Data Driven Discovery from Satellite Remote Sensing: System Development and Analysis of Vegetation Indices

Praveen Kumar
Department of Civil & Env. Eng.
University of Illinois
Urbana, Illinois 61801

NASA ESTO Project

Team

Praveen Kumar
Assoc. Prof.
Civil & Env. Eng.

Ricky Robertson
Post Doc Civil Eng.

David Tcheng
NCSA

David Clutter
NCSA

Peter Bajcsy
Res. Sci. NCSA

Amanda White
Grad Student
Civil & Env. Eng.

Rob Kooper
NCSA

Vikas Mehra
Grad
Student
Civil & Env.
Eng.

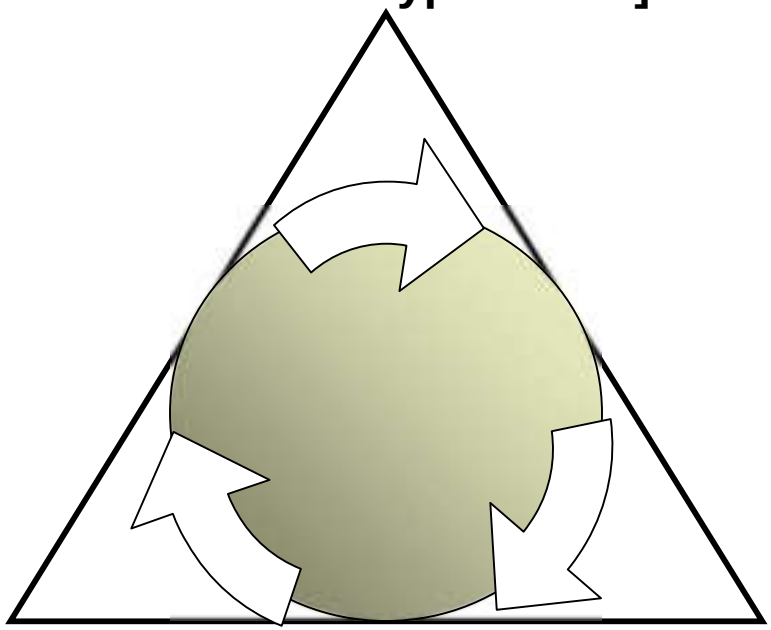
Wei-Wen Feng
Grad Student
Comp. Sci.

Ben Ruddell
Grad Student
Civil & Env. Eng.

Pratyush Sinha
Grad Student
Civil and Env. Eng.

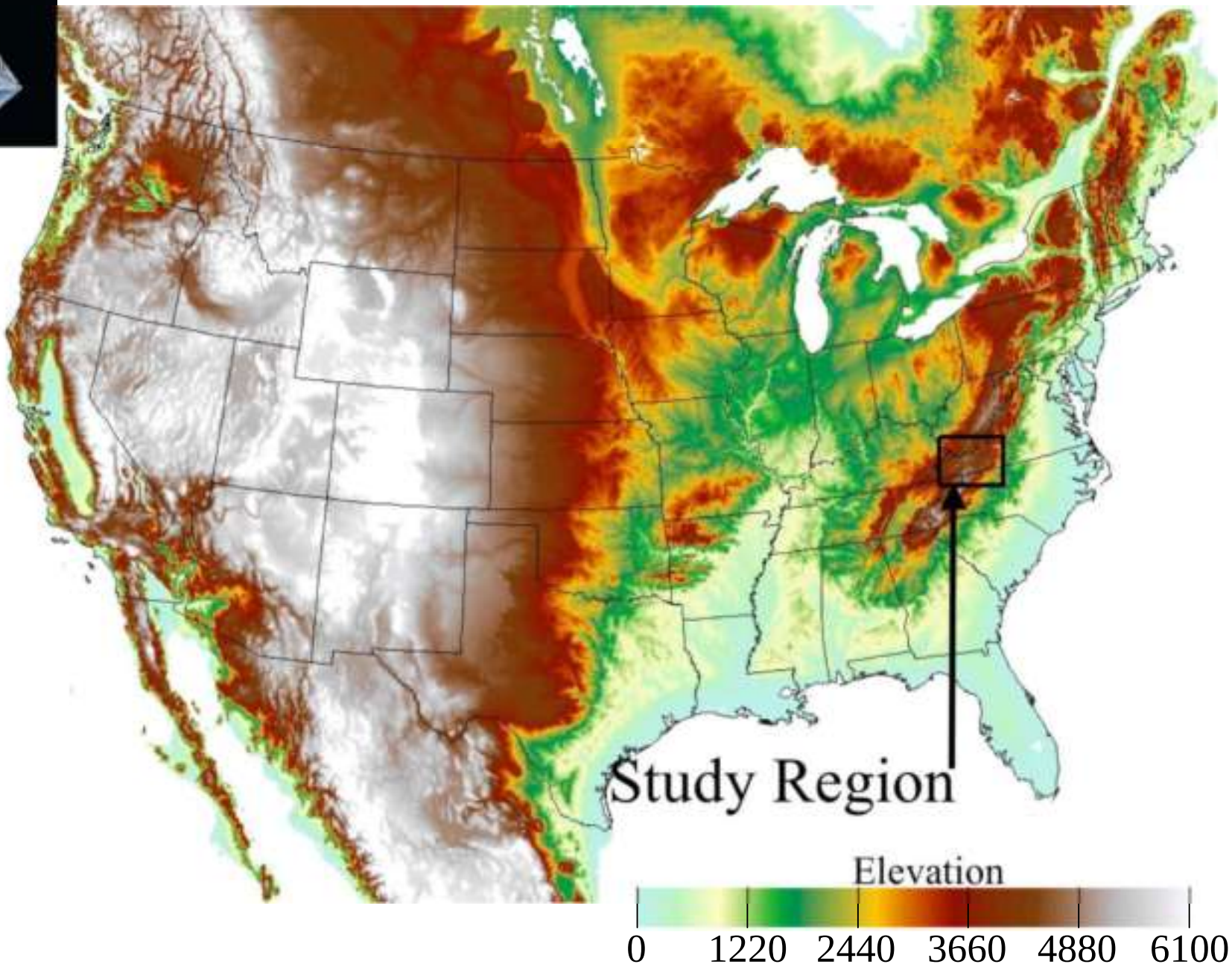
Data Mining Success Needs

Science Drivers
[A priori or data Driven hypotheses]

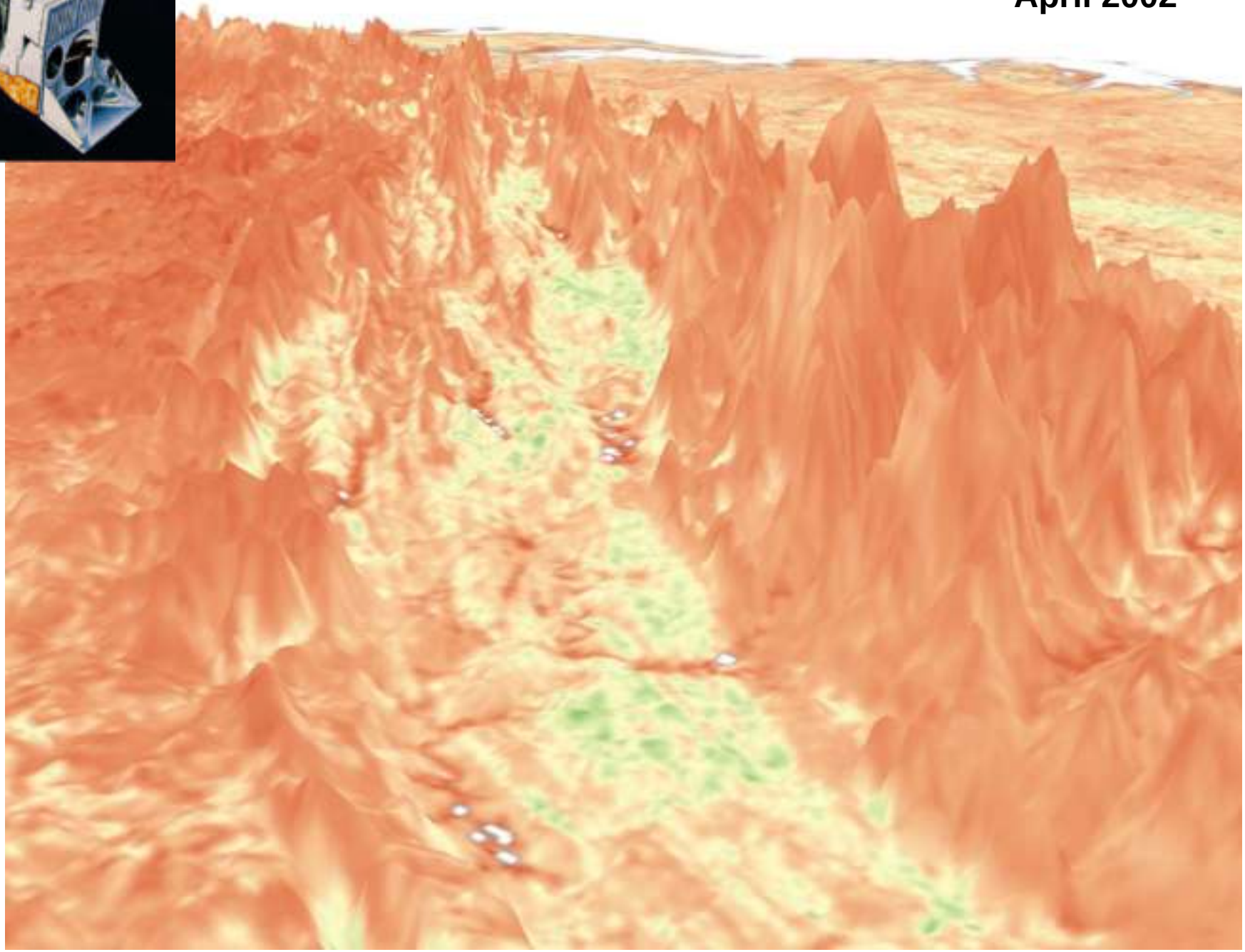


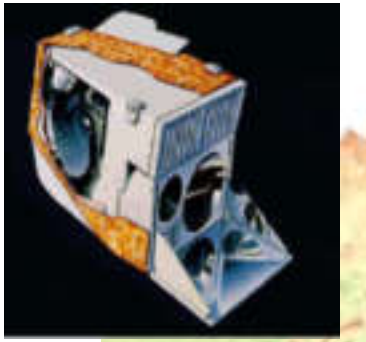
Observing/ Data System
[Sensors, resolution data access]

Data Mining Technology
[Computer Science/ Statistics]

