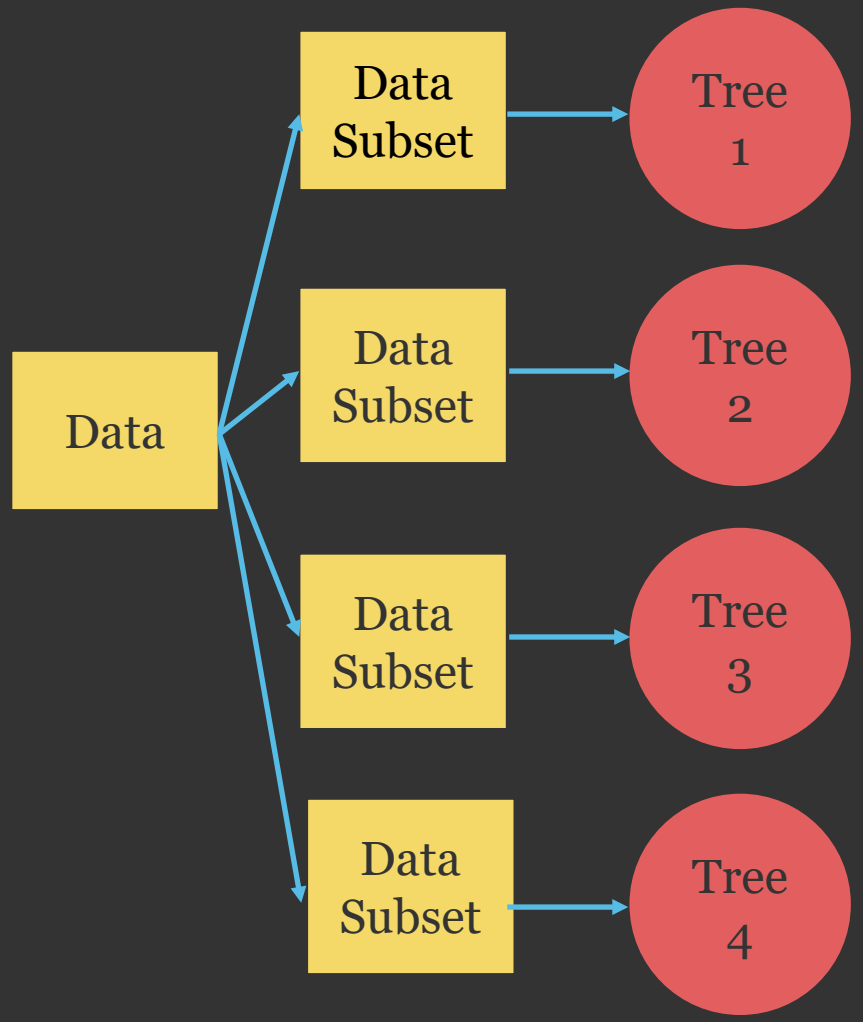
Variable	Abbreviation	Description/Equation
Linear Aspect	AspLn	$270 - \frac{360}{2\pi} \times \arctan^2\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$
Cosine Aspect Transformation	AspCos	Cos(Aspect); measure of eastwardness
Sine Aspect Transformation	AspSin	Sin(Aspect); measure 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	$\sigma^2(z)$
Surface Area Ratio	SAR	$\frac{\text{Cell Size}^2}{\text{Cos(Slope in Degrees)}}$
Surface Relief Ratio	SRR	$\frac{z_{mean} - z_{min}}{z_{max} - z_{min}}$
Heat Load Index	HLI	Index for annual direct incoming solar radiation based on latitude, slope, and aspect
Site Exposure Index	SEI	$\text{Slope} \times \cos\left(\pi \frac{\text{Aspect} - 180}{180}\right)$

Feature Space



Random Forests (RF)

- Uses decision trees
- Uses the Gini Index of Impurity
- Ensemble decision tree method
- Uses random subset of predictor variables for splitting at each node
- Uses random subset of training data in each tree
- Attempts to reduce correlation between trees
- Ensemble of weak classifiers



