

Spatial Modeling Using High Performance Computing

Matthew Marsik^{1*} et al

*correspondence: mmarsik@ufl.edu

¹Land Use and Environmental Change Institute, University of Florida

Mapping Forest Management at Regional Scales in the Southeast U.S.

Matthew Marsik^{1*}, Michael Binford^{1,2}, Sherry Fu^{1,2}, Di Yang^{1,2}, Caroline Staub^{1,3}, Jaclyn Hall⁴

*correspondence: mmarsik@ufl.edu

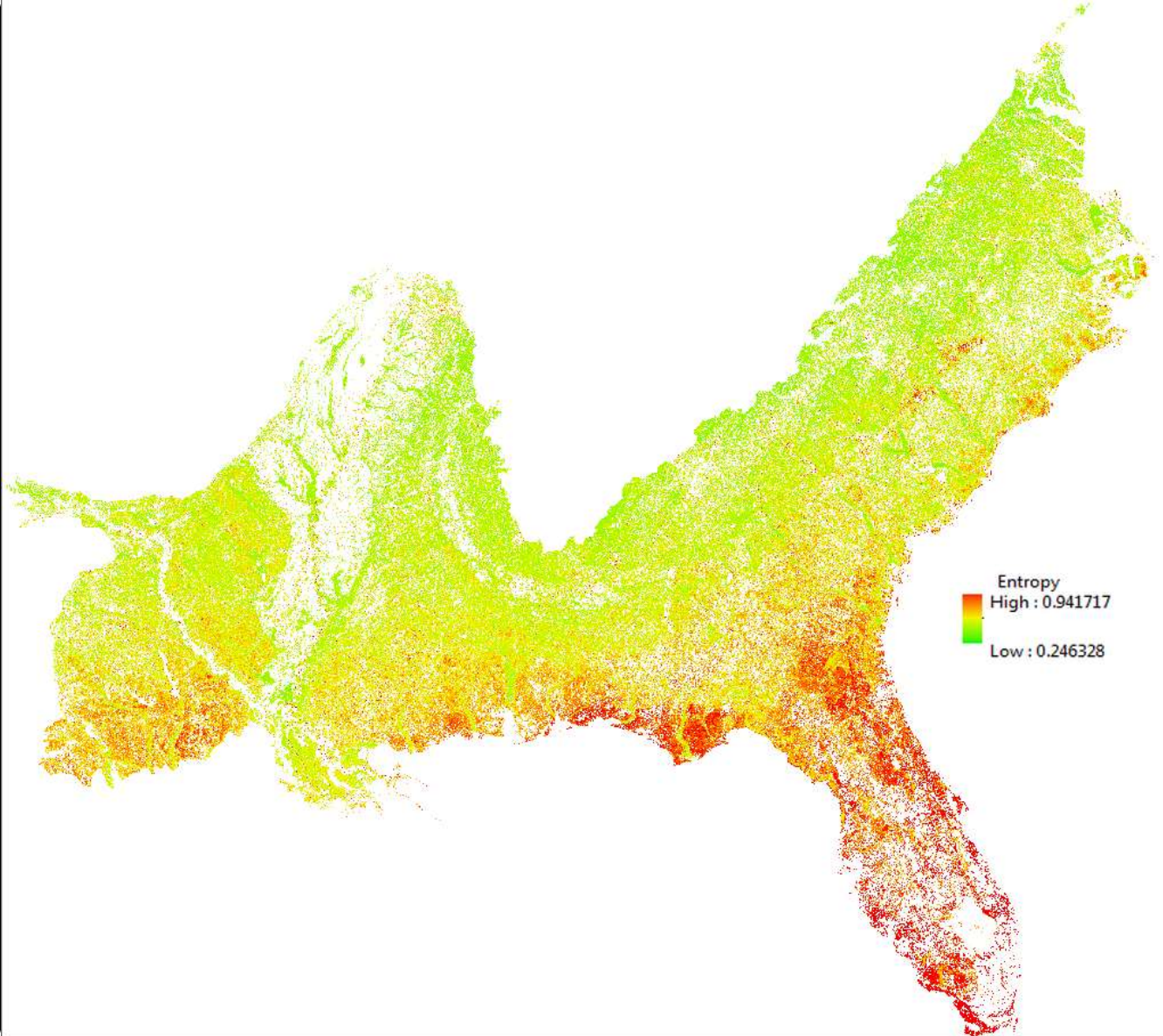
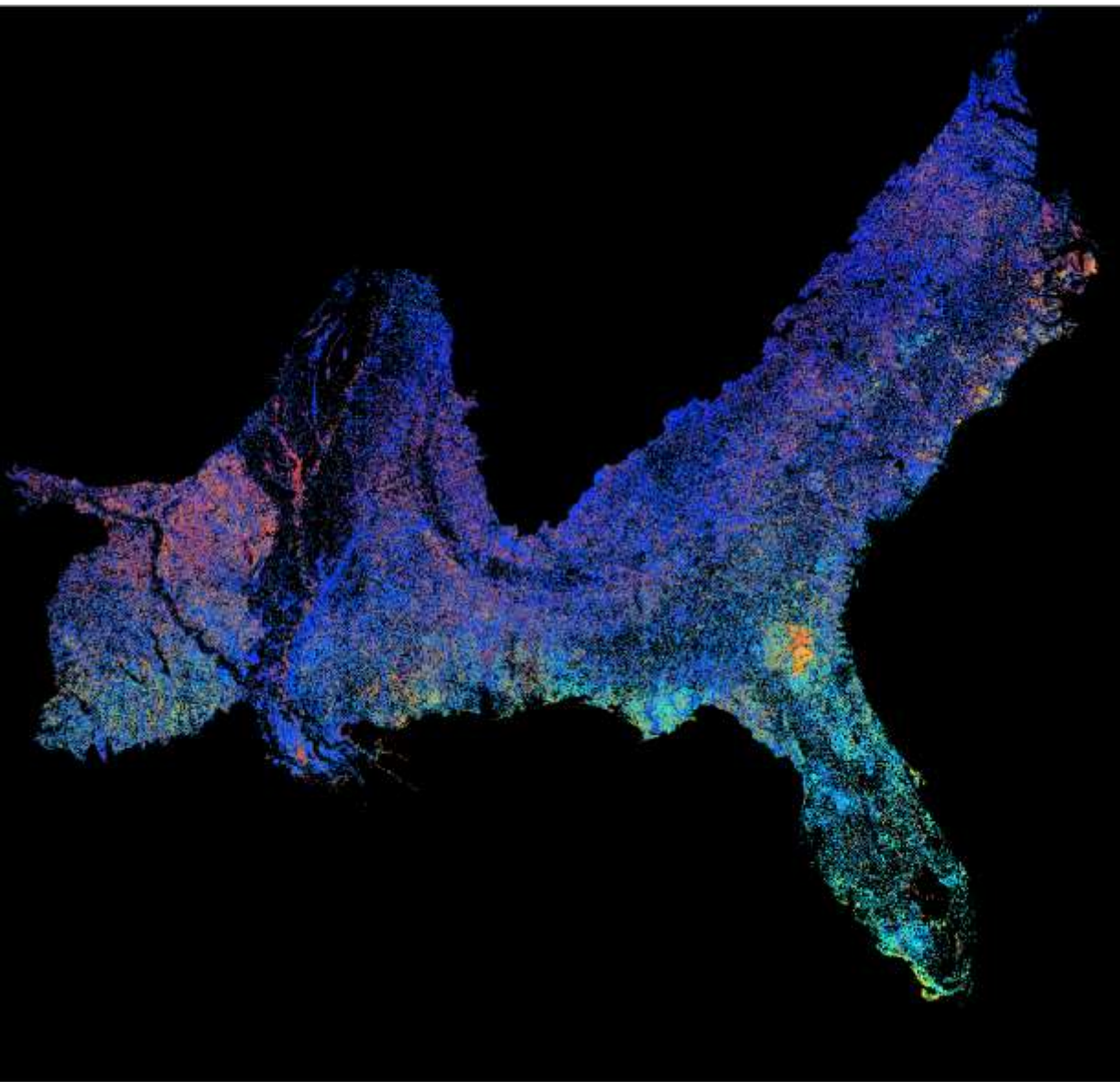
¹Land Use and Environmental Change Institute, University of Florida, ²Department of Geography, University of Florida, ³Agricultural and Biological Engineering, University of Florida, ⁴Health Policy & Outcomes, College of Medicine, University of Florida