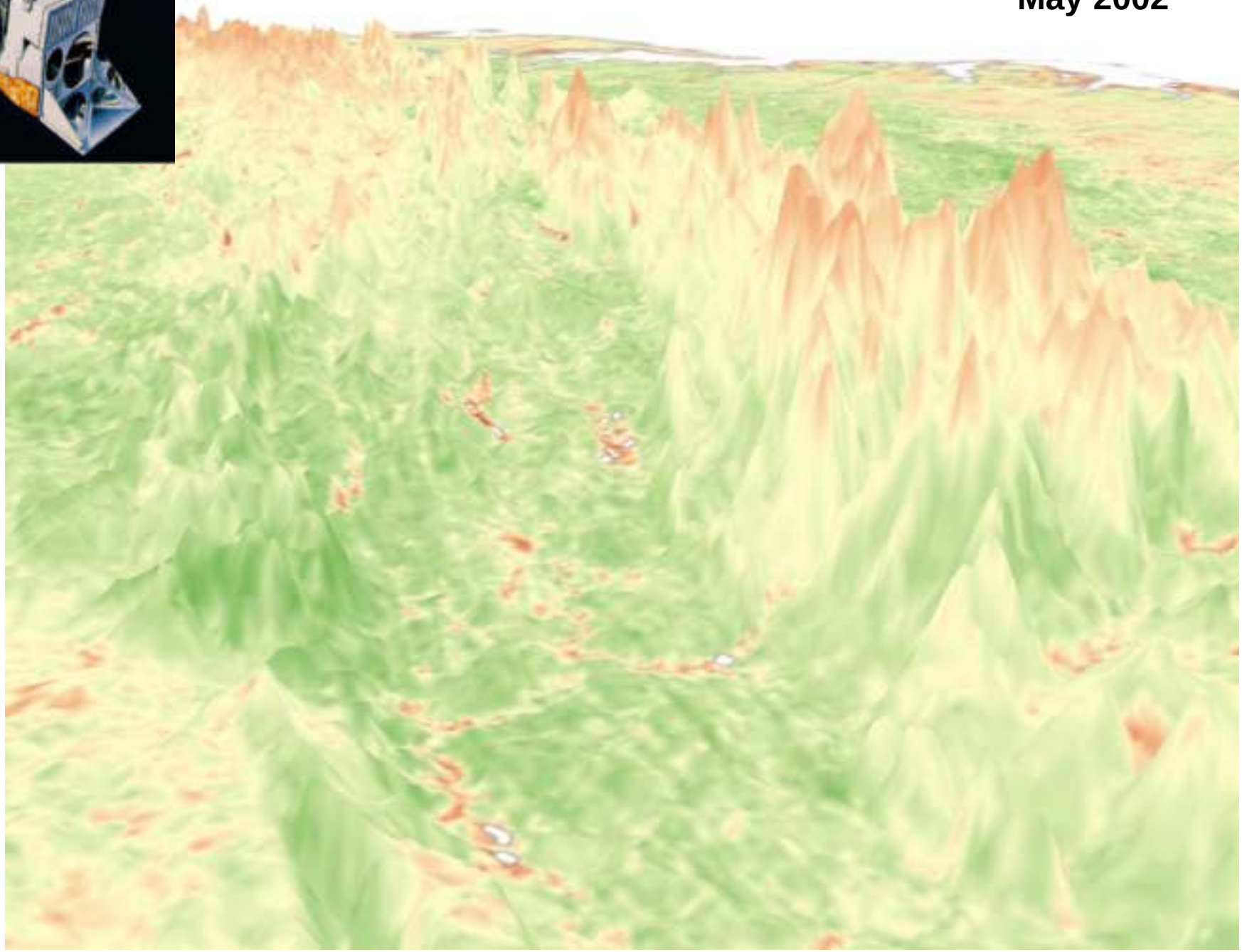

Study Region
Elevation
0
1220
2440
3660
4880
6100

April 2002

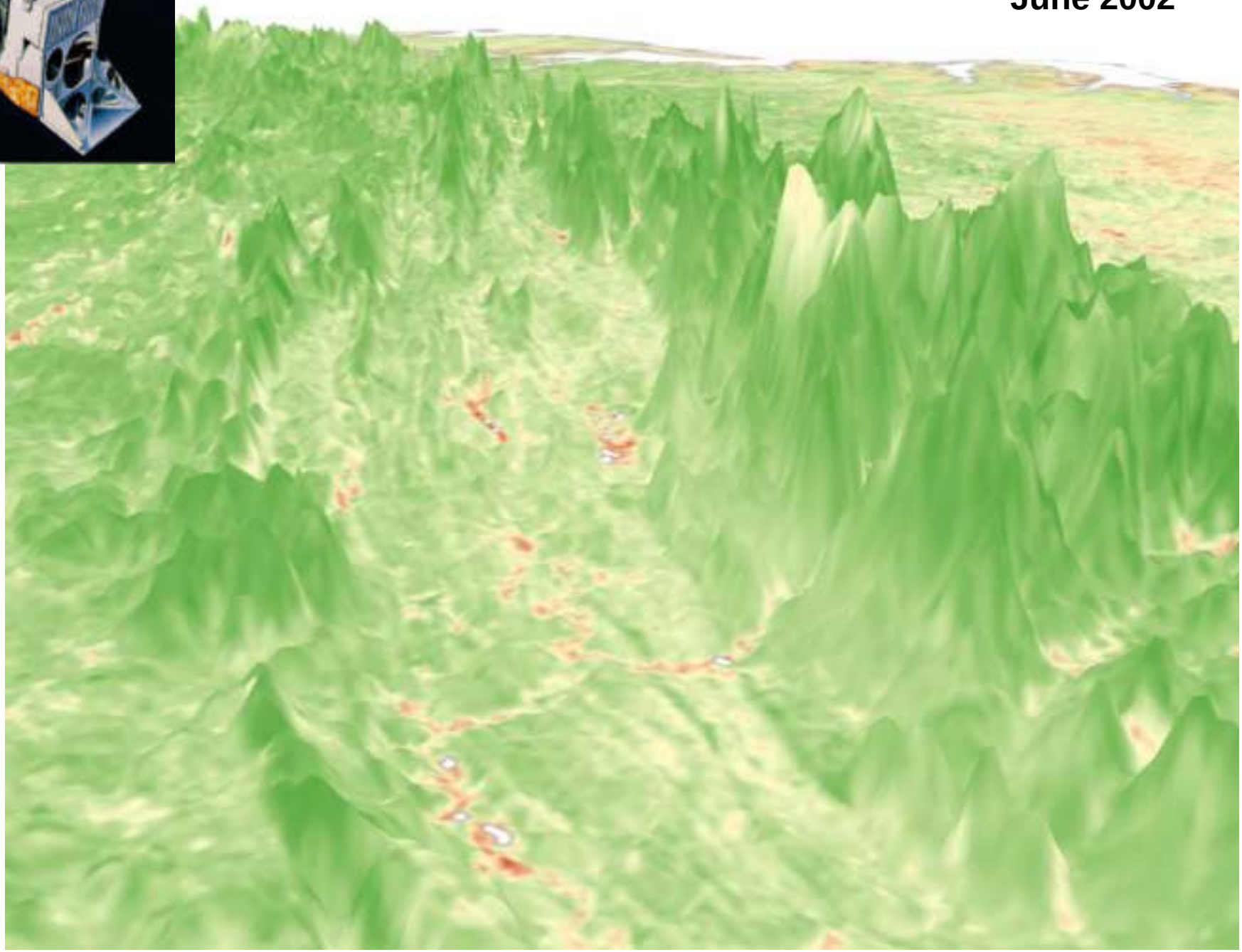




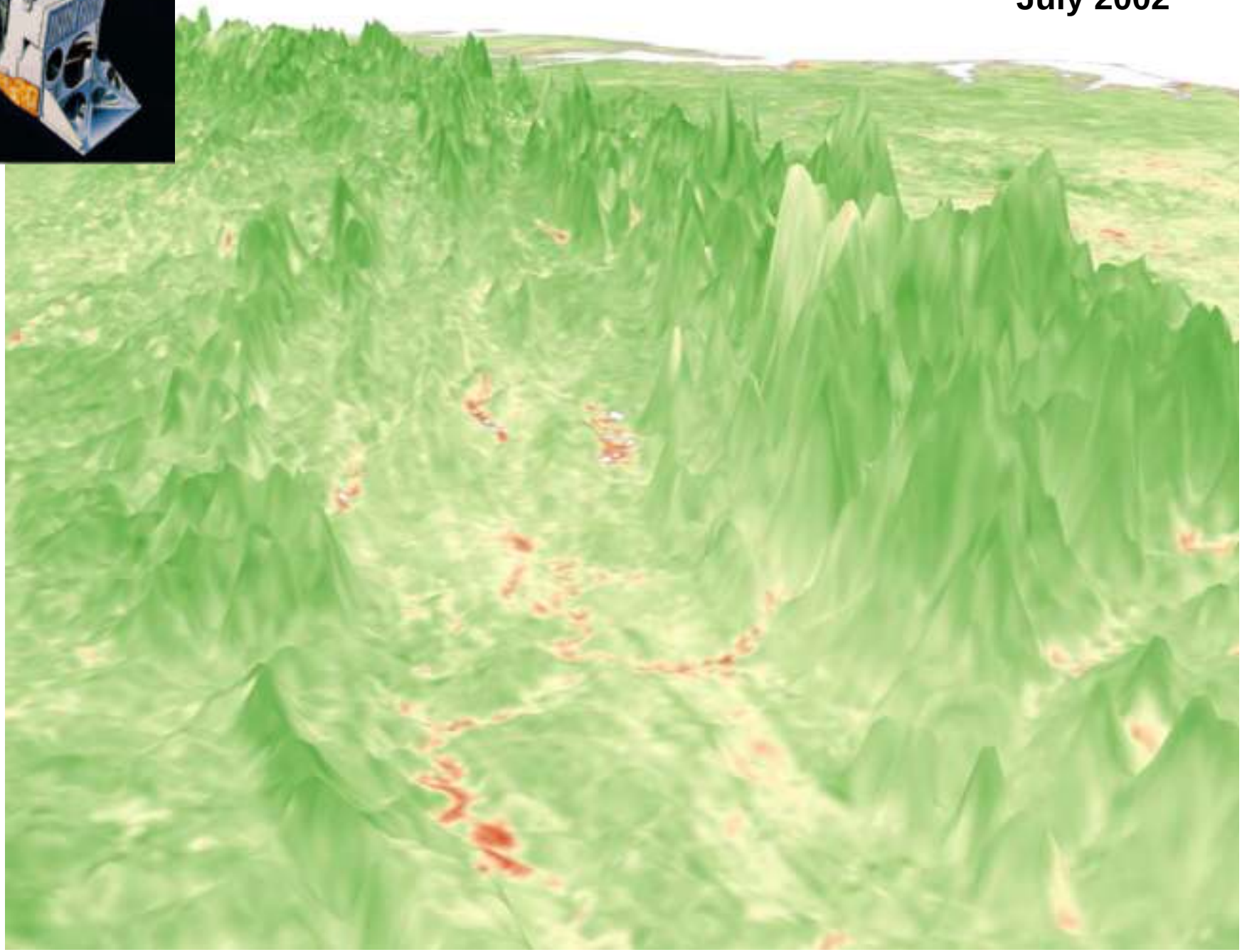
May 2002



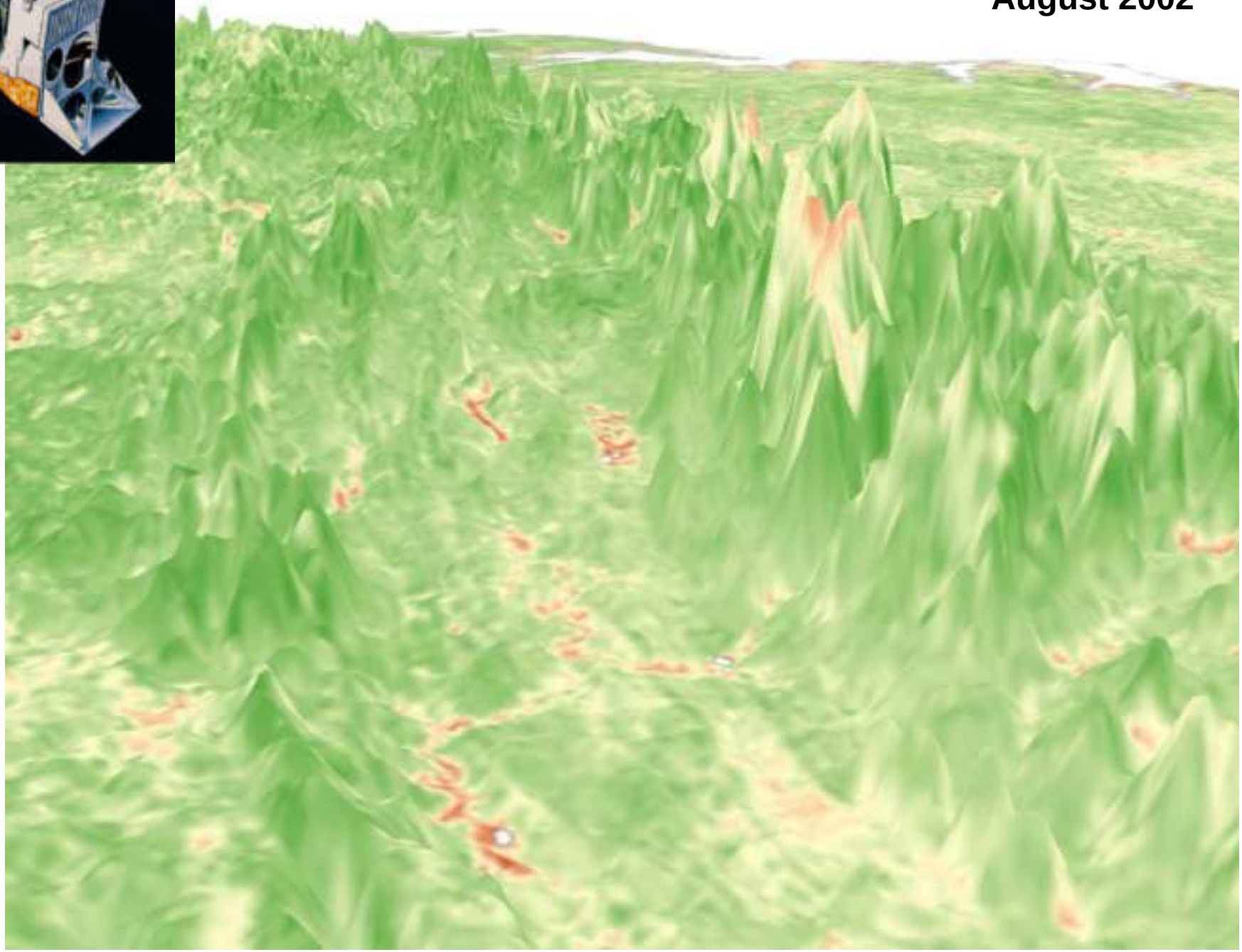
June 2002

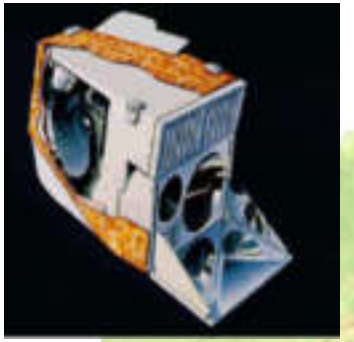


July 2002

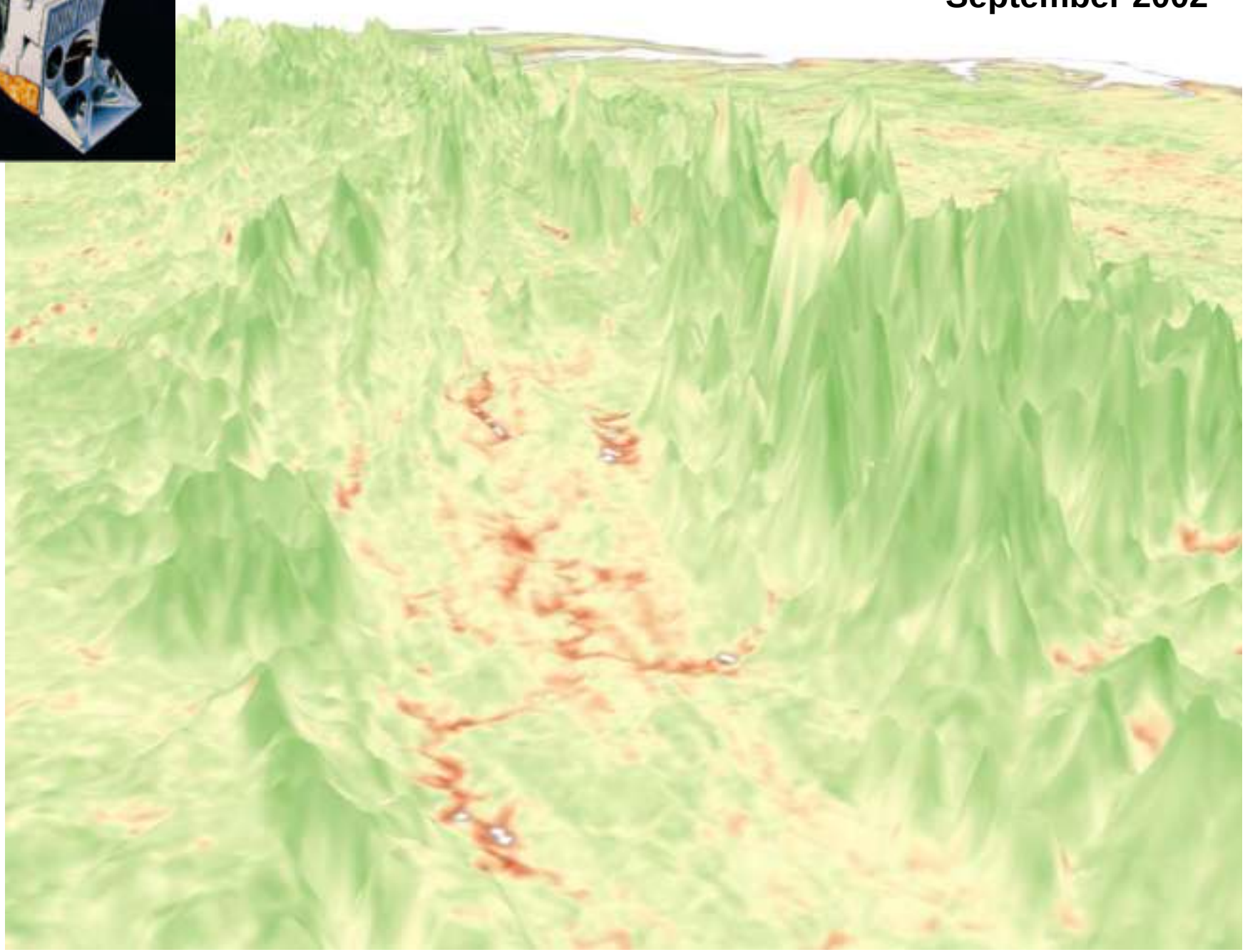


August 2002

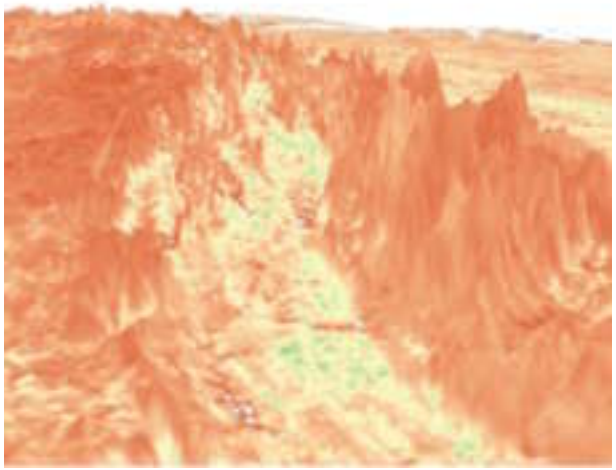




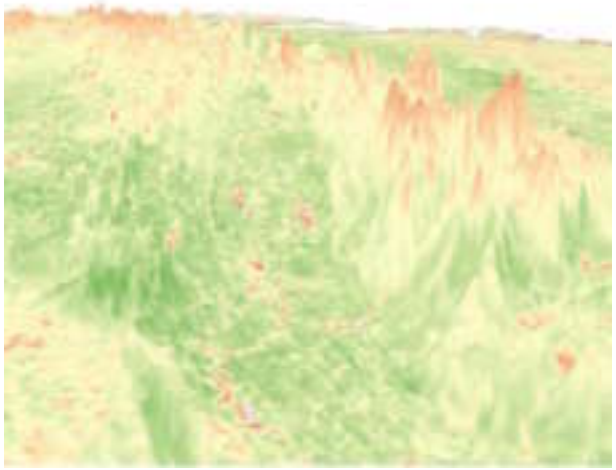