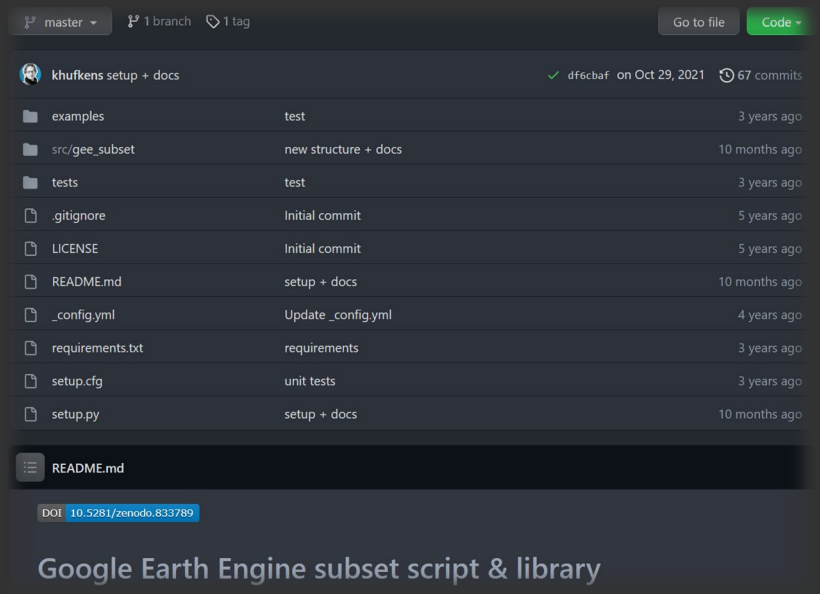
Take the majority vote

Subset of predictors used in each tree

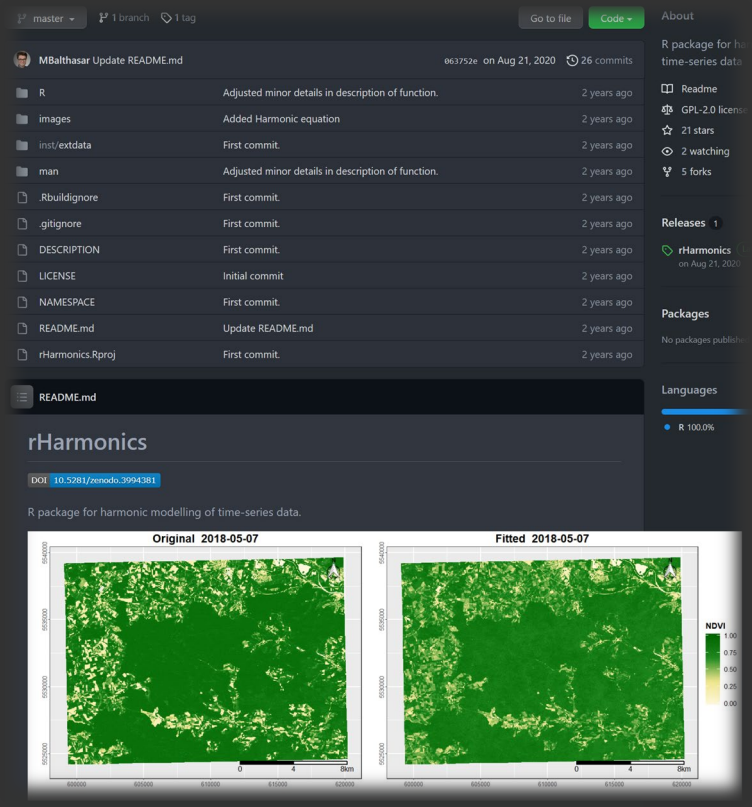
Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.

Comparisons

- Summer Band Medians
- Fall Band Medians
- Spring Band Medians
- Harmonic Regression Coefficients



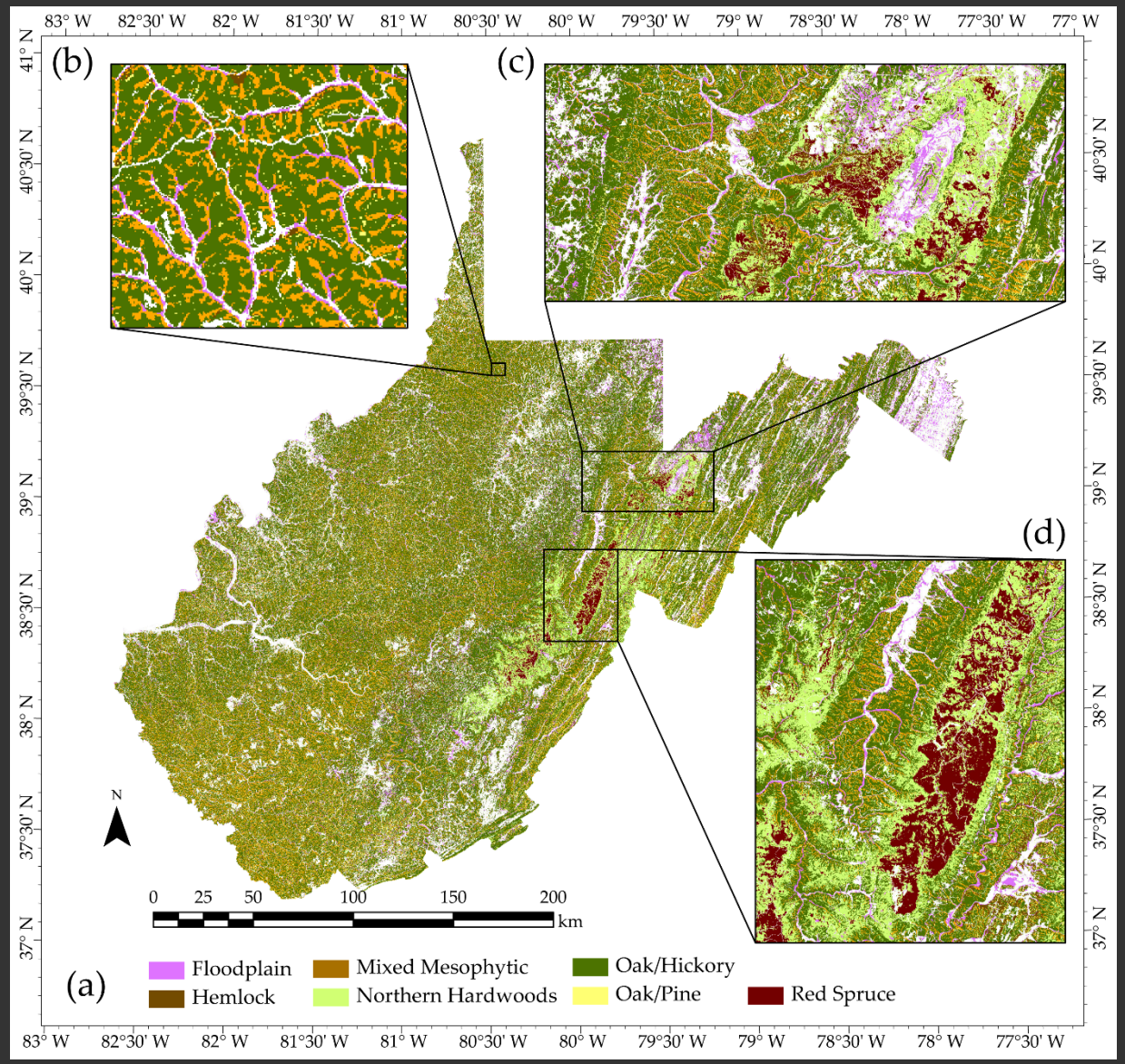
https://github.com/bluegreen-labs/gee_subset