University of Florida High Performance Computing Center - HiPerGator I & II

- ▶ Managed by UF Research Computing since 2004
 - ▶ Initially 21,000 cores (HiPerGator 1)
 - ▶ HiPerGator 2 adds 30,000 more cores
- ▶ Primarily Dell PowerEdge racks
 - ▶ Intel Xeon and AMD Opteron CPUs
- ▶ 3 PB of storage
- ▶ Linux with Moab job manager

Breaks for Additive Season Trends on UF HPC Cluster

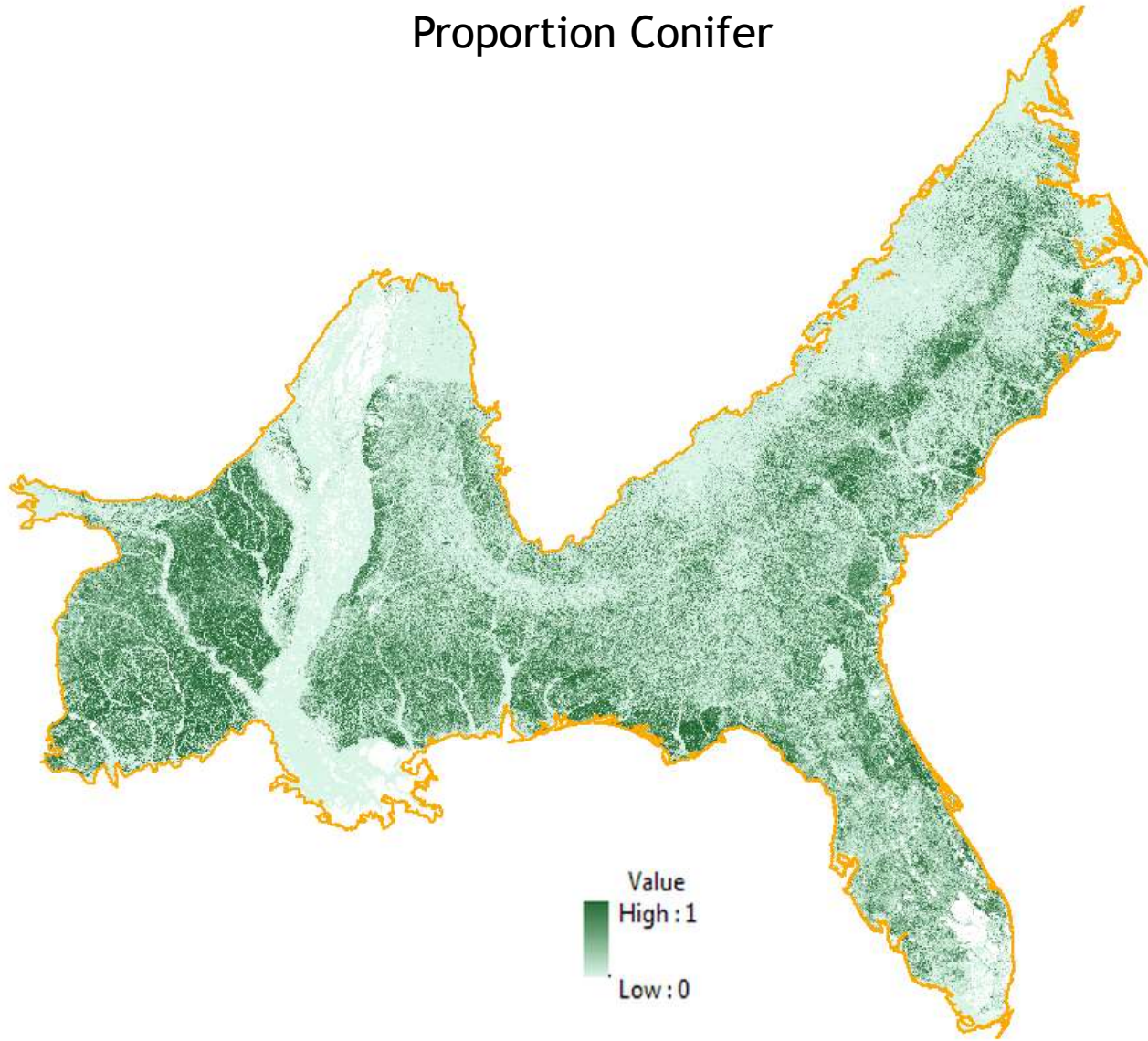
- ▶ BFAST (Zaccarelli et al., 2013) computed on raster stack of MODIS MOD13Q1 EVI 16-day composites
 - ▶ Per pixel time series analysis of vegetation phenology
 - ▶ 360 bands, ~16 million cells, ~9.8 million forested cells
- ▶ HPC request
 - ▶ 128 cores, 4GB per core, 2TB scratch space
 - ▶ 307 hours walltime, 13,662 CPU hours, 3 seconds/pixel processing time
- ▶ R 3.2
 - ▶ Rmpi package
 - ▶ BFAST package
- ▶ MPI computing workflow
 - ▶ 1 master/2 workers per core