September 2002



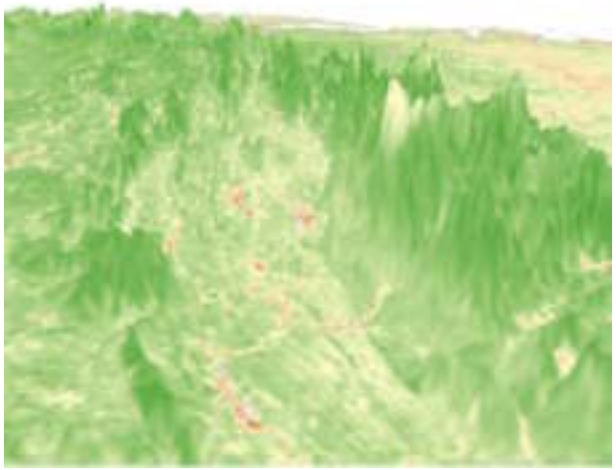
April 2002



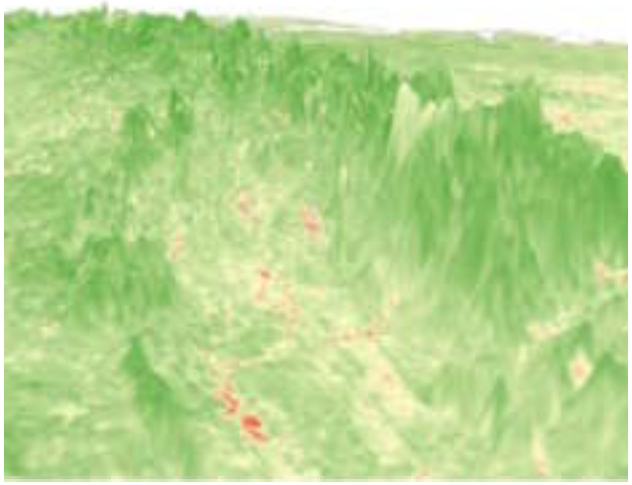
May 2002



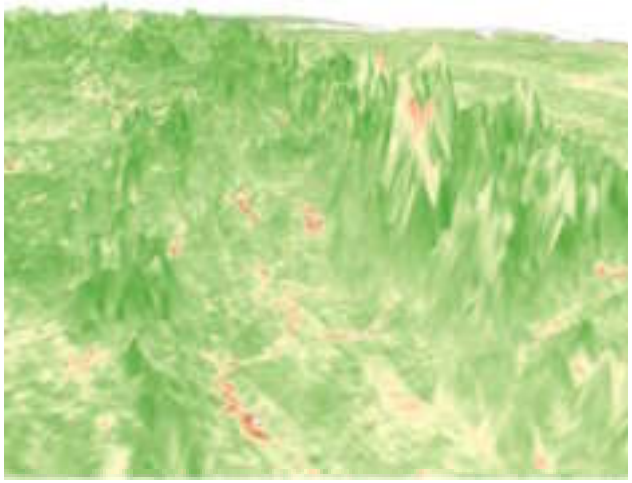
June 2002



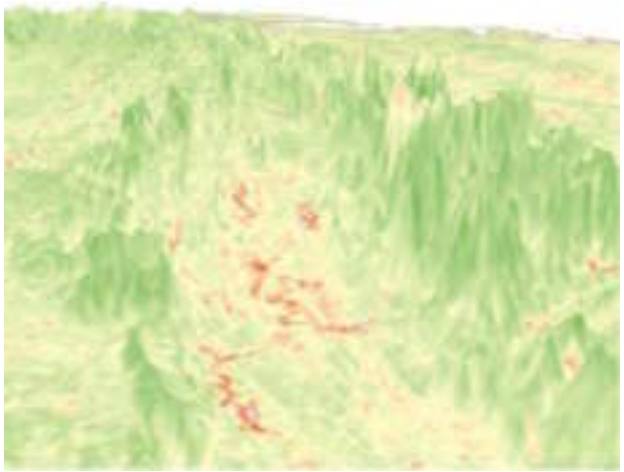
July 2002



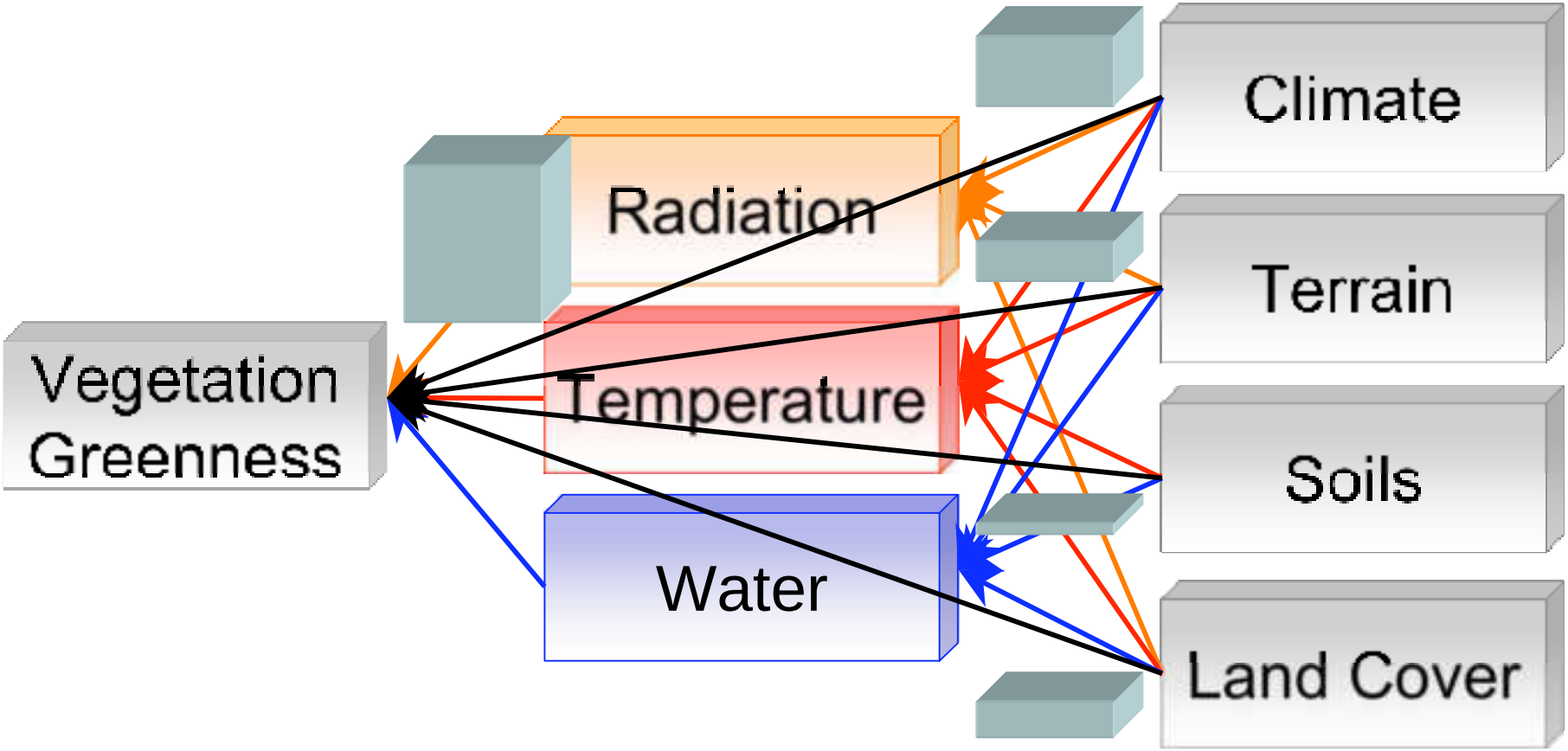
August 2002

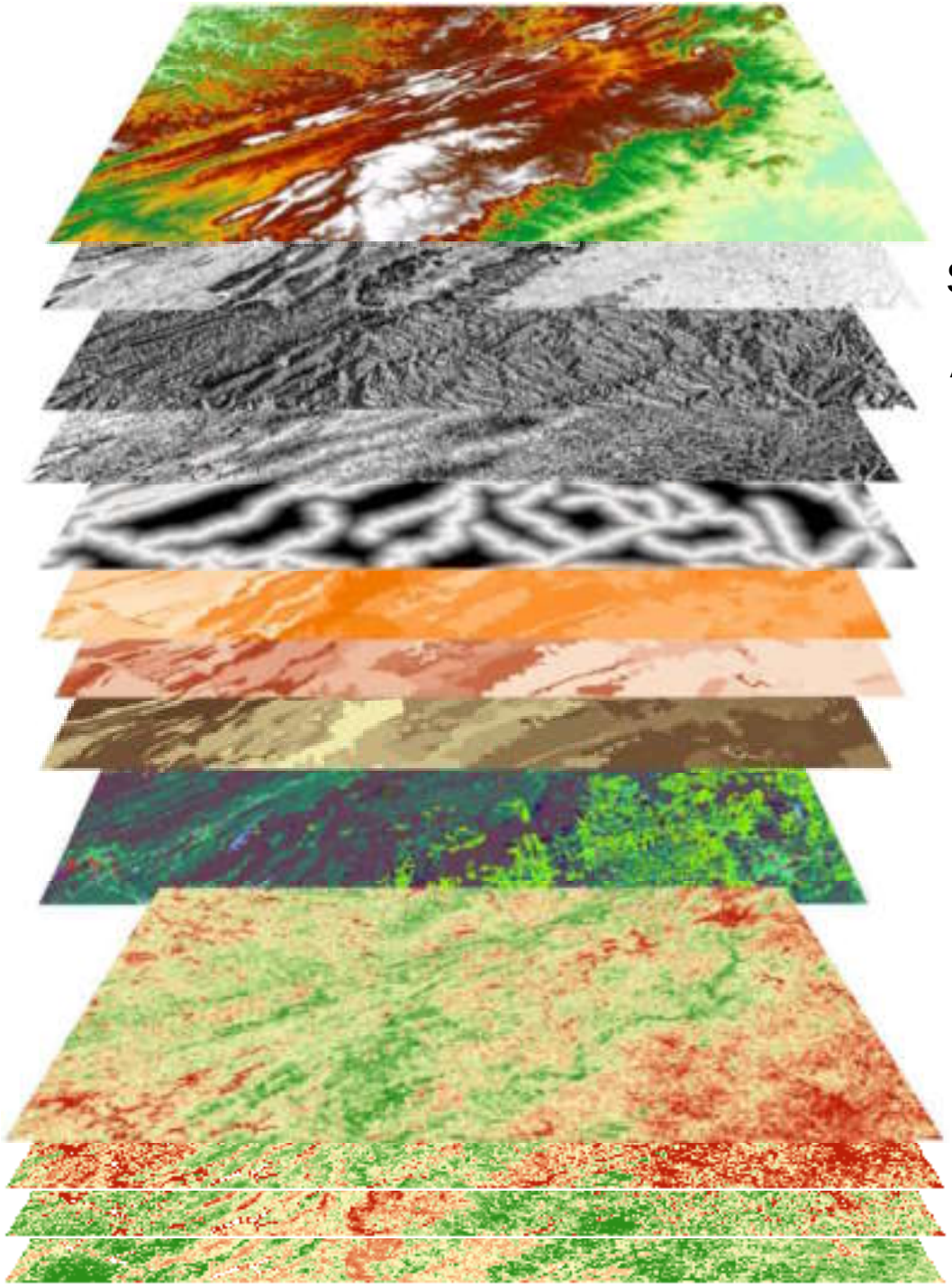


September 2002



What are the dominant influences on vegetation greenness?