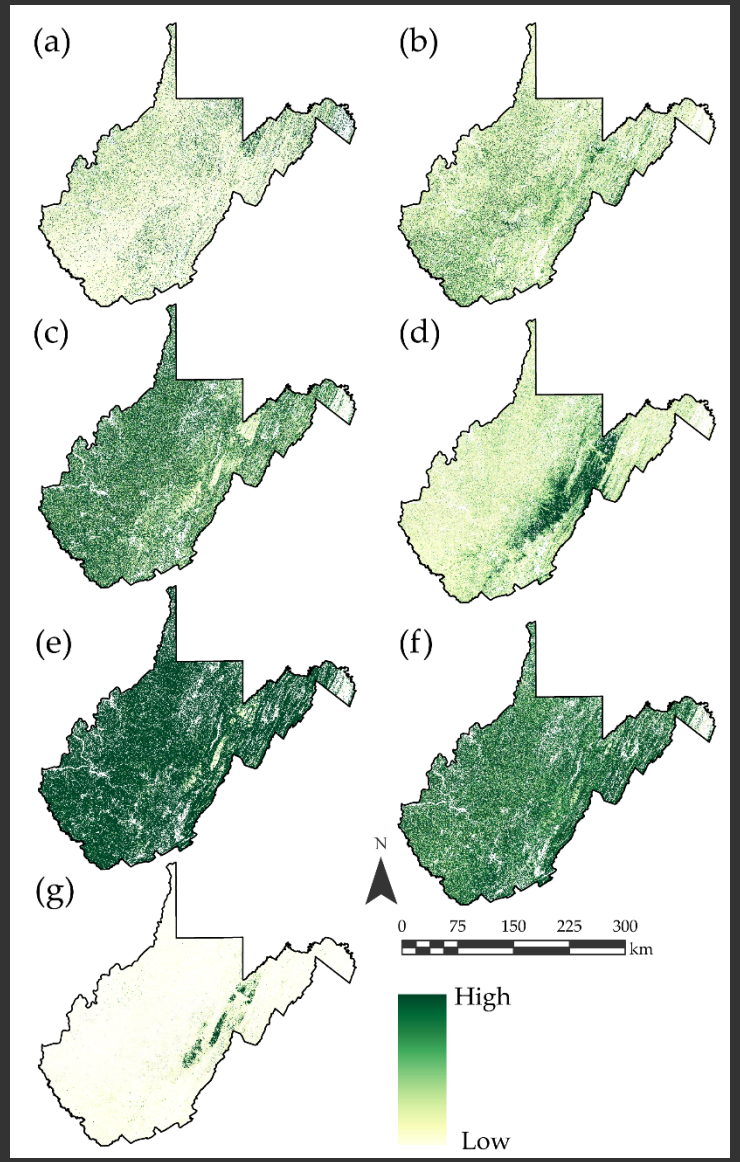


https://github.com/MBalthasar/rHarmonics/

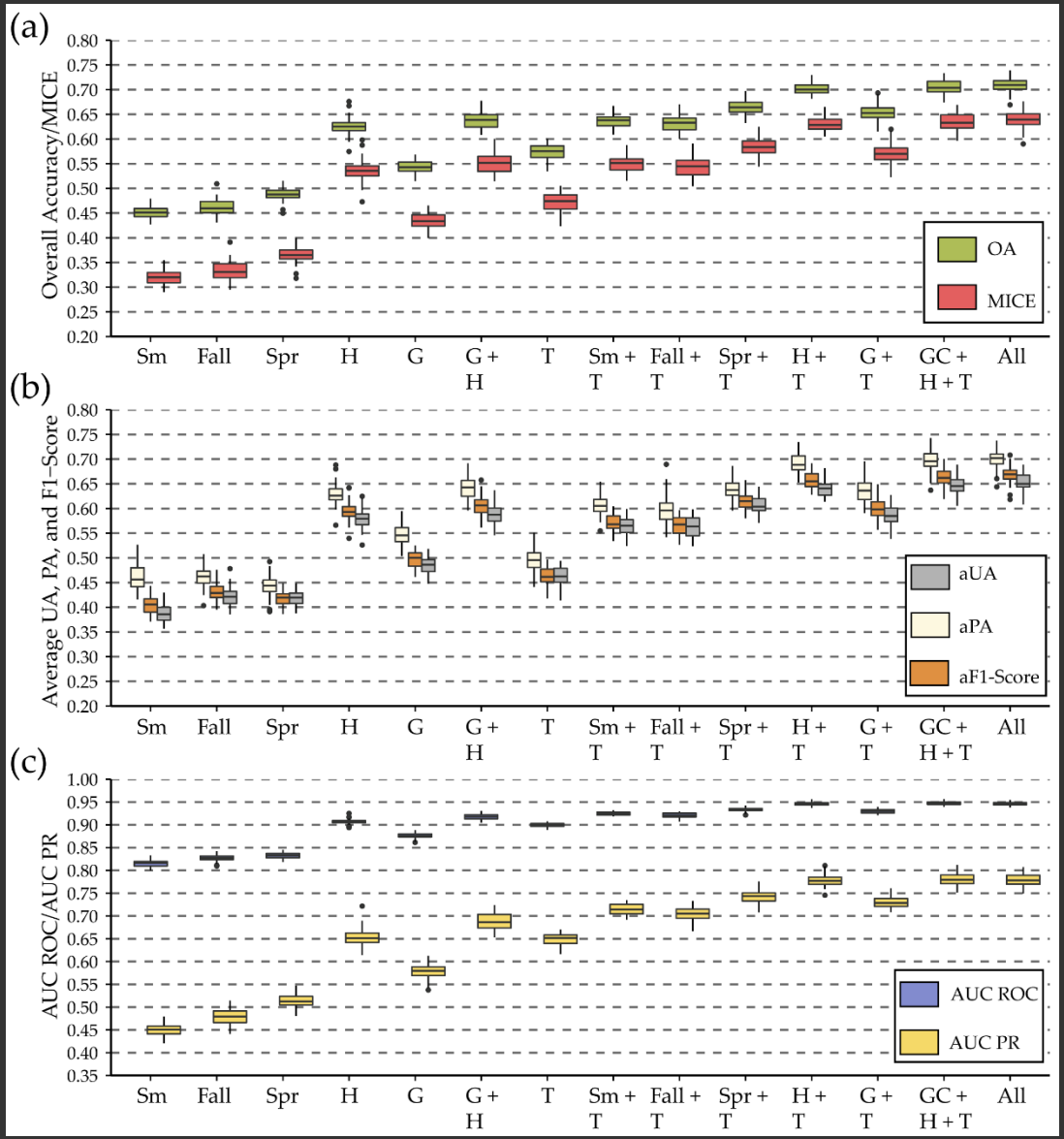
Feature Set	Abbreviation	Number of Variables
GLAD Phenology Type C	G	188
Digital Terrain Variables	T	17
Harmonic Regression Coefficients	H	32
Summer	Sm	10
Fall	Fall	10
Spring	Spr	10

