Supplemental Index for

Classifying Rocky Land Cover Using Random Forest Modeling: Lessons Learned and Potential Applications in Washington, USA

Joe V. Celebrezze, Okikiola M. Alegbeleye, Doug A. Glavich, Lisa A. Shipley, and Arjan J. H. Meddens

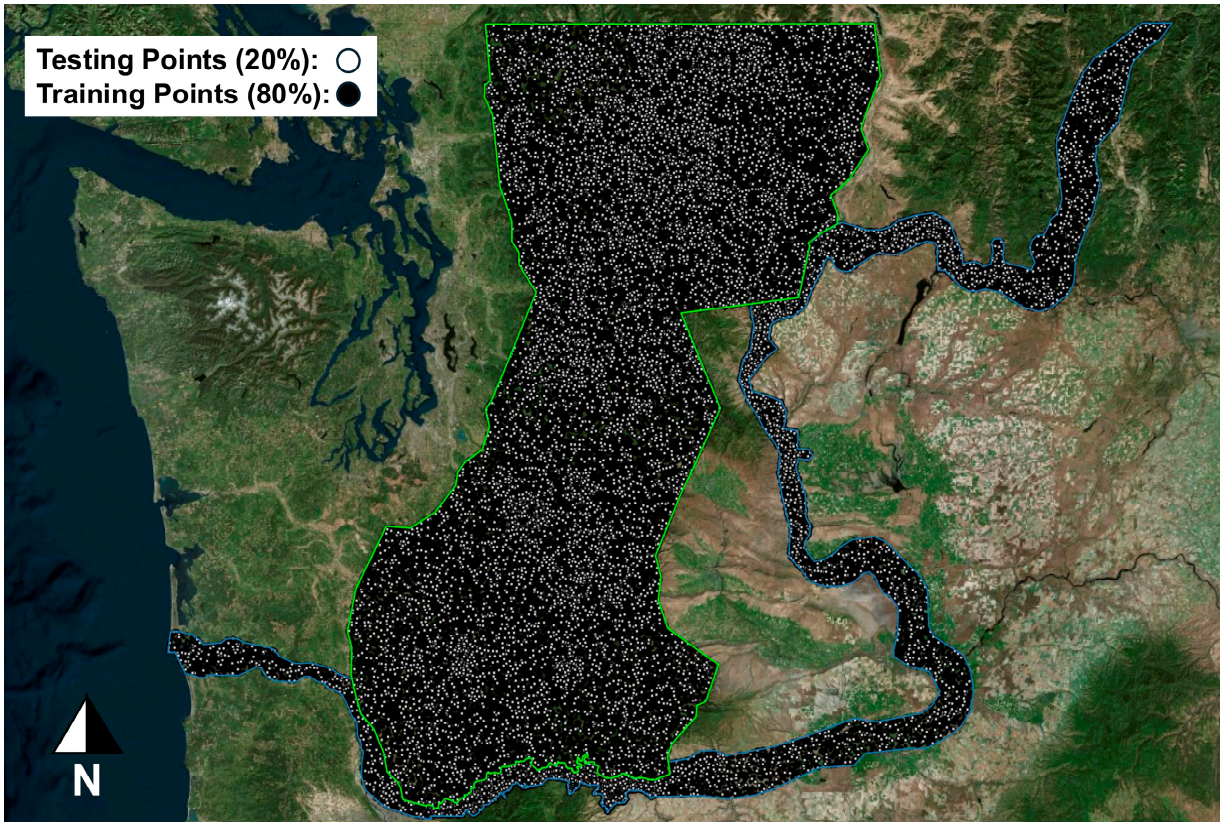

Figure S1: Training (80%) and testing (20%) points, randomly sampled stratified by land cover type, are evenly distributed across the landscape for Cascades and Columbia regions