INPUTS

Elevation

Slope

Aspect

Compound Topographic Index

Distance to Nearest Stream

Percent Sand

Percent Silt

Percent Clay

Land Use/Land Cover

Possibly many more …

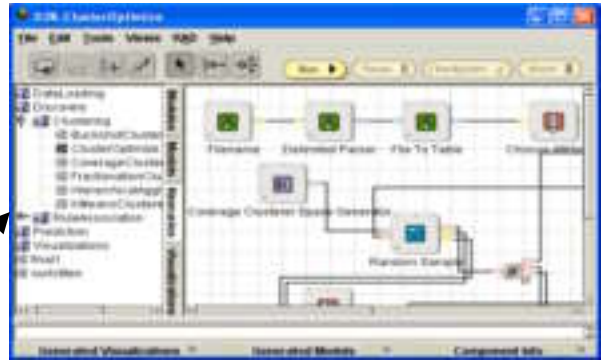
OUTPUT

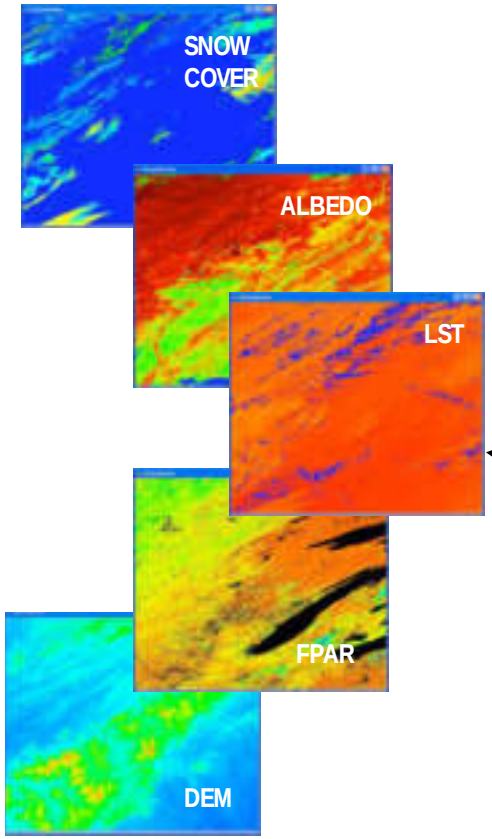
April 2000
April 2001
April 2002
April 2003

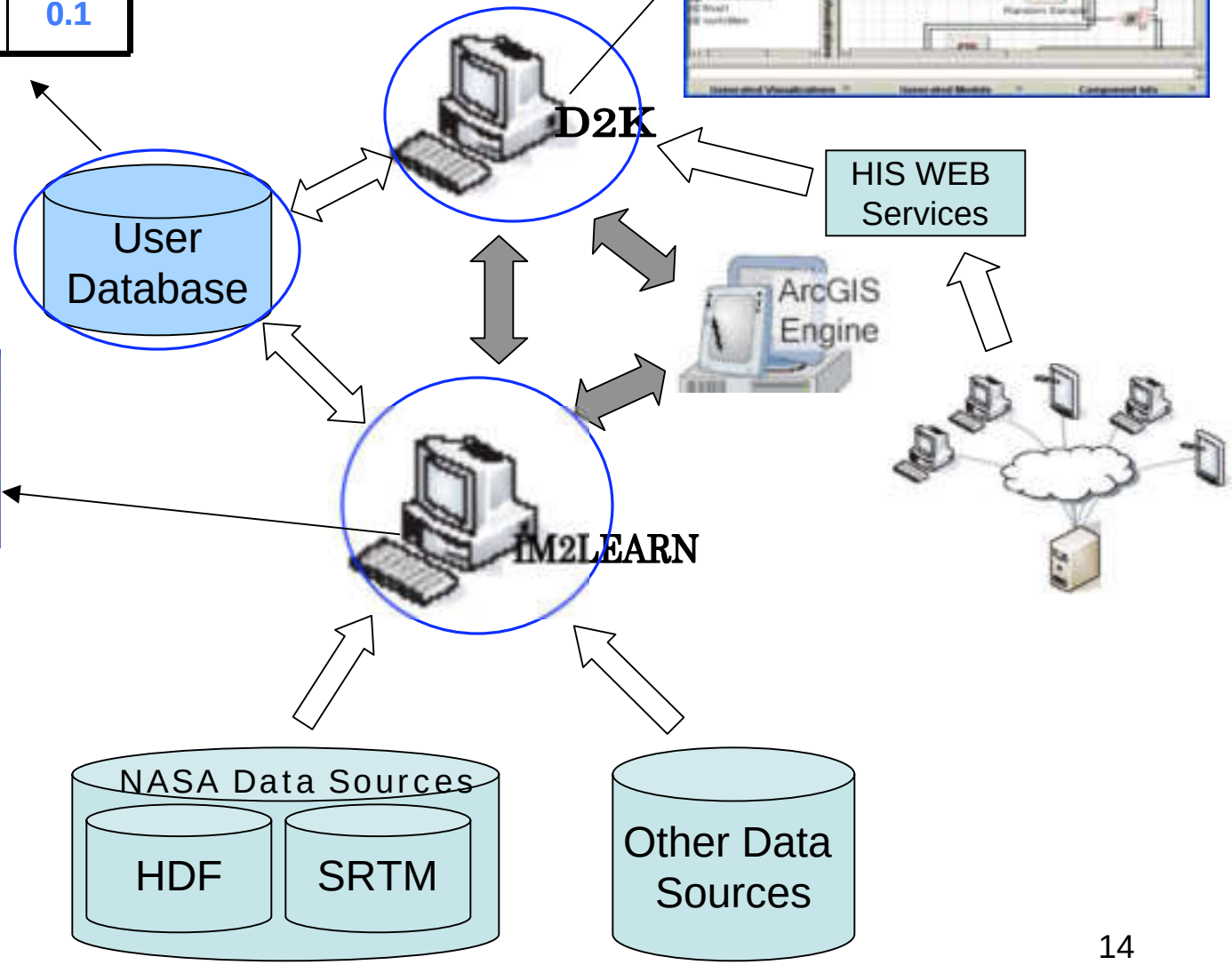
Average Enhanced Vegetation Index for April

Pixel	EVI	Albedo	LAI
1	0.4	0.2	0.3
2	0.3	0.2	0.1
3	0.2	0.3	0.1

GeoLearn System Architecture

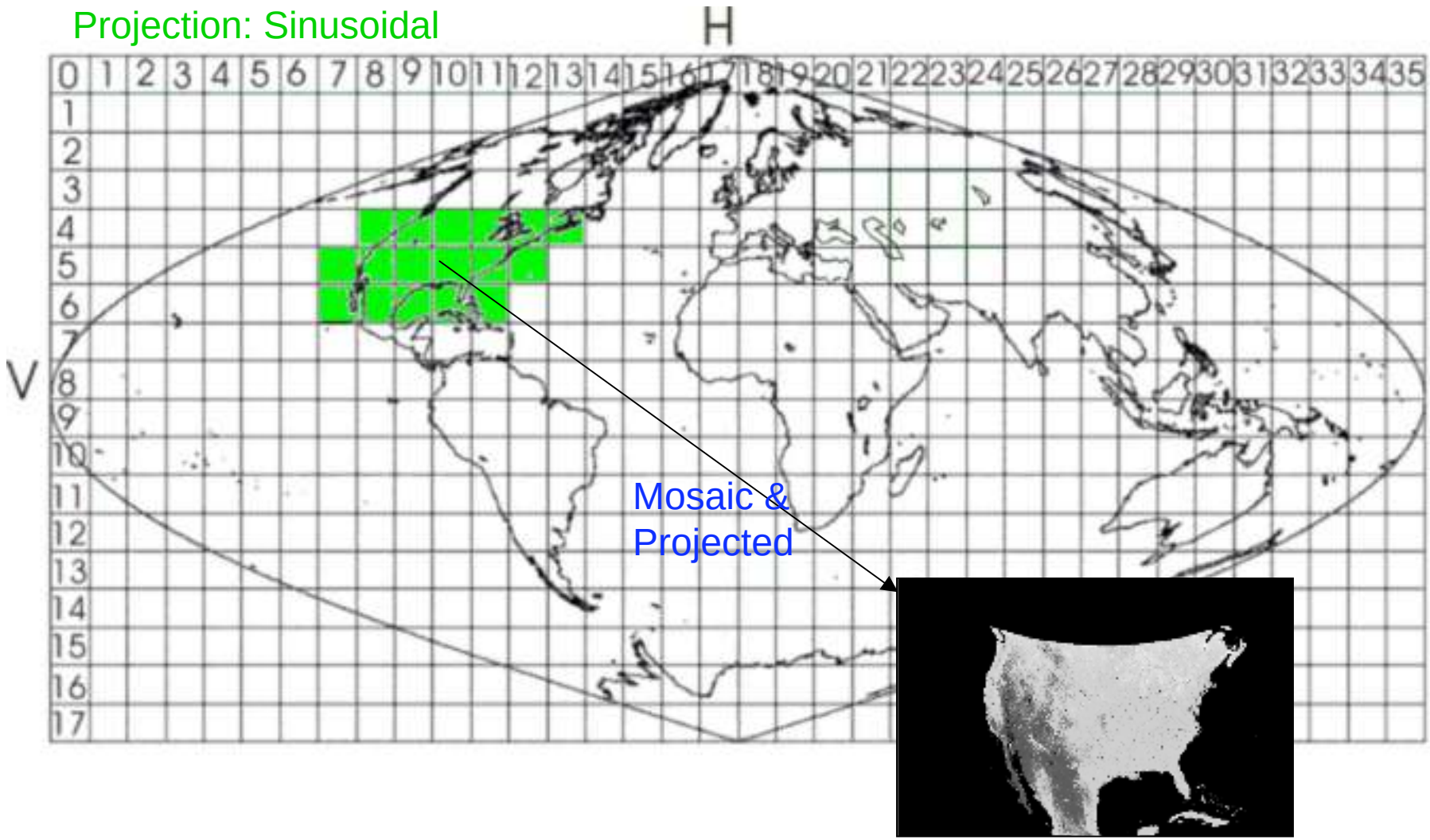






US MODIS Tiles

Projection: Sinusoidal

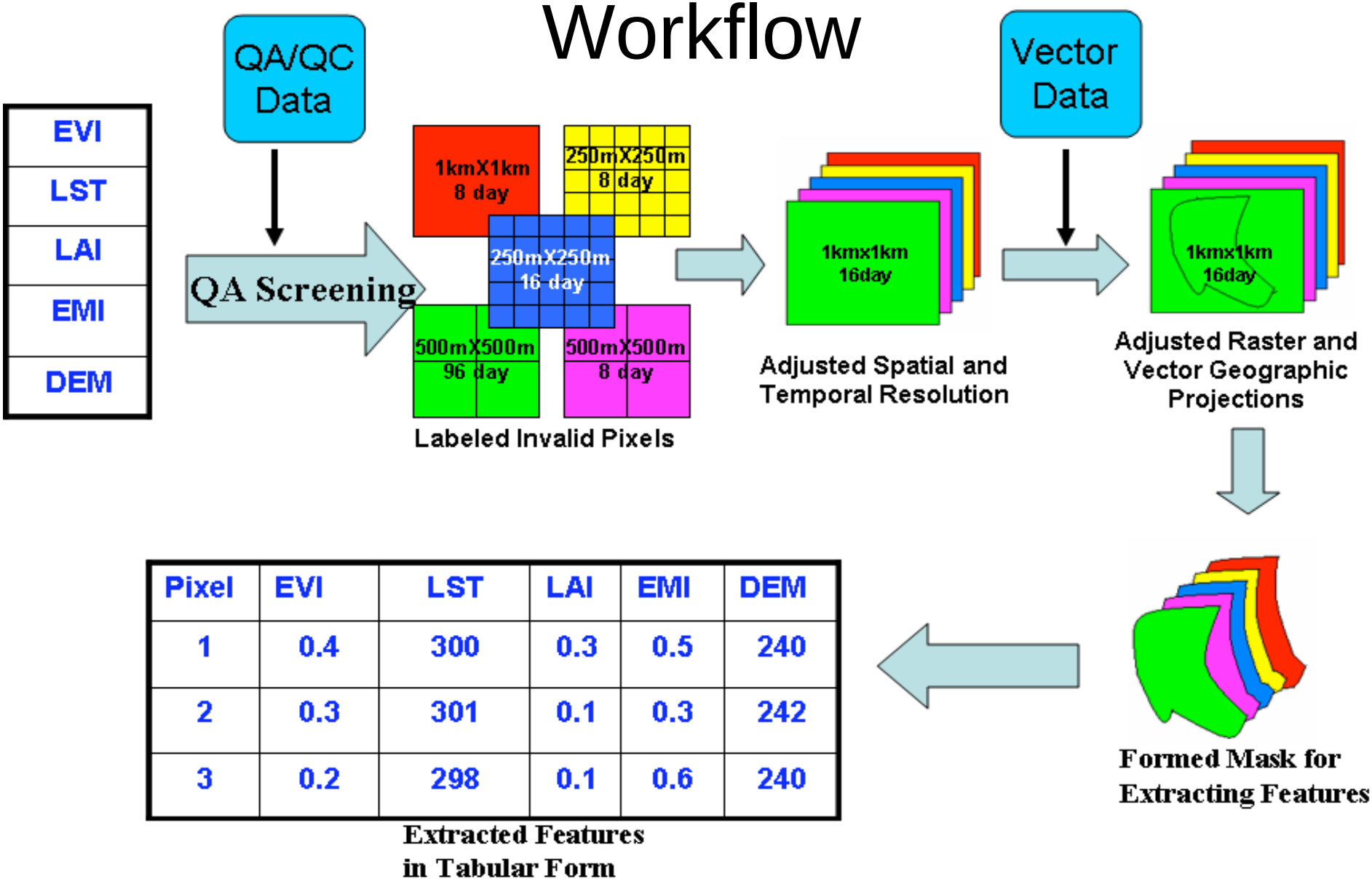


Albers Equal Area Conic

Spatio-Temporal Resolution issues in HDF data processing

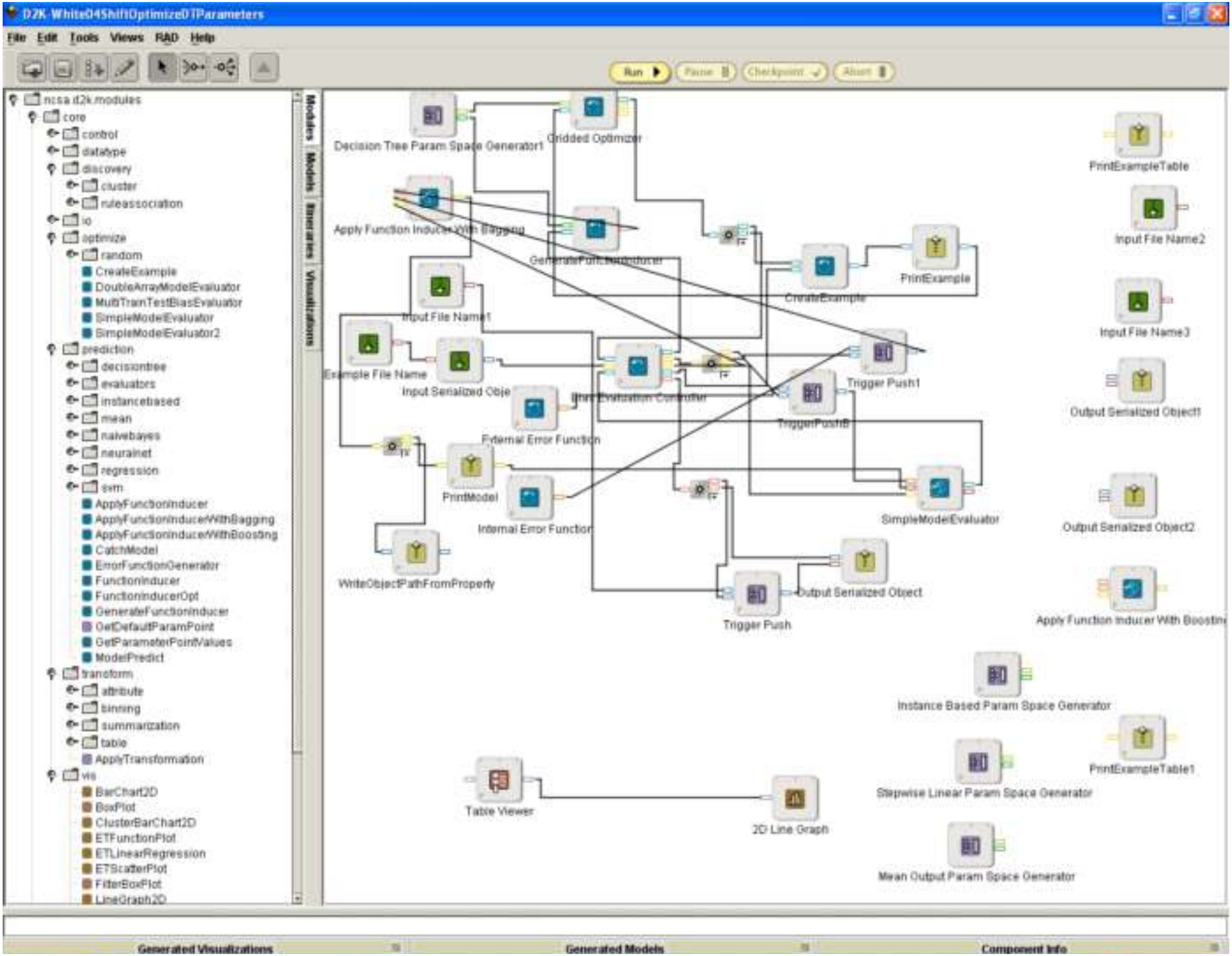
Datasets Name	Projection	Resolution (meters)	Geospatial data type	Time Period (days)	Data Size (MB)
EVI / NDVI	Sinusoidal	250	Grid	16	495
LAI / FPAR	Sinusoidal	1000	Grid	8	6
LAND COVER	Sinusoidal	1000	Grid	96	22
LST / EMISSIVITY	Sinusoidal	1000	Grid	8	20
SNOW COVER	Sinusoidal	500	Grid	8	11
ALBEDO	Sinusoidal	1000	Grid	16	68