Results


(a)
(b)
(c)
(d)
83° W
82°30' W
82° W
81°30' W
81° W
80°30' W
80° W
79°30' W
79° W
78°30' W
78° W
77°30' W
77° W
41° N
40°30' N
40° N
39°30' N
39° N
38°30' N
38° N
37°30' N
37° N
N
0 25 50 100 150 200
km
Floodplain
Hemlock
Mixed Mesophytic
Northern Hardwoods
Oak/Hickory
Oak/Pine
Red Spruce


(a)
(b)
(c)
(d)
(e)
(f)
(g)
N
0 75 150 225 300
km
High
Low

Results



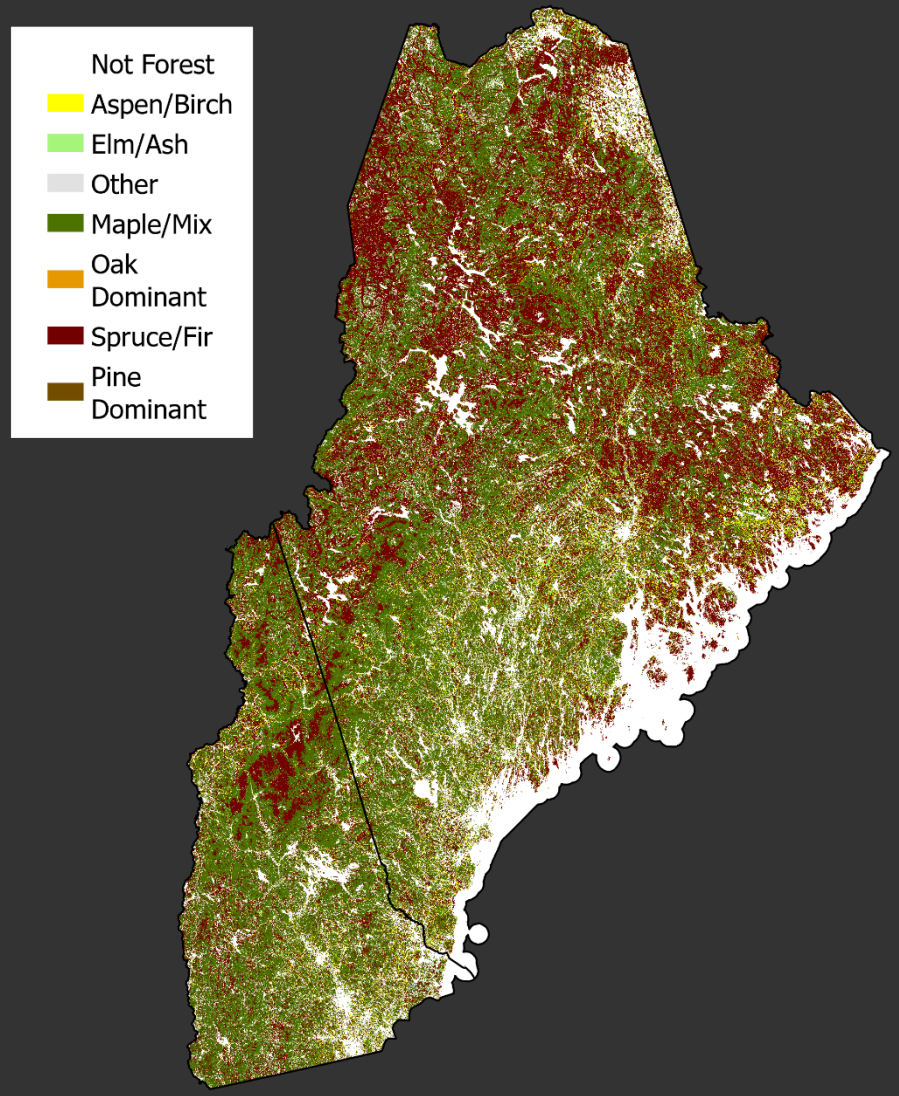
Set	Number of Classes	OA	MICE	Top 3	aUA	aPA	aFS	AUC ROC	AUC PR
G	7	0.543	0.433	0.886	0.484	0.547	0.497	0.875	0.579
G + T	7	0.653	0.570	0.938	0.587	0.637	0.601	0.930	0.730
G	6	0.648	0.496	0.933	0.501	0.601	0.527	0.906	0.698
G + T	6	0.762	0.660	0.966	0.615	0.673	0.631	0.953	0.837