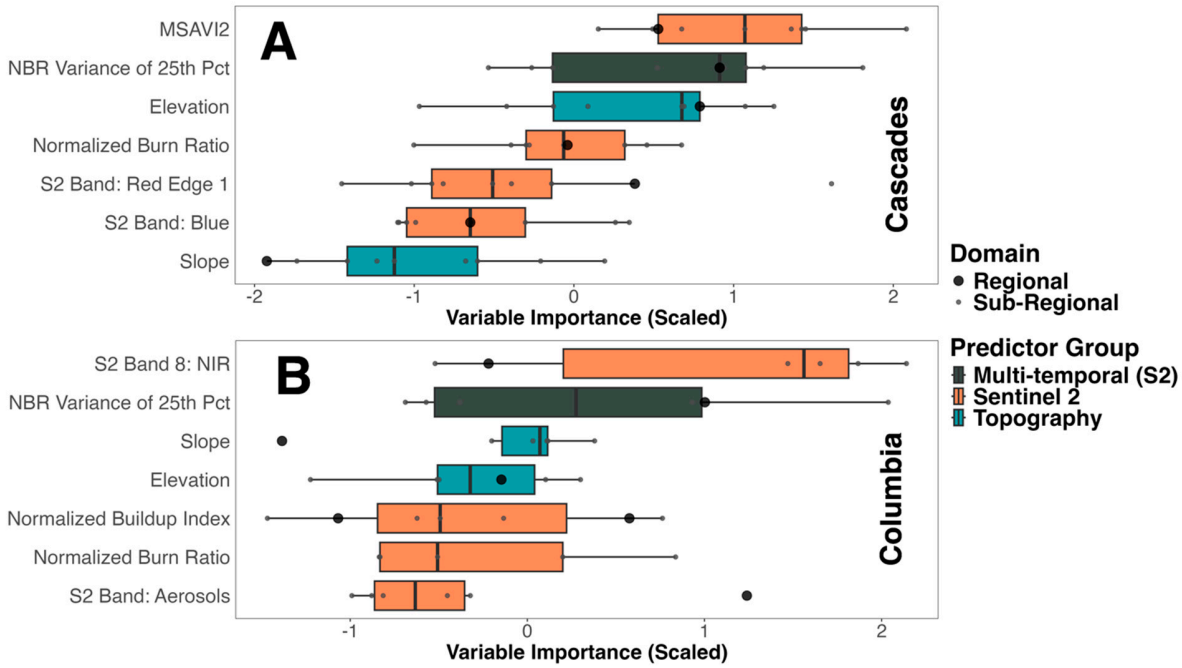

Figure S2. Variable importance boxplots for (A) Cascades and (B) Columbia regions showing the scaled variable importance for the optimized models with boxes colored by predictor groups. Large black points represent scaled variable importance for regional models, while smaller gray points represent sub-regional models