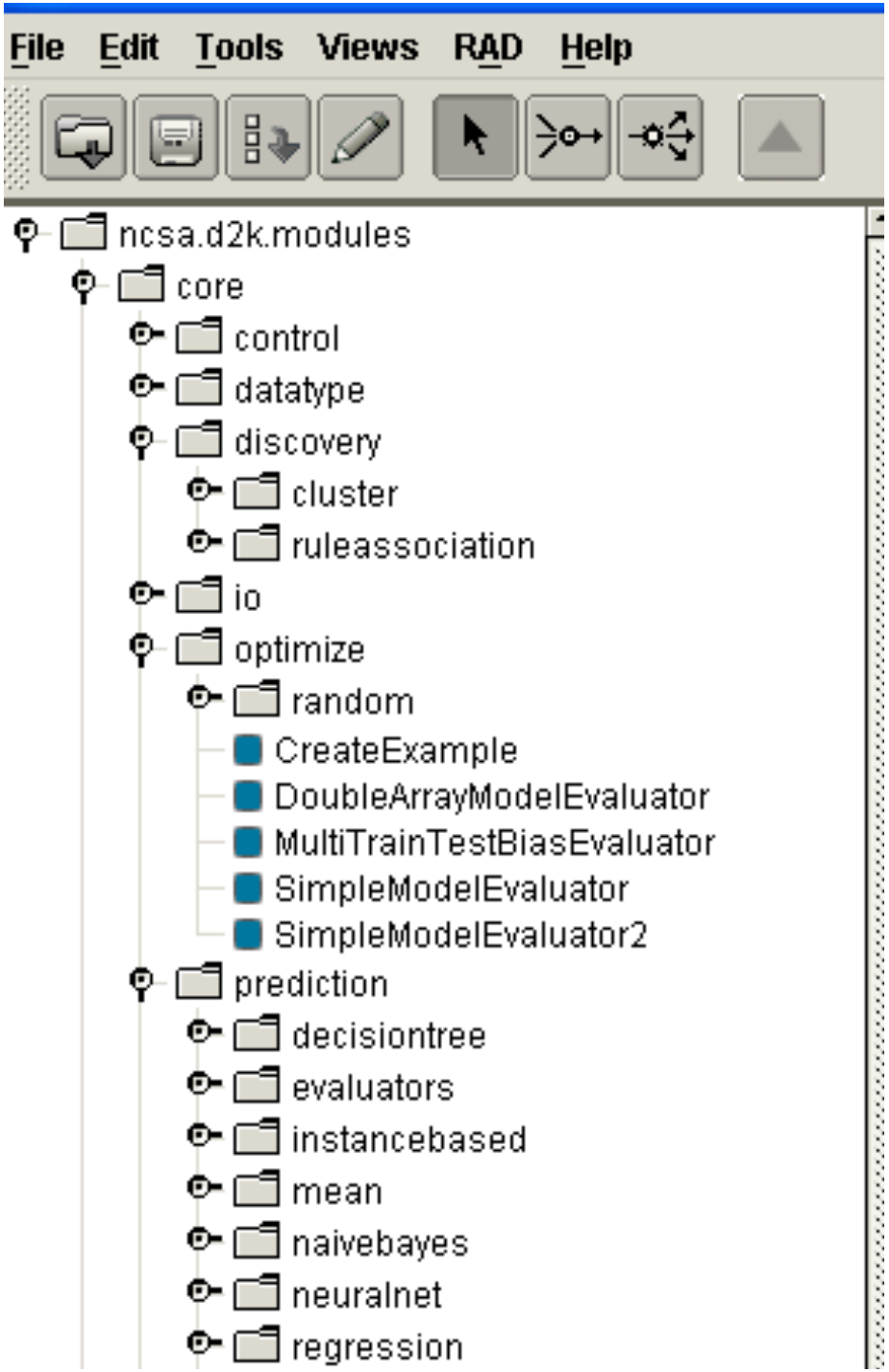
Remote Sensing Data processing Workflow

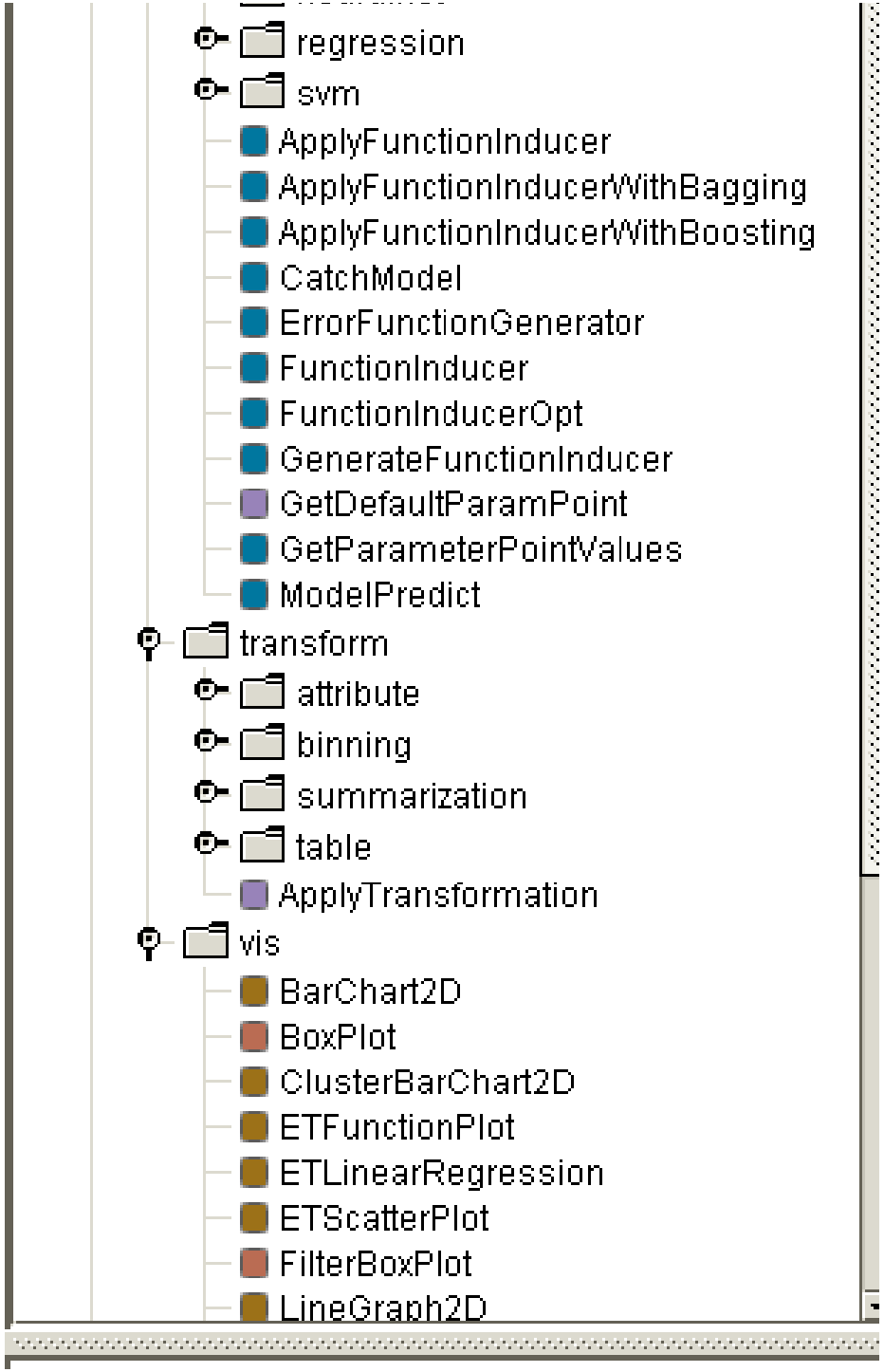


Pixel	EVI	LST	LAI	EMI	DEM
1	0.4	300	0.3	0.5	240
2	0.3	301	0.1	0.3	242
3	0.2	298	0.1	0.6	240

Extracted Features in Tabular Form


D2K-White04ShiftOptimizeDTParameters
File Edit Tools Views RAD Help
Run
Pause
Checkpoint
Abort
ncsa.d2k.modules
core
control
datatype
discovery
cluster
ruleassociation
io
optimize
random
CreateExample
DoubleArrayModelEvaluator
MultiTrainTestBiasEvaluator
SimpleModelEvaluator
SimpleModelEvaluator2
prediction
decisiontree
evaluators
instancebased
mean
naivebayes
neuralnet
regression
svm
ApplyFunctionInducer
ApplyFunctionInducerWithBagging
ApplyFunctionInducerWithBoosting
CatchModel
ErrorFunctionGenerator
FunctionInducer
FunctionInducerOpt
GenerateFunctionInducer
GetDefaultParamPoint
GetParameterPointValues
ModelPredict
transform
attribute
binning
summarization
table
ApplyTransformation
vis
BarChart2D
BoxPlot
ClusterBarChart2D
ETFunctionPlot
ETLinearRegression
ETScatterPlot
FilterBoxPlot
LineGraph2D
Modifiers
Models
Itineraries
Visualizations
Decision Tree Param Space Generator1
Gridded Optimizer
Apply Function Inducer With Bagging
GenerateFunctionInducer
Input File Name1
CreateExample
PrintExample
PrintExampleTable
Input File Name2
Input File Name3
Example File Name
Input Serialized Obje
Trigger Push1
TriggerPushB
Output Serialized Object1
External Error Function
PrintModel
Internal Error Function
SimpleModelEvaluator
Output Serialized Object2
WriteObjectPathFromProperty
Output Serialized Object
Trigger Push
Apply Function Inducer With Boosting
Instance Based Param Space Generator
PrintExampleTable1
Stepwise Linear Param Space Generator
Table Viewer
2D Line Graph
Mean Output Param Space Generator
Generated Visualizations
Generated Models
Component Info


File
Edit
Tools
Views
RAD
Help
ncsa.d2k.modules
core
control
datatype
discovery
cluster
ruleassociation
io
optimize
random
CreateExample
DoubleArrayModelEvaluator
MultiTrainTestBiasEvaluator
SimpleModelEvaluator
SimpleModelEvaluator2
prediction
decisiontree
evaluators
instancebased
mean
naivebayes
neuralnet
regression


regression
svm
ApplyFunctionInducer
ApplyFunctionInducerWithBagging
ApplyFunctionInducerWithBoosting
CatchModel
ErrorFunctionGenerator
FunctionInducer
FunctionInducerOpt
GenerateFunctionInducer
GetDefaultParamPoint
GetParameterPointValues
ModelPredict
transform
attribute
binning
summarization
table
ApplyTransformation
vis
BarChart2D
BoxPlot
ClusterBarChart2D
ETFunctionPlot
ETLinearRegression
ETScatterPlot
FilterBoxPlot
LineGraph2D

MODIS VI QA bit layout

1-0-0-11-0-0-0-00-0011-01

0 1 VI Quality **(00:Ideal 01:CheckOtherQA 10:Cloudy 11:NoProduct)**

2 3 4 5 VI Usefulness **(0000:HighestQuality 0001-1100:DescendingQuality 1101:NoAtmCorr 1110:LowestQuality 1111:NotUseful)**

6 7 Aerosol Quality **(00:Climatology 01:Low 10:Average 11:High)**

8 Adjacency Correction **(0:No 1:Yes)**

9 Atmospheric BRDF Correction **(0:No 1:Yes)**

10 Mixed Cloud **(0:No 1:Yes)**

11 12 Land/Water Flag **(00:Ocean 01:Coast 10:Wetland 11:Land)**

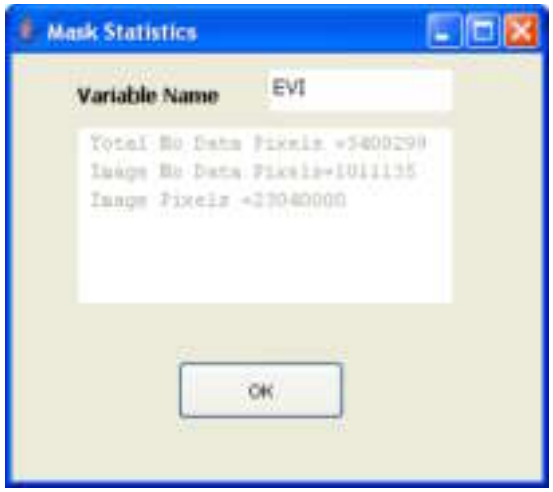
13 Possible Snow/Ice **(0:No 1:Yes)**

14 Possible Shadow **(0:No 1:Yes)**

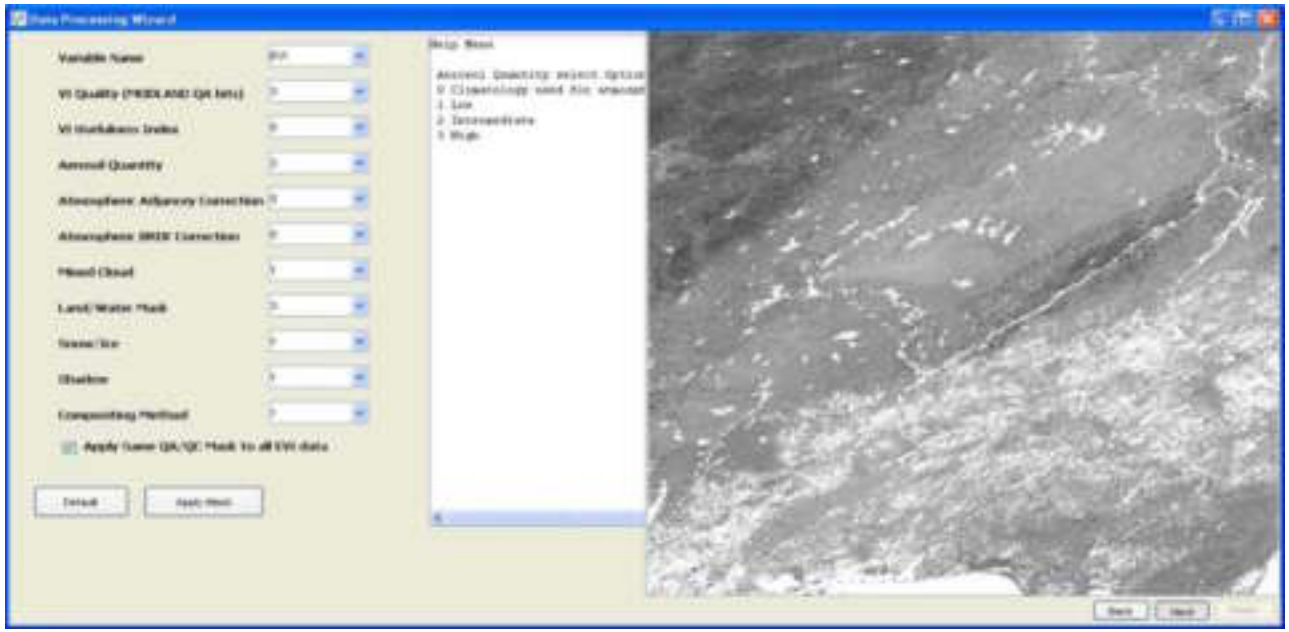
15 Composite Method **(0:BRDF 1:CV-MVC)**