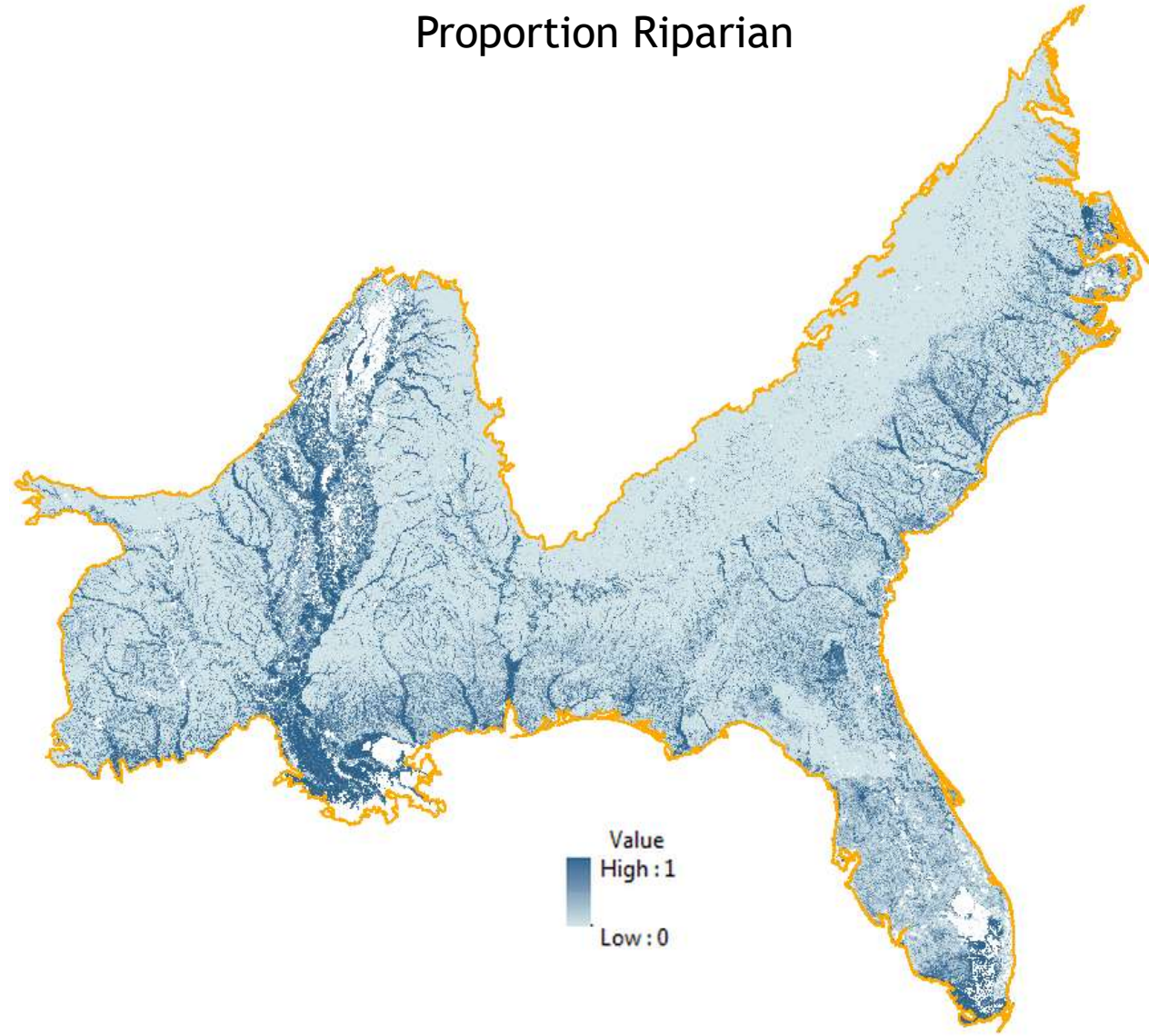
Multi-scale Proportion Landscape Calculation