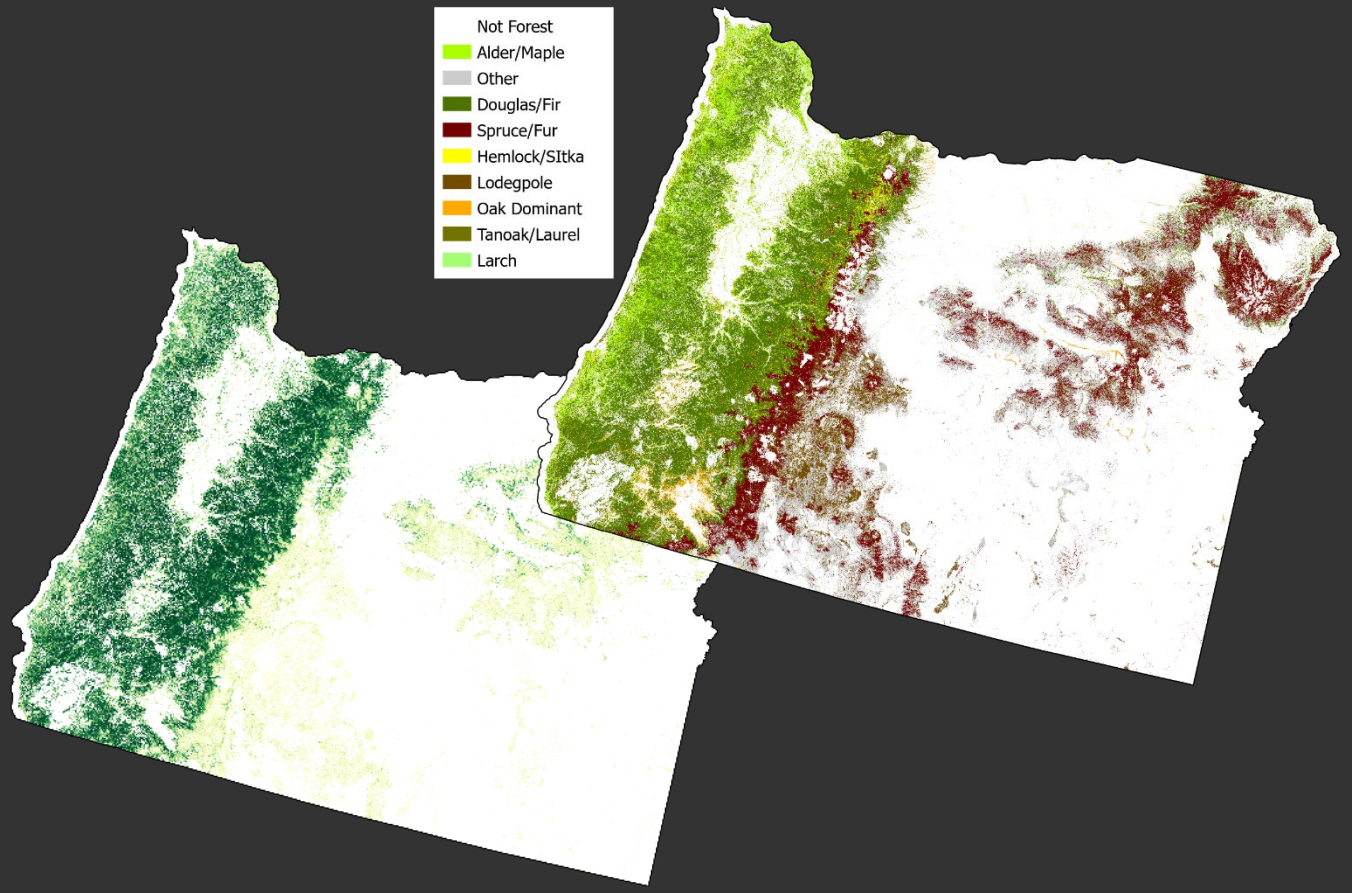
		Reference								
		Floodplain	Hemlock	Mix. Meso	North. Hard.	Oak/Hick.	Oak/Pine	Red Spruce	Totals	UA
Prediction	Floodplain	137	3	1	1	4	7	1	154	0.890
	Hemlock	2	15	3	1	2	5	1	29	0.517
	Mix. Meso.	0	17	40	5	18	2	0	82	0.488
	North. Hard.	0	1	2	16	4	0	6	29	0.552
	Oak/Hick.	1	7	28	15	140	51	0	242	0.579
	Oak/Pine	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		

Key Findings

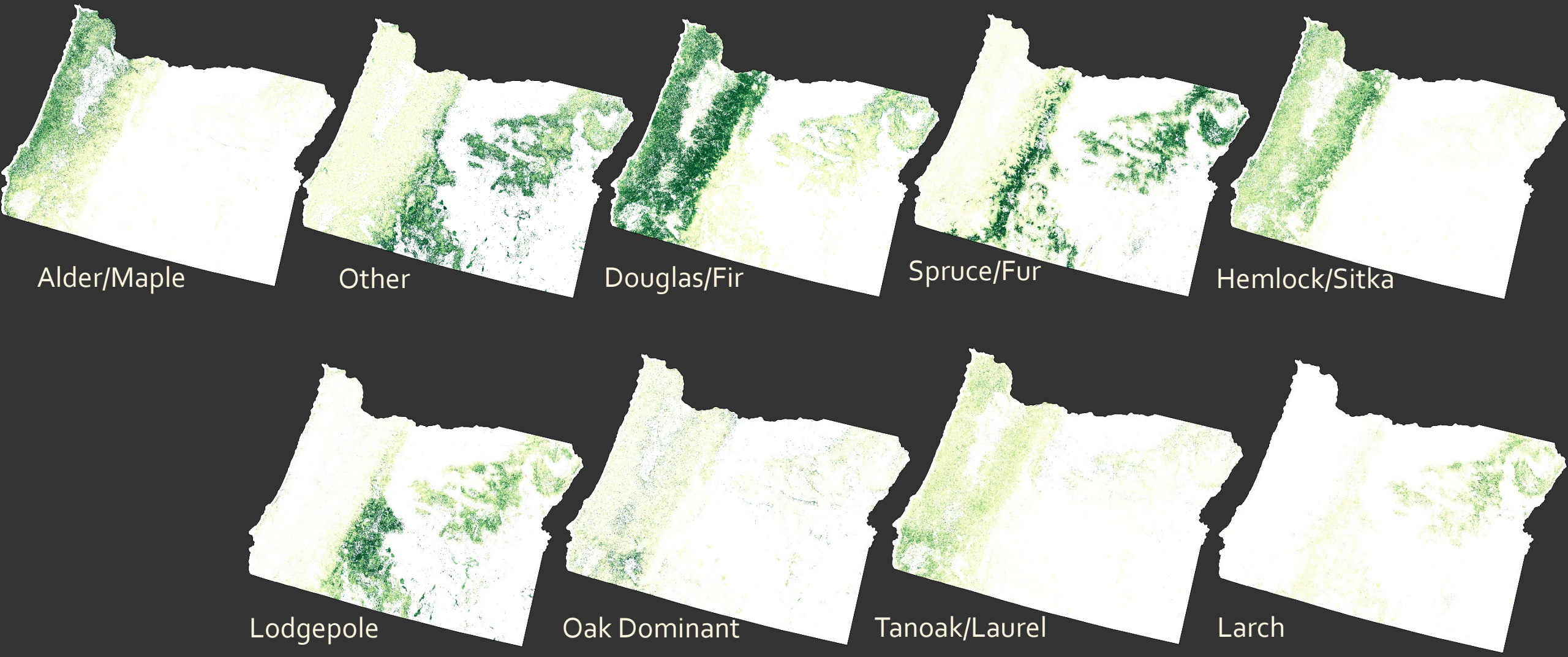
- Including digital terrain variables are valuable
- The number of classes and class definitions can have a large impact on the accuracy
- Highlights the value of supplementing "hard" classification products with associated probabilistic predictions
- GLAD Phenology Metrics were generally of value; however, they did not provide the level of accuracy obtained using harmonic regression coefficients