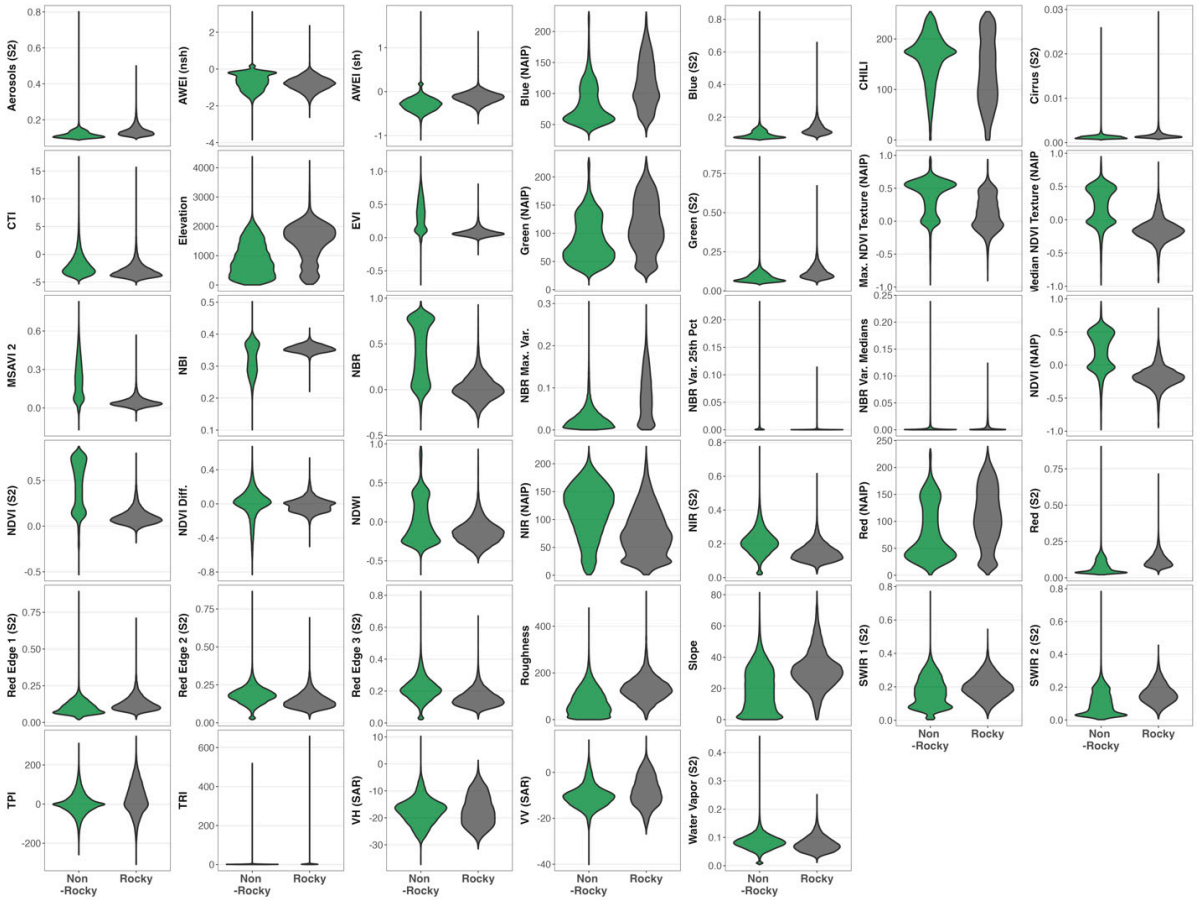

Figure S3. Violin plots displaying predictor values and their distributions across all points for both regions, split between rocky and non-rocky classifications showing all 40 predictors

Table S1. Accuracy statistics using *testing data* for full (all 40 predictors; see Table 2) and optimized (7 predictors, see Figures 5 & 6) models for sub-regions (italicized) and regions showing overall accuracy and user's and producer's accuracies and the F score for classifying non-rocky (NR) and rocky (R) land cover. For the *Portland* sub-region, there were too few rocky areas to classify the area; however, it was included in the Columbia region.

Sub-Region / Region	Full Model							Optimized Model						
	Overall Accuracy (%)	User's Accuracy (%)		Producer's Accuracy (%)		F Score		Overall Accuracy (%)	User's Accuracy (%)		Producer's Accuracy (%)		F Score	
		NR	R	NR	R	NR	R		NR	R	NR	R	NR	R
N. Cascades	95.7	96.5	94.9	94.8	96.6	0.956	0.957	95.6	97.0	94.2	94.0	97.1	0.955	0.956
Winthrop	97.7	98.0	96.6	98.9	94.0	0.985	0.953	96.9	97.4	95.5	98.6	92.1	0.980	0.938
Forbidden	98.1	98.4	97.6	98.8	96.7	0.986	0.976	98.11	98.5	97.3	98.6	97.1	0.986	0.972
Baker	98.0	97.8	98.4	99.2	95.8	0.985	0.971	97.32	97.6	96.8	98.3	95.4	0.980	0.961
Snoqualmie	98.0	98.8	96.8	98.2	97.8	0.985	0.973	96.73	98.0	94.6	96.9	96.4	0.974	0.955
Rainier	98.4	99.2	96.1	98.6	97.8	0.989	0.970	98.11	99.1	95.2	98.3	97.5	0.987	0.964
Adams	97.5	98.5	94.4	98.3	95.1	0.984	0.948	97.33	98.2	94.7	98.4	94.1	0.983	0.944
St. Helens	97.7	98.2	94.2	99.3	87.0	0.987	0.905	97.55	98.0	94.1	99.3	85.5	0.986	0.896
Cascades	97.9	98.7	96.1	98.3	96.9	0.985	0.965	97.36	98.2	95.4	98.0	95.9	0.981	0.954
Colville	97.0	96.8	100	100	64.5	0.984	0.784	94.57	95.5	89.0	98.2	76.1	0.968	0.820
Chelan	95.3	95.1	96.5	99.2	81.6	0.971	0.884	92.24	93.3	87.6	97.0	75.2	0.951	0.809
Quincy	93.6	96.0	87.0	95.3	88.8	0.957	0.879	93.08	94.7	88.3	96.1	84.6	0.954	0.864
Tri-Cities	97.5	98.1	91.7	99.2	82.5	0.986	0.868	98.05	99.0	87.0	99.0	87.0	0.990	0.870
CRGNSA	96.5	96.4	97.2	99.4	85.2	0.979	0.908	95.69	96.1	93.7	98.6	84.4	0.973	0.888
Columbia	95.2	96.0	91.3	98.2	82.1	0.971	0.864	93.9	95.5	86.1	97.2	79.2	0.963	0.825

Table S2. Accuracy statistics using *training data* for full (all 40 predictors; see Table 2) and optimized (7 predictors, see Figures 5 & 6) models for sub-regions (italicized) and regions showing overall accuracy and user's and producer's accuracies and the F score for classifying non-rocky (NR) and rocky (R) land cover. For the *Portland* sub-region, there were too few rocky areas to classify the area; however, it was included in the Columbia region.