- ▶ Calculate proportion of riparian and conifer classes in each 250 m cell
 - ▶ 250 m index raster with ~9.8 million unique values and cells
 - ▶ 30 m Landfire vegetation type raster ~60 million cells
 - ▶ Repeat for four time steps: 2002, 2006, 2010, and 2012
 - ▶ Repeat for Pacific Northwest
- ▶ Cross tabulation would not run on UF HPC
 - ▶ Constrained to only 1 CPU
- ▶ ArcGIS arcpy and PostGreSQL able to process data
 - ▶ ~20 hours for SE US

Proportion Conifer



Proportion Riparian



Training Data



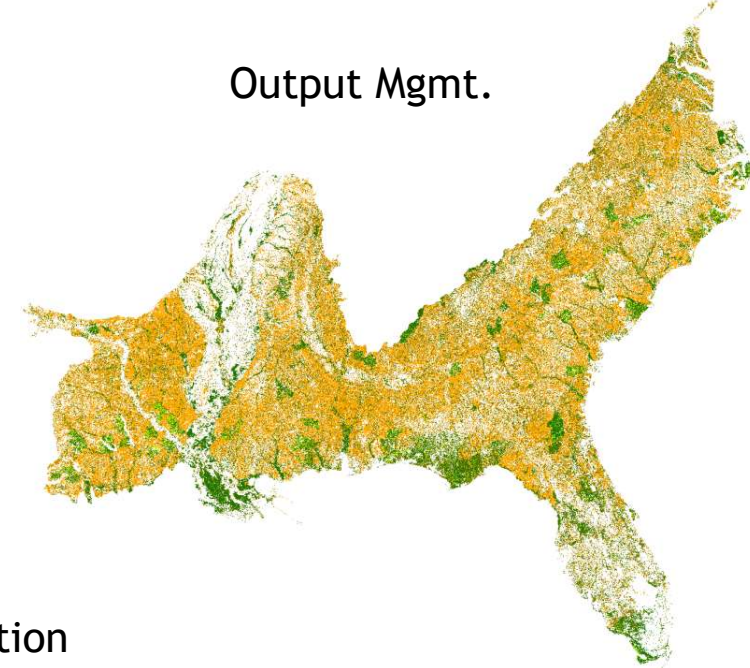
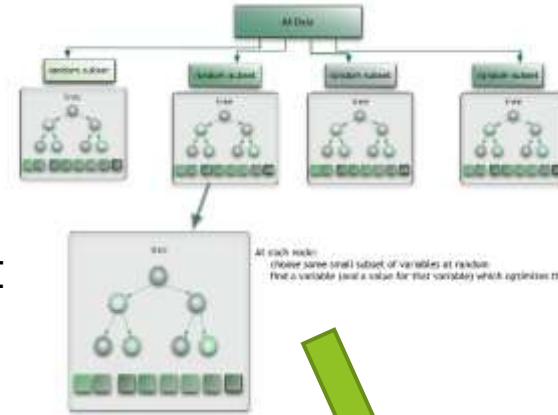
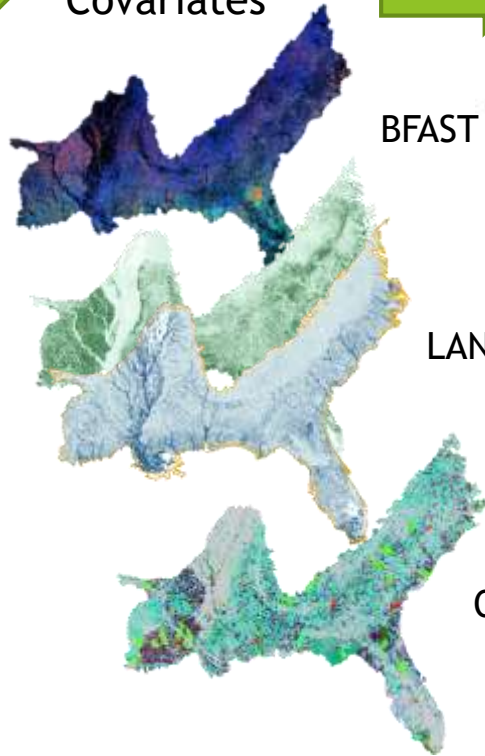
Covariates