Ref: http://edcdaac.usgs.gov/modis/mod13a1v4.asp

Quality Assurance/Quality Control Mask



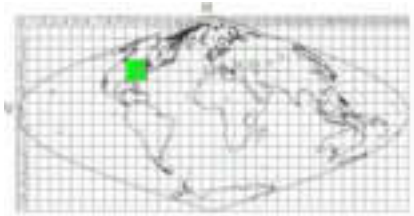
EVI Dataset before applying mask



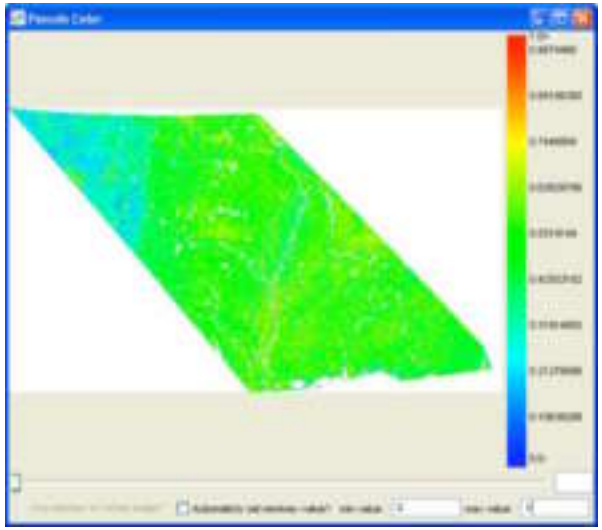
EVI Dataset after applying QA/QC mask

White region shows bad pixels got removed due to mask

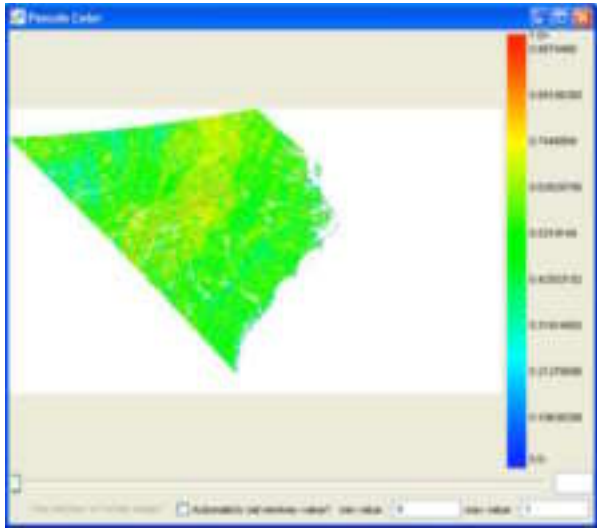
Mosaicking



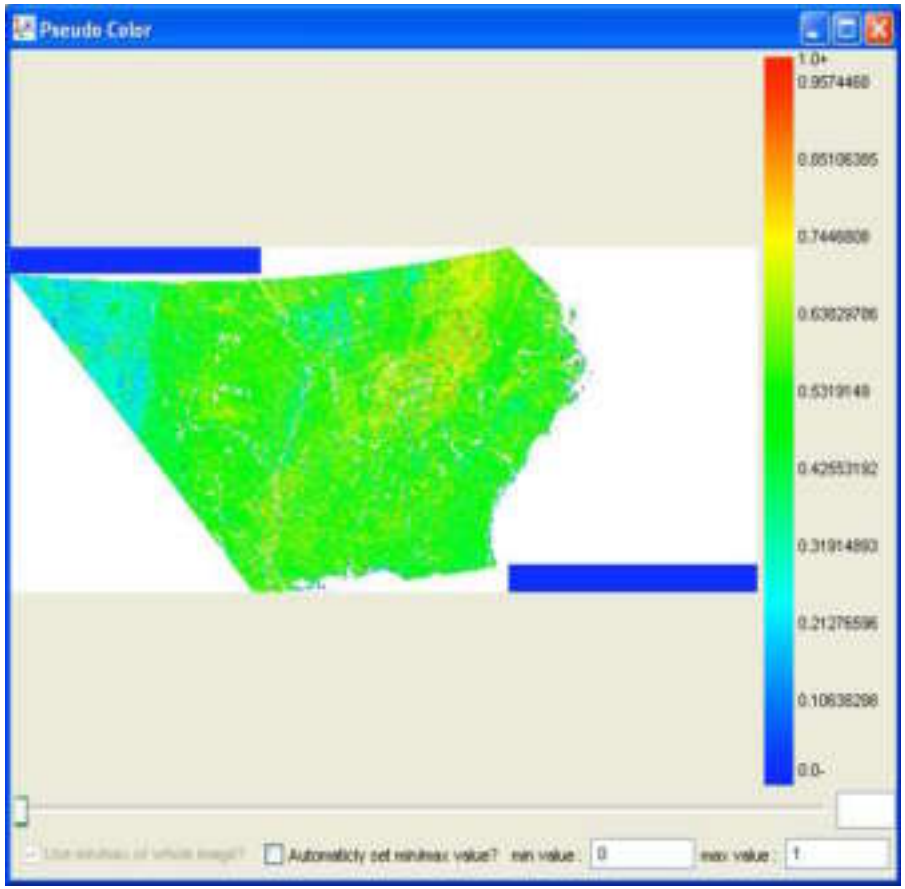
EVI (h10v5)

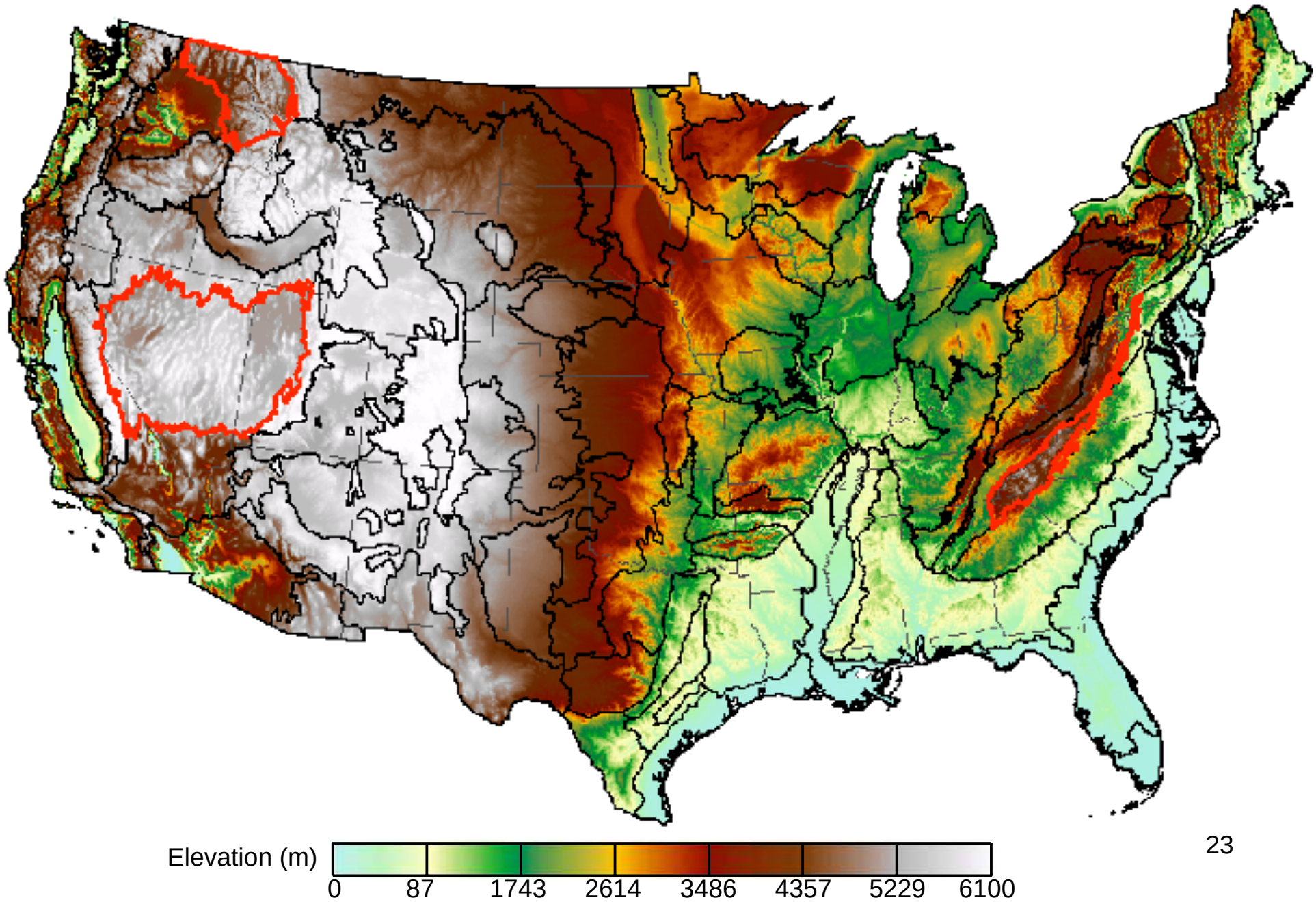


EVI (h11v5)