Moving Forward



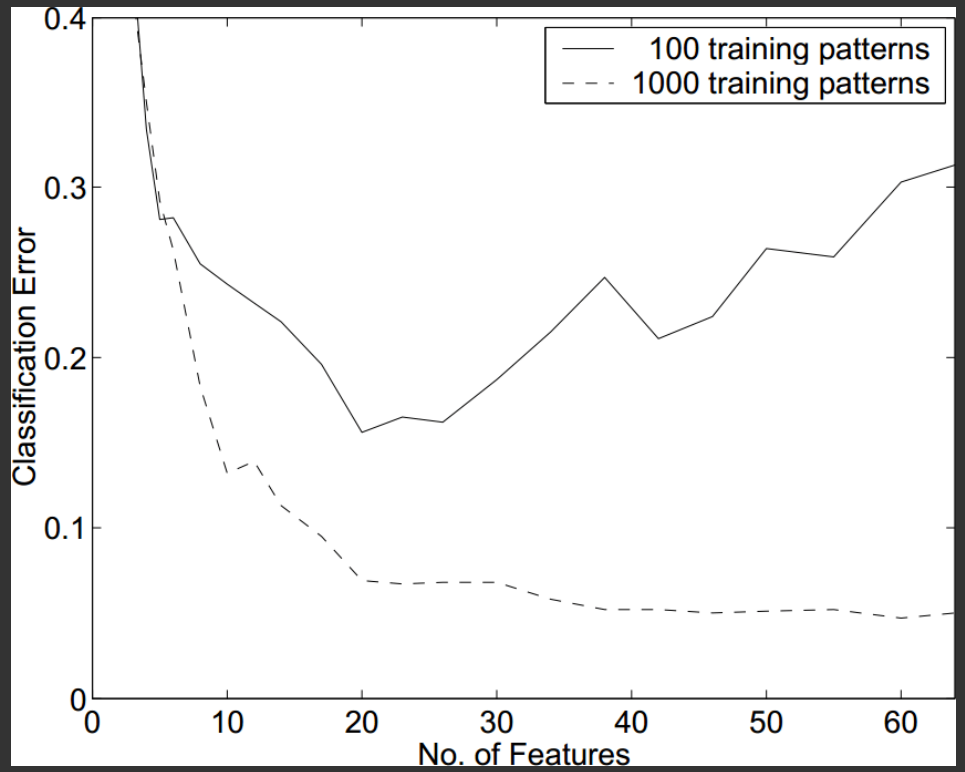


Moving Forward



Lessons Learned

Developing Predictor Variables



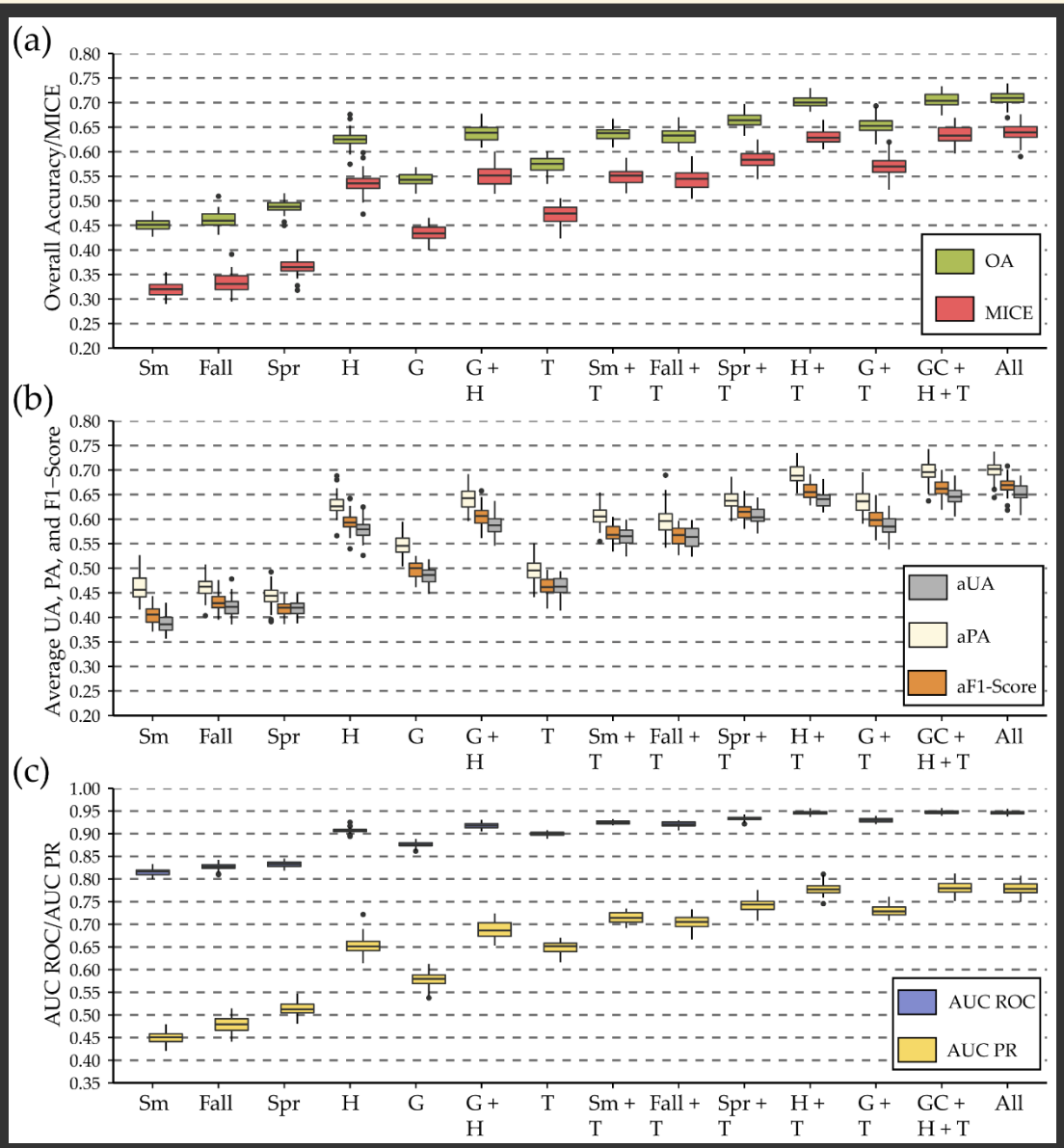
Jain, Anil K., Robert P. W. Duin, and Jianchang Mao. "Statistical pattern recognition: A review." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.1 (2000): 4-37.