Sub-Region / Region	Full Model							Optimized Model						
	Overall Accuracy (%)	User's Accuracy (%)		Producer's Accuracy (%)		F Score		Overall Accuracy (%)	User's Accuracy (%)		Producer's Accuracy (%)		F Score	
		NR	R	NR	R	NR	R		NR	R	NR	R	NR	R
N. Cascades	99.9	99.9	99.8	99.8	99.9	0.999	0.999	99.8	99.9	997	99.6	99.9	0.998	0.998
Winthrop	100	100	99.9	100	100	1.0	1.0	100	100	99.9	100	99.9	1.0	0.999
Forbidden	100	100	99.9	99.9	100	1.0	0.999	100	100	100	100	100	1.0	1.0
Baker	100	100	100	100	100	1.0	1.0	100	100	100	100	99.9	1.0	1.0
Snoqualmie	100	100	100	100	99.9	1.0	0.999	99.9	99.9	100	100	99.7	0.999	0.999
Rainier	100	100	99.9	100	99.9	1.0	0.999	100	100	99.9	100	99.9	1.0	0.999
Adams	99.9	100	99.9	100	99.9	1.0	0.999	99.9	99.9	99.9	100	99.8	0.999	0.998
St. Helens	100	100	99.8	100	100	1.0	0.999	100	100	100	100	100	1.0	1.0
Cascades	99.9	100	99.9	100	99.9	1.0	0.999	99.9	99.9	99.9	100	99.8	0.999	0.999
Colville	99.9	99.9	100	100	99.2	1.0	0.996	99.8	99.8	99.8	100	99.2	0.999	0.995
Chelan	100	100	100	100	100	1.0	1.0	99.9	99.9	100	100	99.7	1.0	0.998
Quincy	100	100	100	100	100	1.0	1.0	99.9	99.9	100	100	99.7	0.999	0.998
Tri-Cities	100	100	100	100	100	1.0	1.0	100	100	100	100	100	1.0	1.0
CRGNSA	100	100	100	100	99.8	1.0	0.999	100	100	100	100	99.9	1.0	0.999
Columbia	100	100	100	100	99.9	1.0	0.999	99.9	99.9	100	100	99.7	1.0	0.999

Table S3. Case Study 1: accuracy statistics and variable importance, for the seven most important predictors (variable importance value in parentheses), for the 1m-resolution classification of rocky (R) and non-rocky (NR) land cover

Accuracy Statistics	Overall Accuracy	User's Accuracy (R)	User's Accuracy (NR)	Producer's Accuracy (R)	Producer's Accuracy (NR)	F-Score (R)	F-Score (NR)
	97.4%	98.5%	96.6%	95.7%	98.9%	0.971	0.977
Variable Importance	1	2	3	4	5	6	7
	$NDVI_{maximum}$ texture (238.7)	$NDVI_{median}$ texture (212.8)	CHILI (190)	$NDVI_{NAIP}$ (188)	NIR NAIP band (183.5)	Blue NAIP band (178.1)	Green NAIP band (133.3)

Table S4. Case Study 3: areas of stability classes (km^2) for each sub-region (italicized) and region

Sub-Region / Region	**Land Area (km^2)**				
	Non-rocky (Stable)	**1/4 Rocky (Unstable)**	**2/4 Rocky (Unstable)**	**3/4 Rocky (Unstable)**	**Rocky (Stable)**
N. Cascades	7057	120.4	85.6	101.0	658.6
Winthrop	14,082	64.7	355.5	39.3	238.7
Forbidden	8868	112.6	86.0	129.9	903.5
Baker	16,498	132.8	92.6	125.6	799.0
Snoqualmie	14,283	97.0	63.2	87.2	536.7
Rainier	17,477	77.1	46.3	60.9	374.7
Adams	18,314	37.2	23.4	28.2	264.9
St. Helens	17,361	11.5	8.7	10.2	72.0
Cascades	113,940	653.2	441.2	582.2	3848
Colville	7538	14.8	4.9	4.8	10.5
Chelan	3938	15.1	6.5	6.4	17.2
Quincy	2130	11.7	7.3	8.6	32.1
Tri-Cities	11,488	3.0	1.8	2.2	8.1
CRGNSA	2585	17.7	9.1	9.3	22.3
Columbia	27,679	62.3	29.6	31.2	90.1