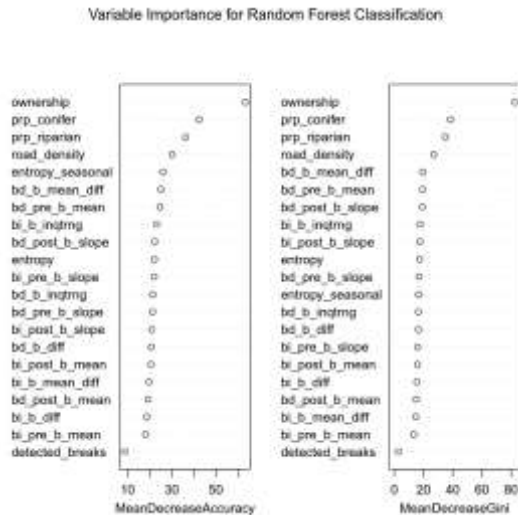
Random Forest



Output Mgmt.



Cross Validation



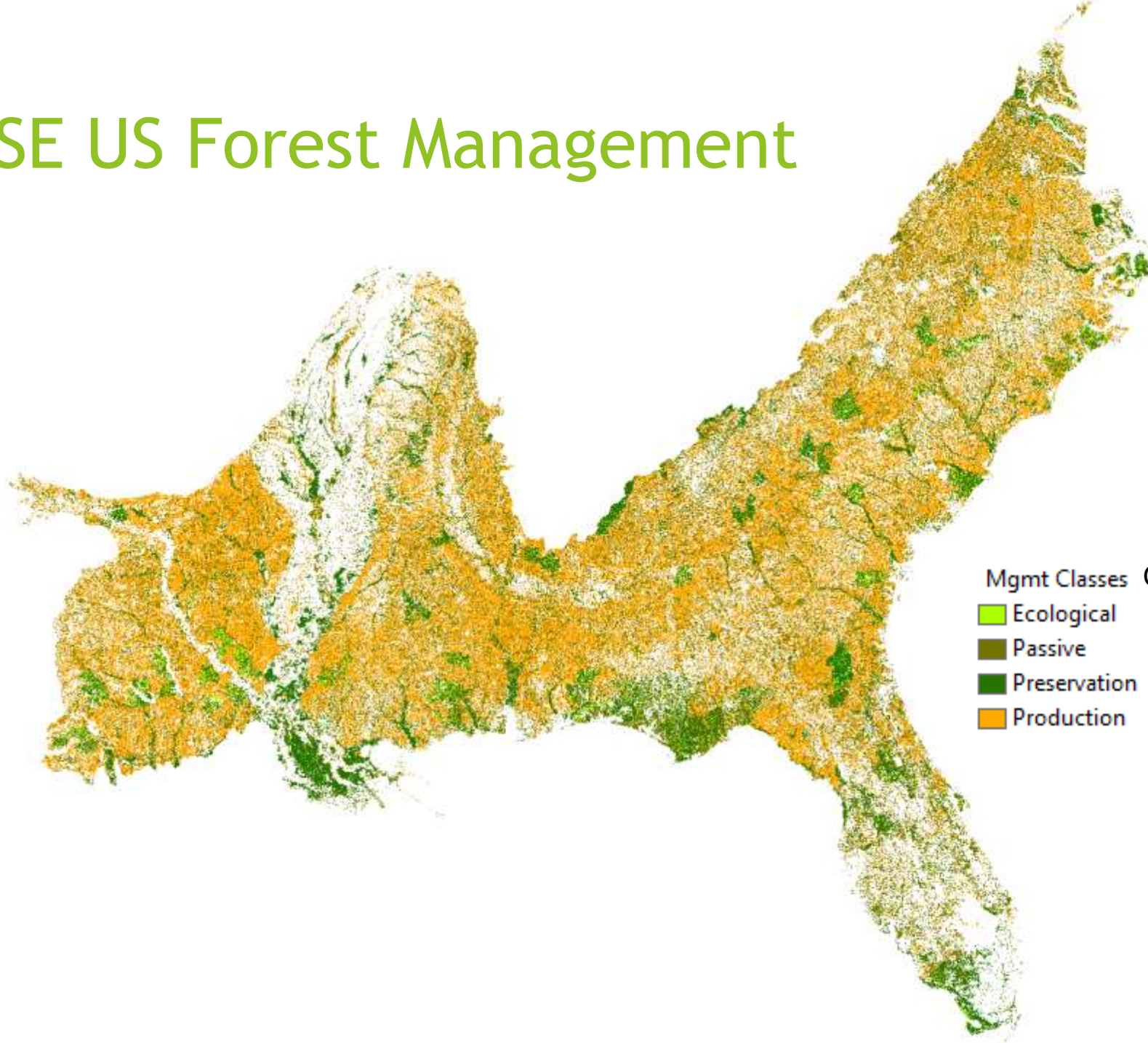
10-fold cross-validation (n=800)

	Ecological	Passive	Preservation	Production	Producers Accuracy
Total		27	110	191	472
Ecological	32	21	7	3	1
Passive	124	3	92	8	21
Preservation	207	1	7	167	32
Production	437	2	4	13	418
Users Accuracy		0.78	0.84	0.87	0.89

external validation (n=200)

	Ecological	Passive	Preservation	Production	Producers Accuracy
Total		2	22	54	122
Ecological	5	2	2	3	0
Passive	31	1	8	9	13
Preservation	57	1	9	30	17
Production	107	3	3	12	92
Users Accuracy		0.00	0.36	0.56	0.75

SE US Forest Management



Mgmt Classes	Cell Count	Percentages
Ecological	215,960	2.20%
Passive	2,230,235	22.73%
Preservation	1,844,837	18.81%
Production	5,519,086	56.26%
Total	9,810,118	100.00%

Summary: SE US Forest Management

- ▶ Data processing on UF's HPC Successful for MODIS MOD13Q1 BFAST analysis
- ▶ All other data processing done on 'high performance' desktop
- ▶ Will refine SE US management map with additional training data
- ▶ Extend random forest mapping to PNW
 - ▶ BFAST (48 hours wall time ~4,200 hours)
 - ▶ Proportion riparian and conifer processed (~6 hours ea. dataset ea. year)