Hughes Phenomenon

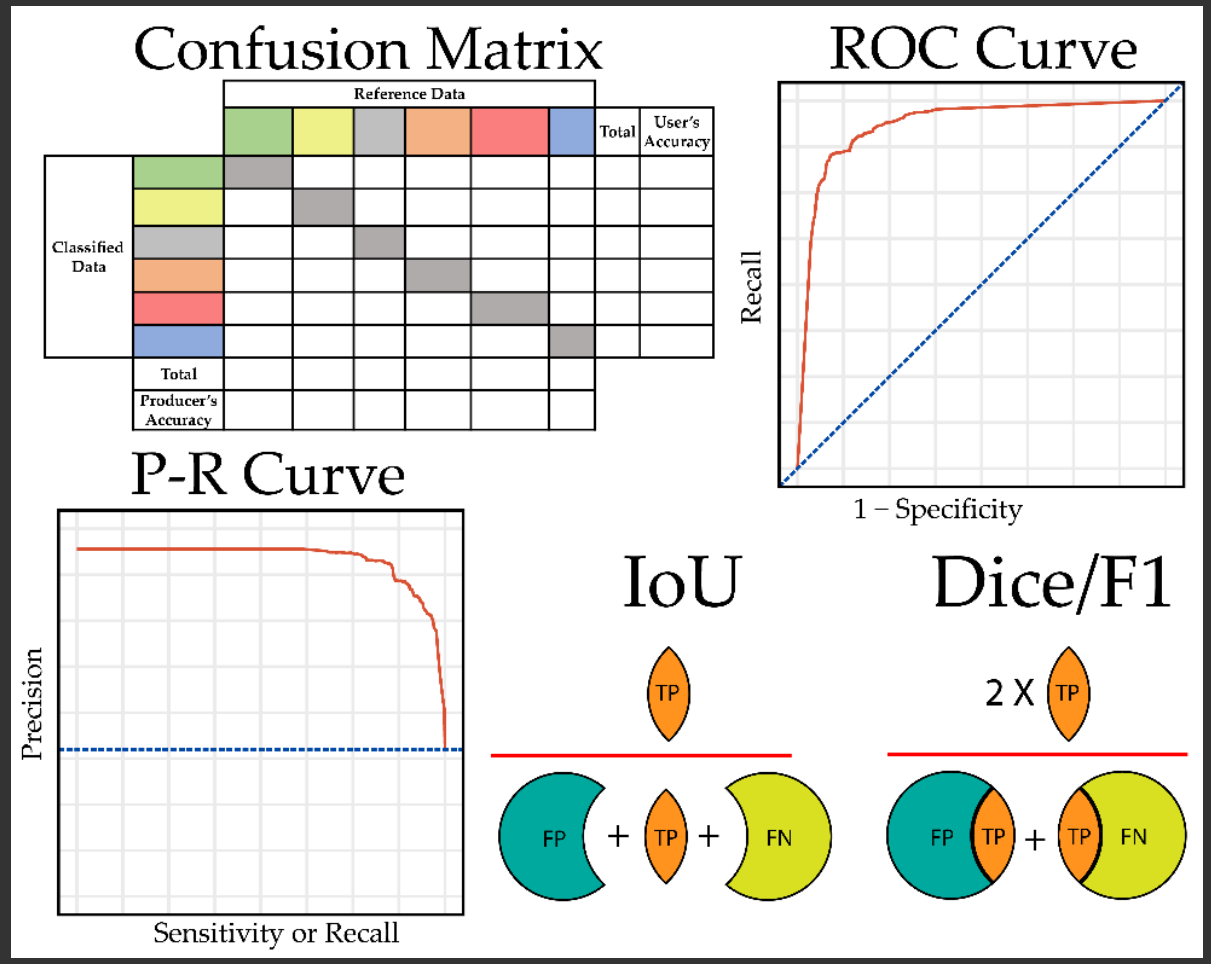
Hughes, G.F. 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory* 14 (1): 55–63. doi:10.1109/TIT.1968.1054102.

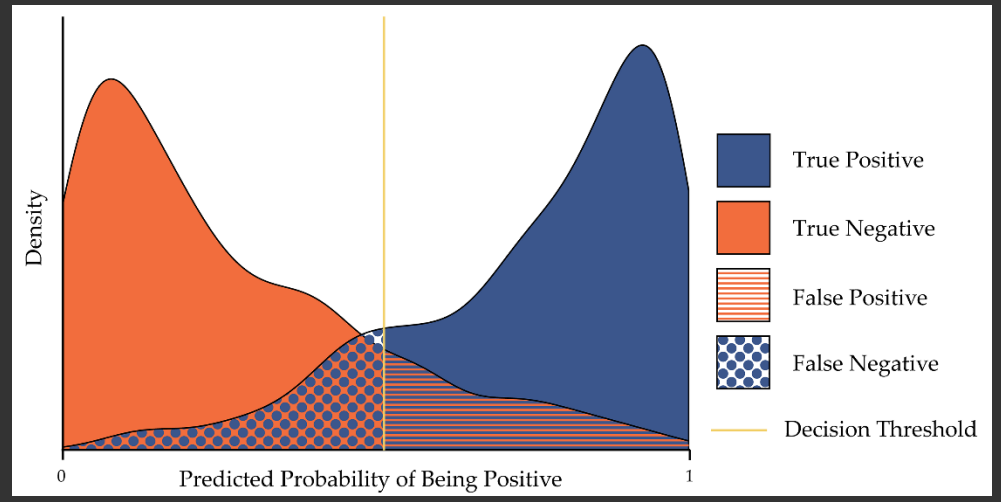
"Curse of dimensionality"