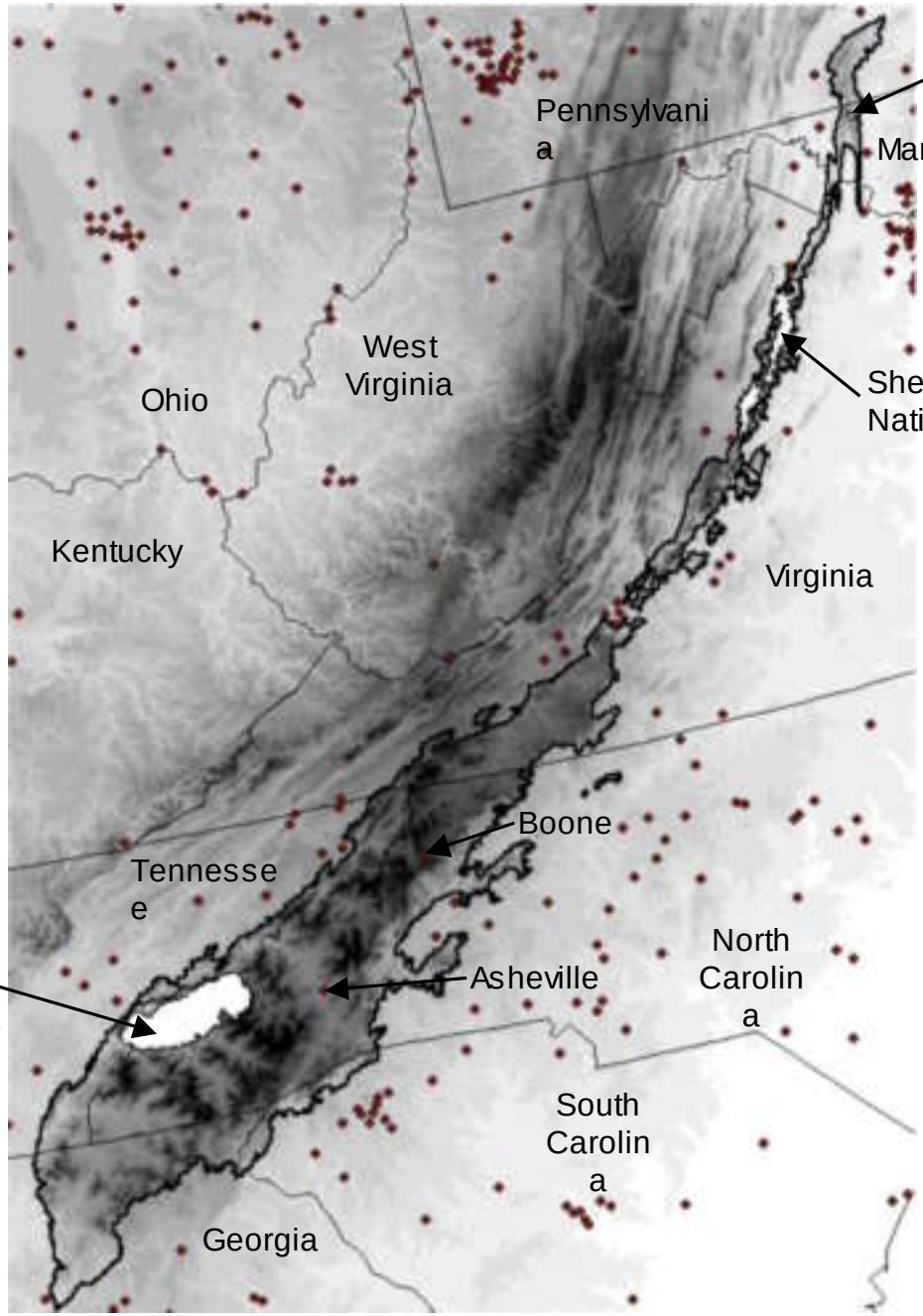


Mosaic EVI




Elevation (m)
0
87
1743
2614
3486
4357
5229
6100

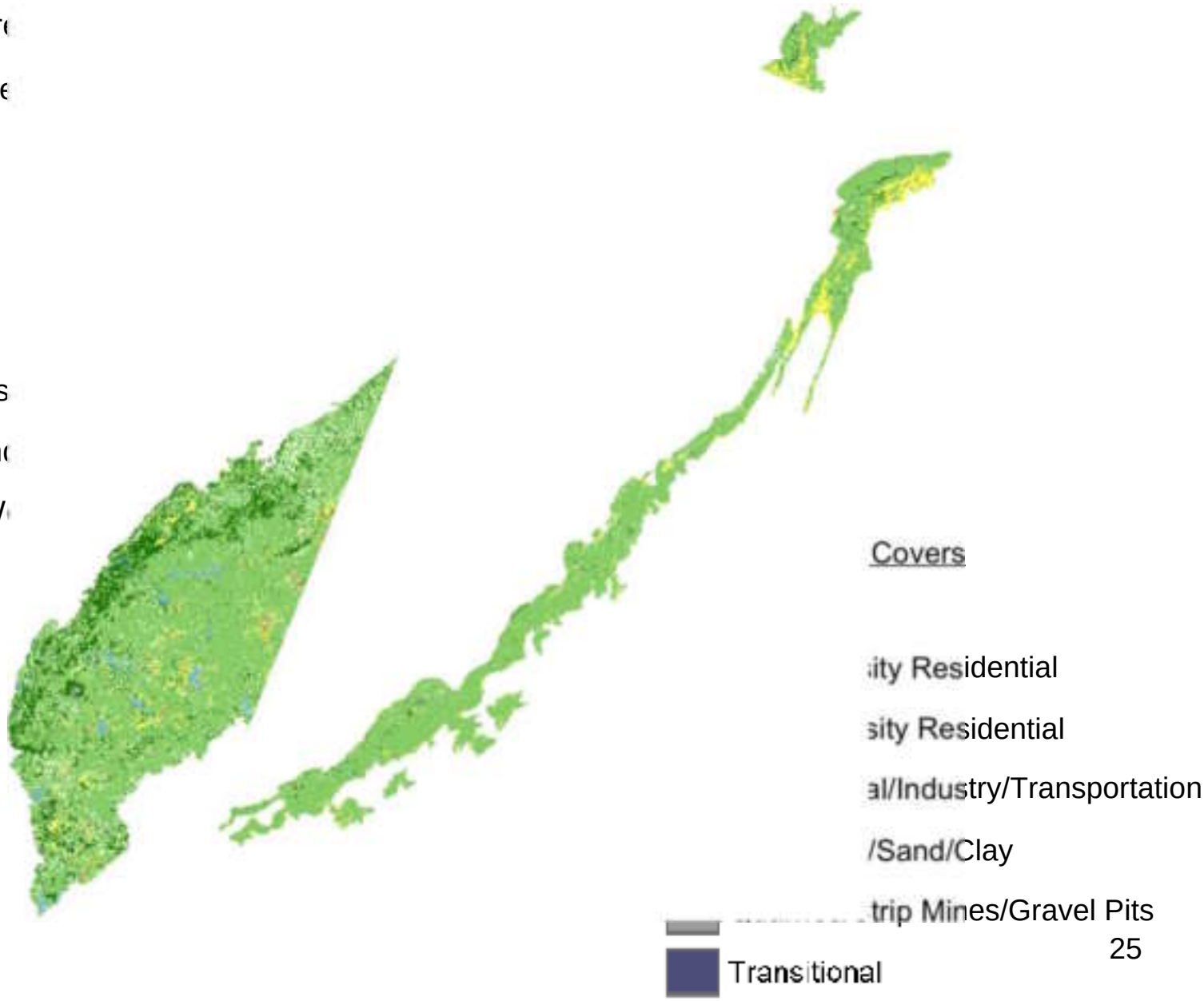
Blue Ridge Ecoregion



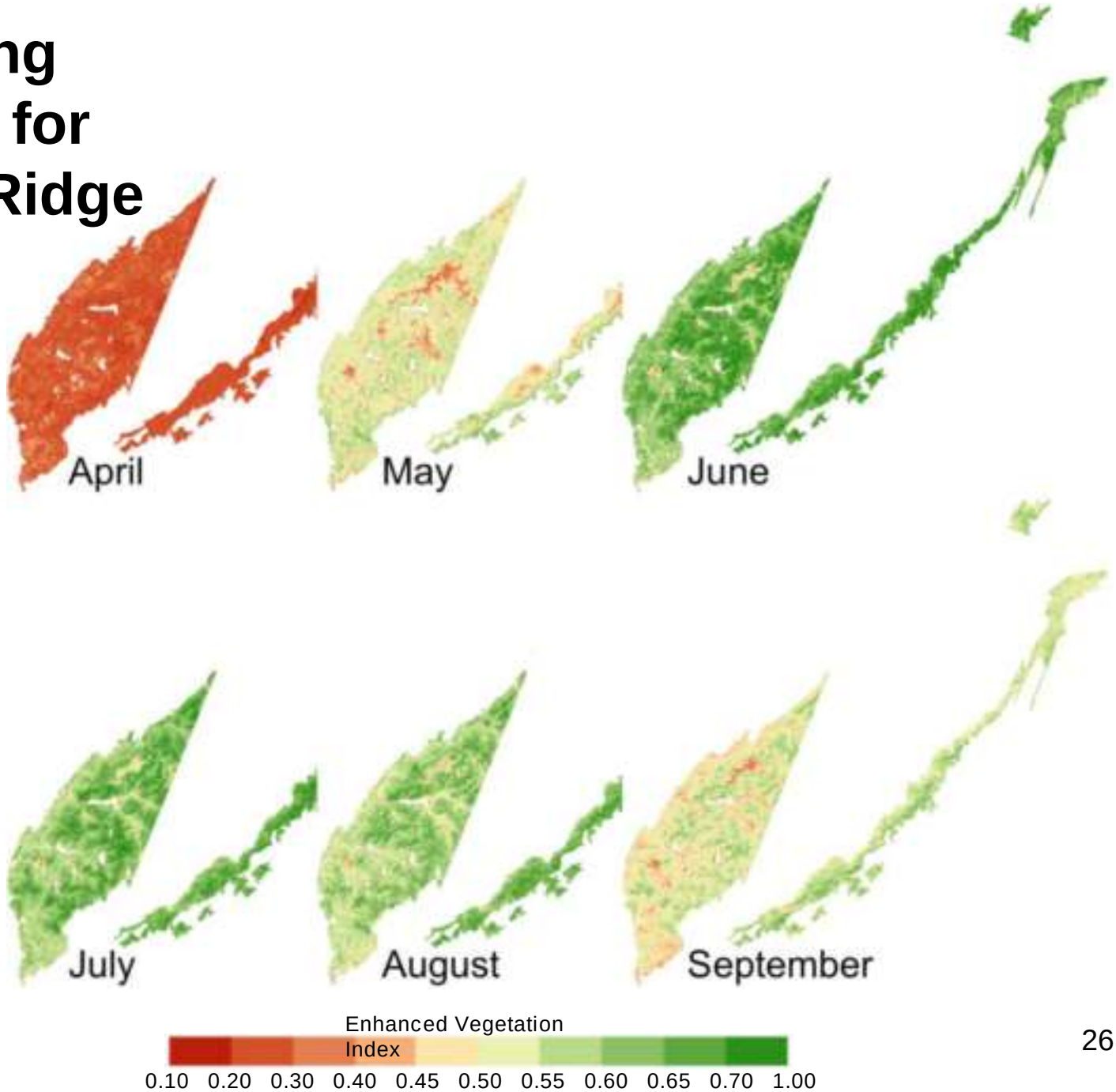
Land Use/Land Cover

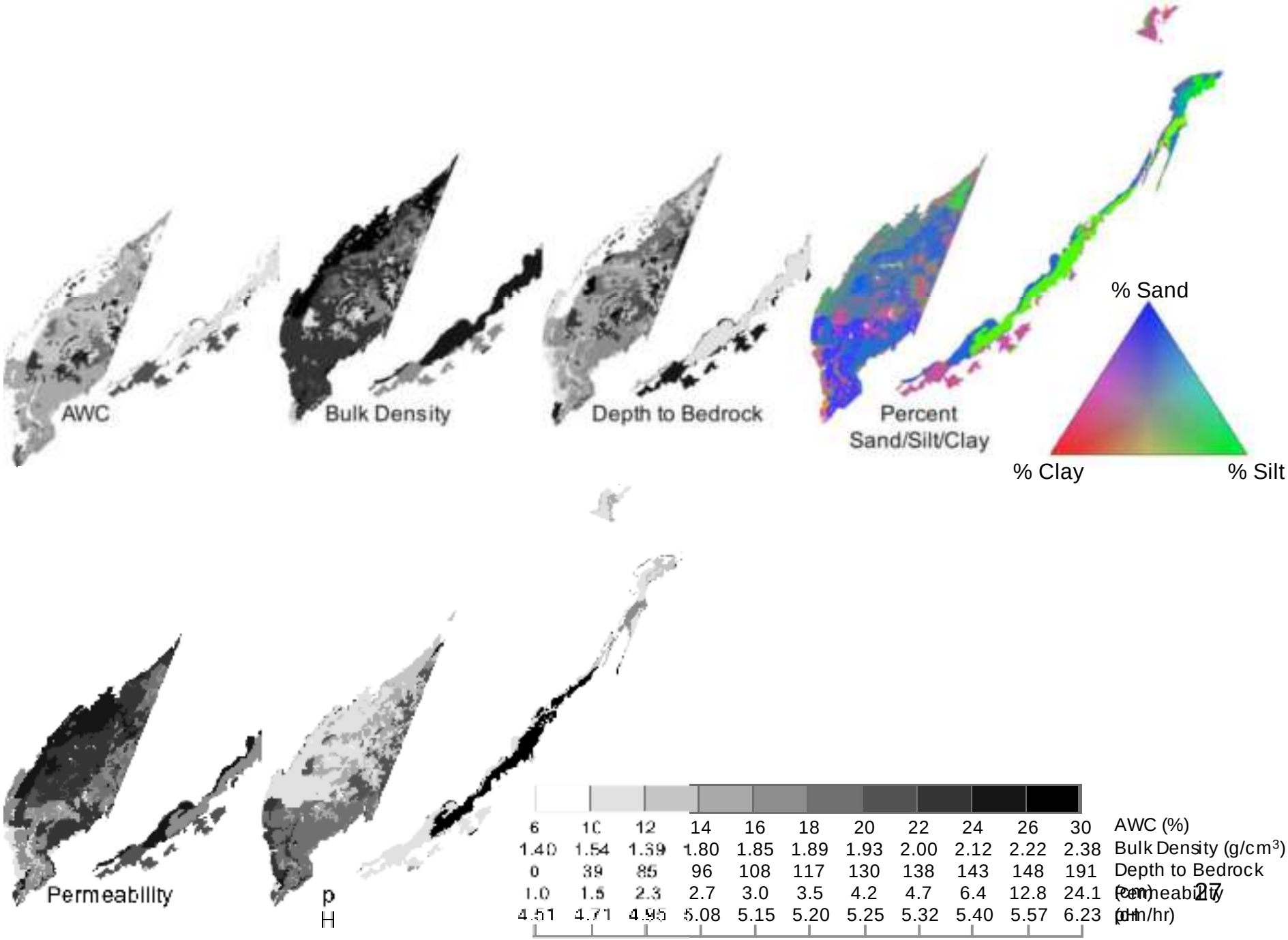
Included Land Covers

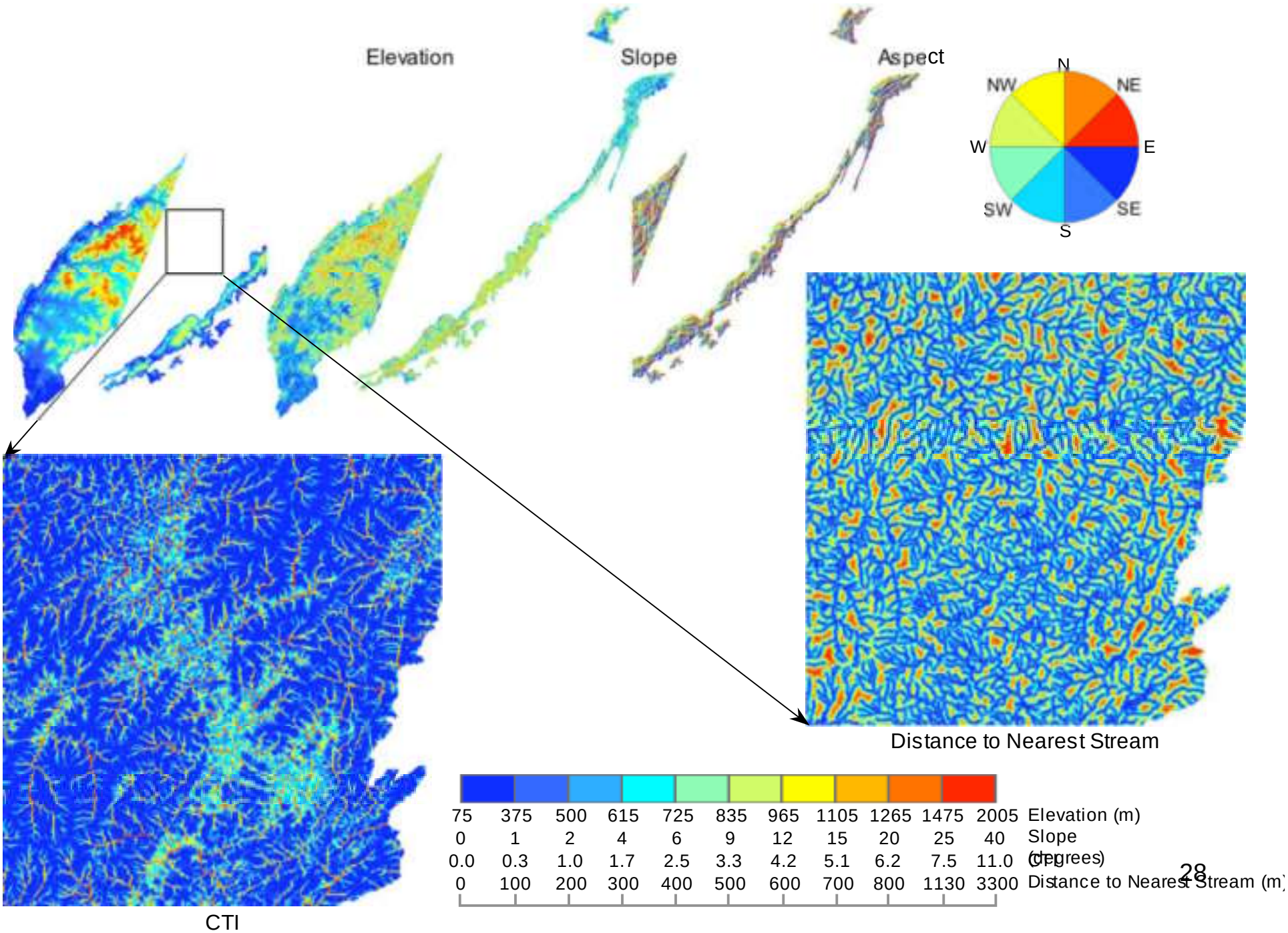
- Deciduous Fore
- Evergreen Fore
- Mixed Forest
- Pasture/Hay
- Row Crops
- Small Grains
- Urban Grasses
- Woody Wetlan
- Herbaceous W