Model Variability

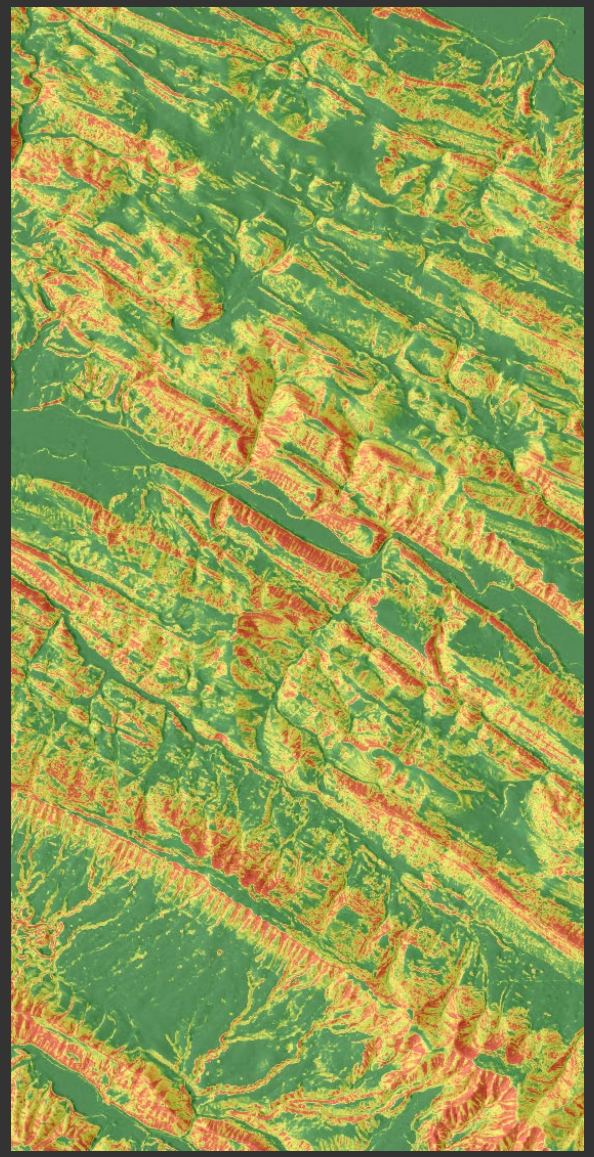


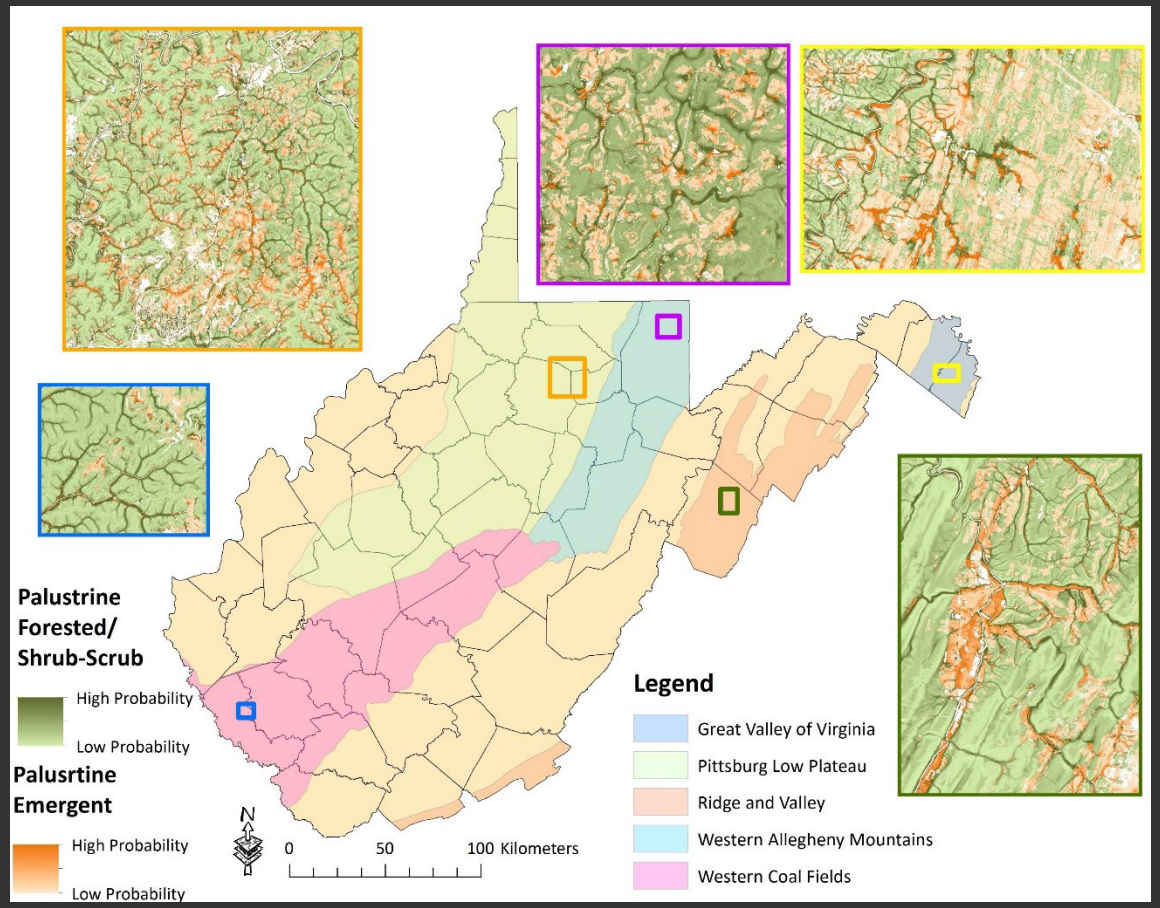
Accuracy Assessment





Difficult Mapping/Modeling Problems





Thanks

Aaron.Maxwell@mail.wvu.edu

GitHub

https://github.com/maxwell-geospatial

YouTube

https://www.youtube.com/channel/UCpS00ZG5FtMyKwGnHaGnvrA/videos

WV VIEW

https://www.wvview.org/index.html