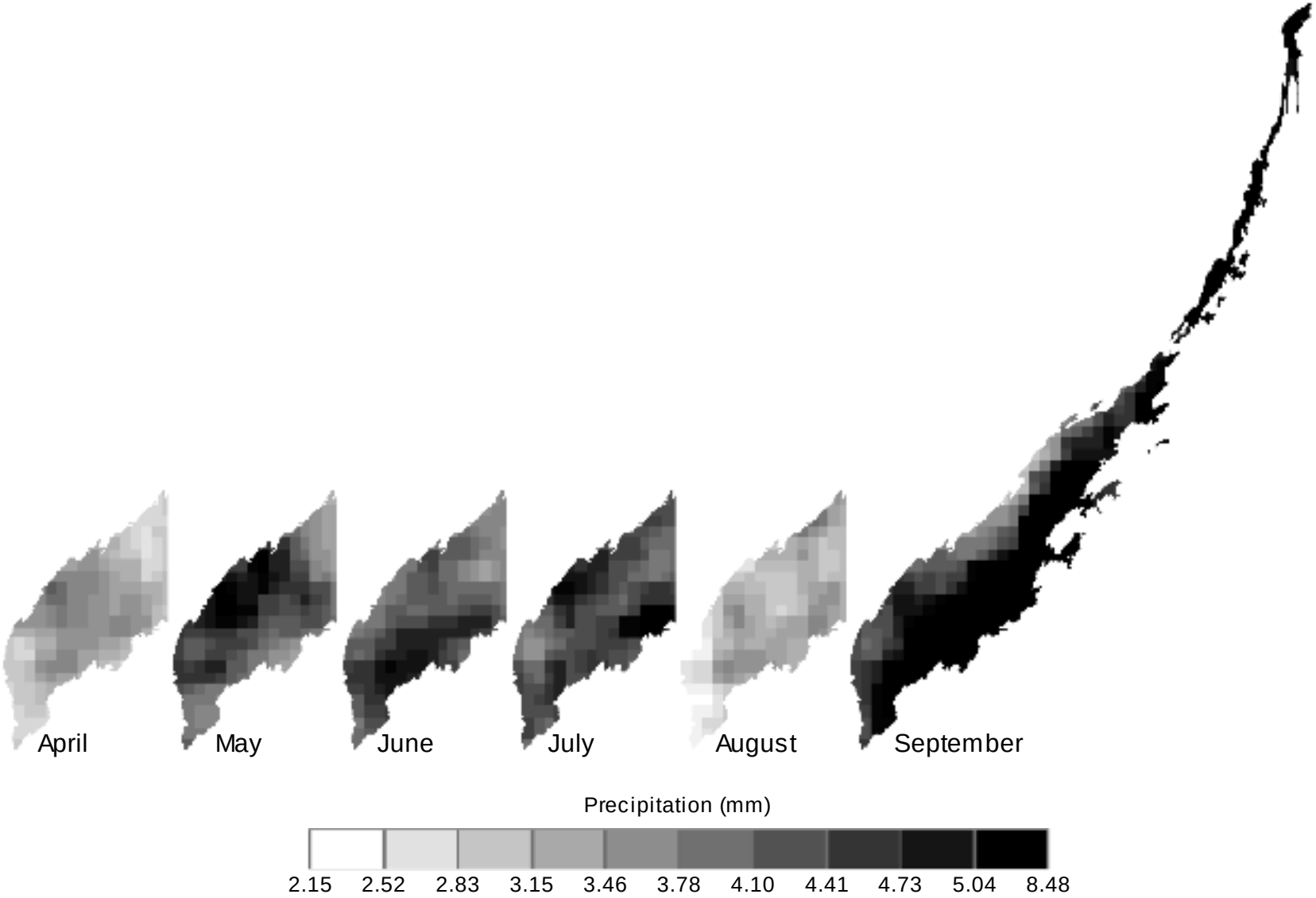


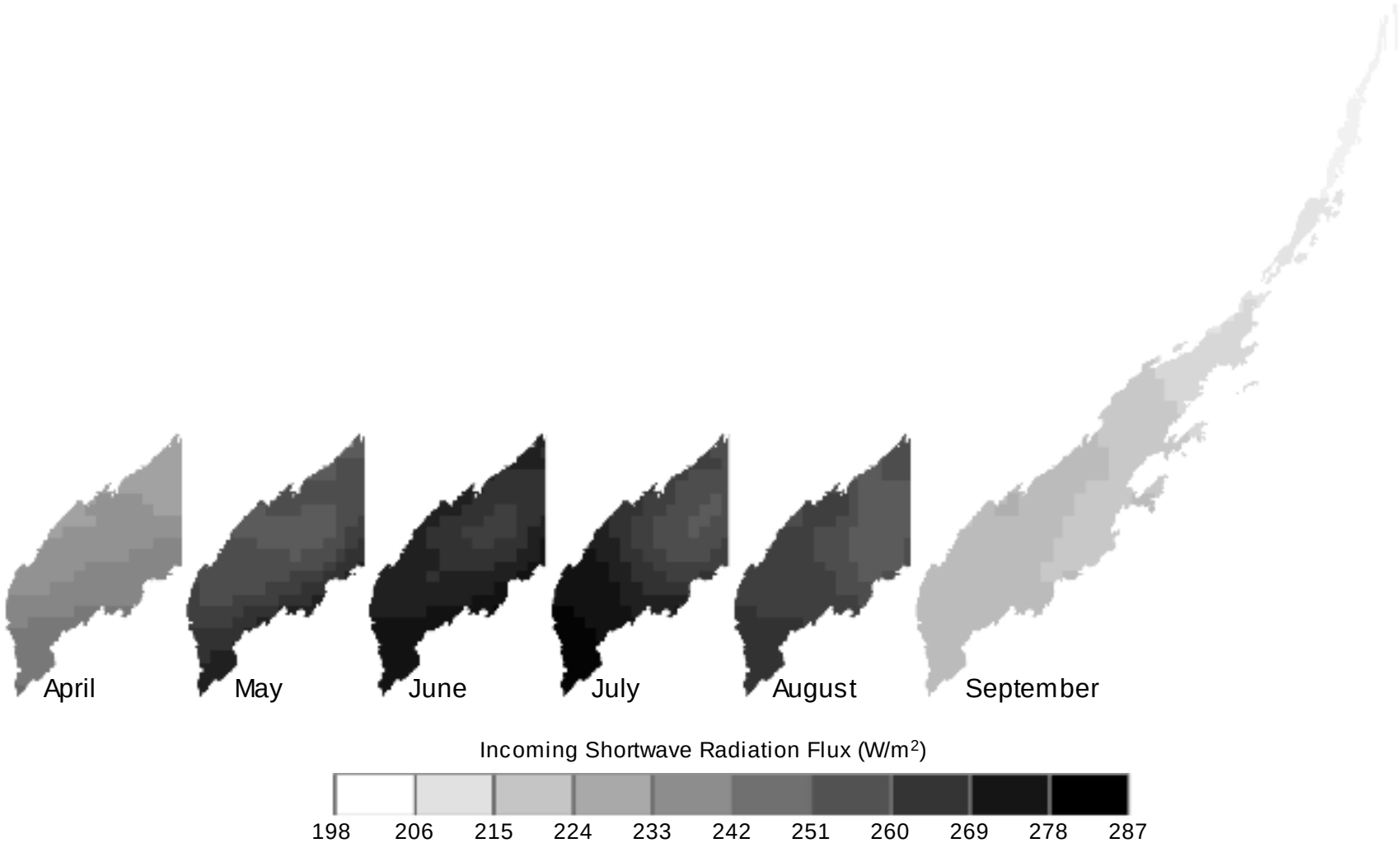
Growing Season for the Blue Ridge

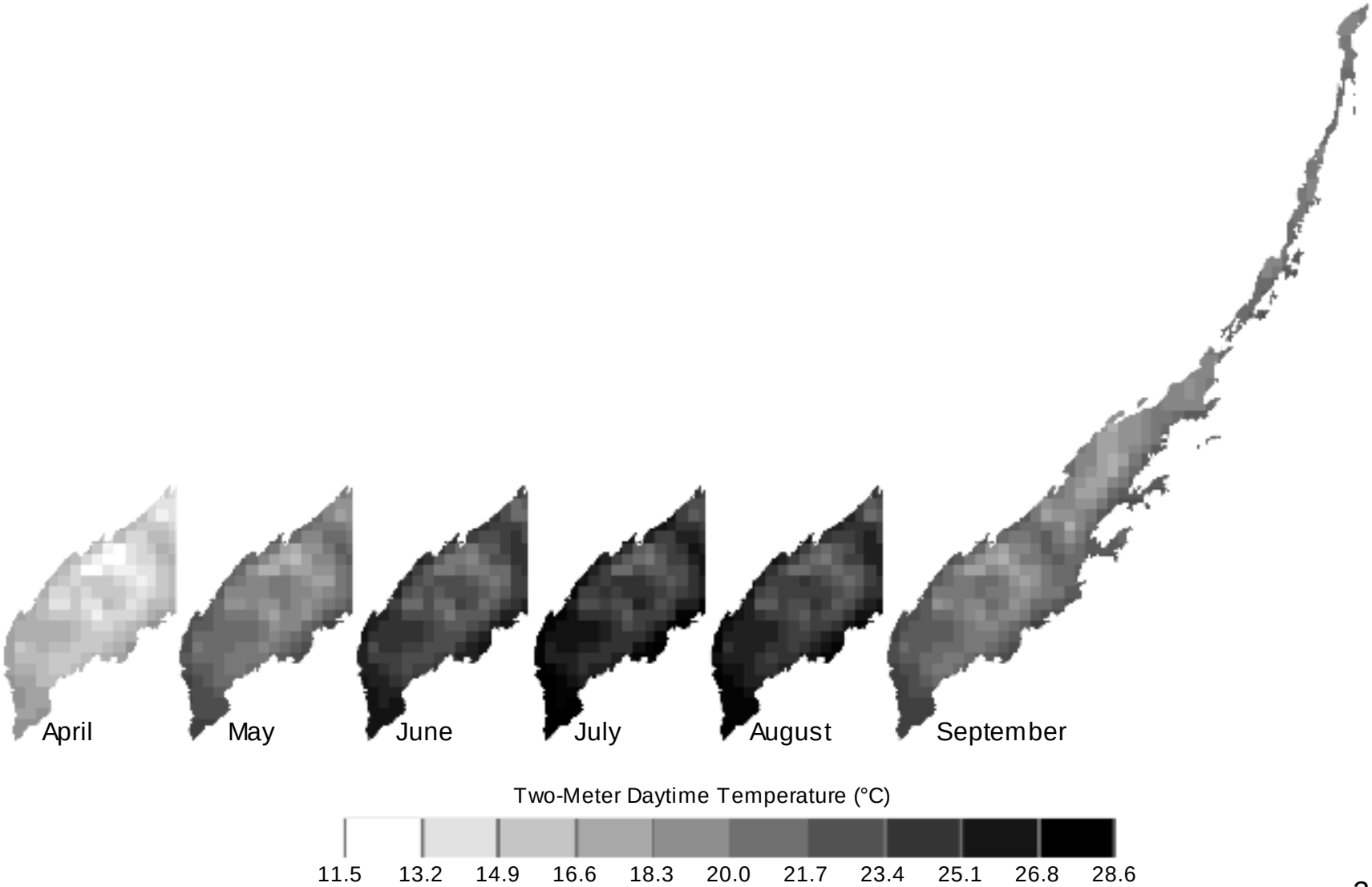



AWC
Bulk Density
Depth to Bedrock
Percent Sand/Silt/Clay
% Sand
% Clay
% Silt
Permeability
pH
6 10 12 14 16 18 20 22 24 26 30 AWC (%)
1.40 1.54 1.69 1.80 1.85 1.89 1.93 2.00 2.12 2.22 2.38 Bulk Density (g/cm³)
0 39 85 96 108 117 130 138 143 148 191 Depth to Bedrock (cm)
1.0 1.6 2.3 2.7 3.0 3.5 4.2 4.7 6.4 12.8 24.1 Permeability (cm/hr)
4.51 4.71 4.95 5.08 5.15 5.20 5.25 5.32 5.40 5.57 6.23 pH


Elevation
Slope
Aspect
N
NE
E
SE
S
SW
W
NW
Distance to Nearest Stream
CTI
75 375 500 615 725 835 965 1105 1265 1475 2005 Elevation (m)
0 1 2 4 6 9 12 15 20 25 40 Slope (degrees)
0.0 0.3 1.0 1.7 2.5 3.3 4.2 5.1 6.2 7.5 11.0 CTI
0 100 200 300 400 500 600 700 800 1130 3300 Distance to Nearest Stream (m)


April
May
June
July
August
September
Precipitation (mm)
2.15
2.52
2.83
3.15
3.46
3.78
4.10
4.41
4.73
5.04
8.48


April
May
June
July
August
September
Incoming Shortwave Radiation Flux (W/m2)
198
206
215
224
233
242
251
260
269
278
287


April
May
June
July
August
September
Two-Meter Daytime Temperature (°C)
11.5
13.2
14.9
16.6
18.3
20.0
21.7
23.4
25.1
26.8
28.6

Regression tree induction is used to model the dependence of the vegetation greenness on physiography and hydrology

INPUTS

Elevation
Slope
Aspect
Distance to nearest stream
Compound topographic index

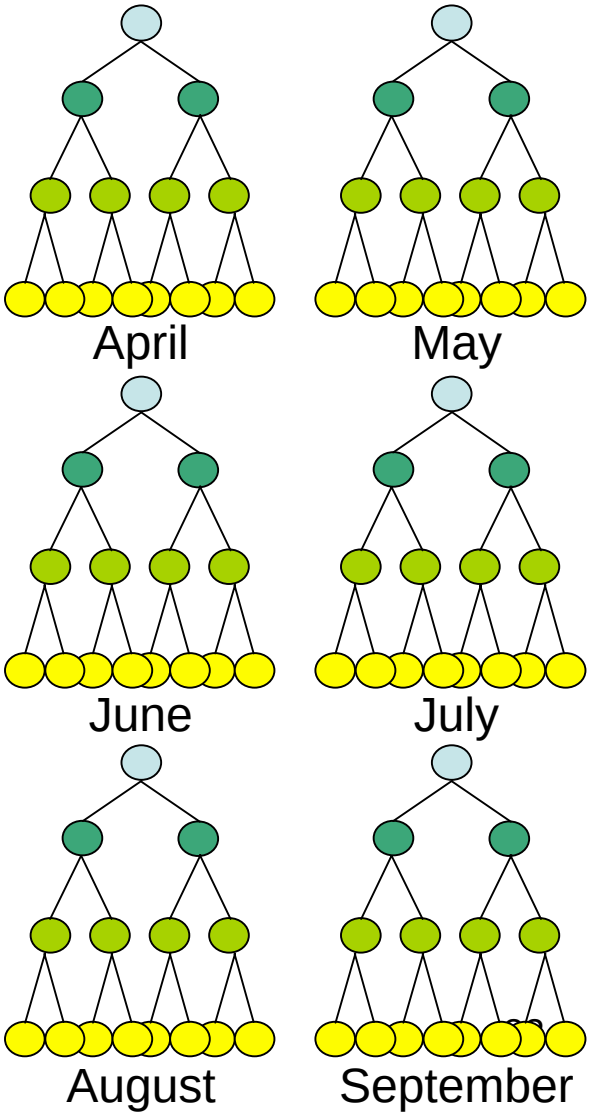
Percent sand/silt/clay
Bulk density
Permeability
Available water content
Bedrock depth
pH

Precipitation
Shortwave radiation
Longwave radiation
Daytime air temperature
Nighttime air temperature

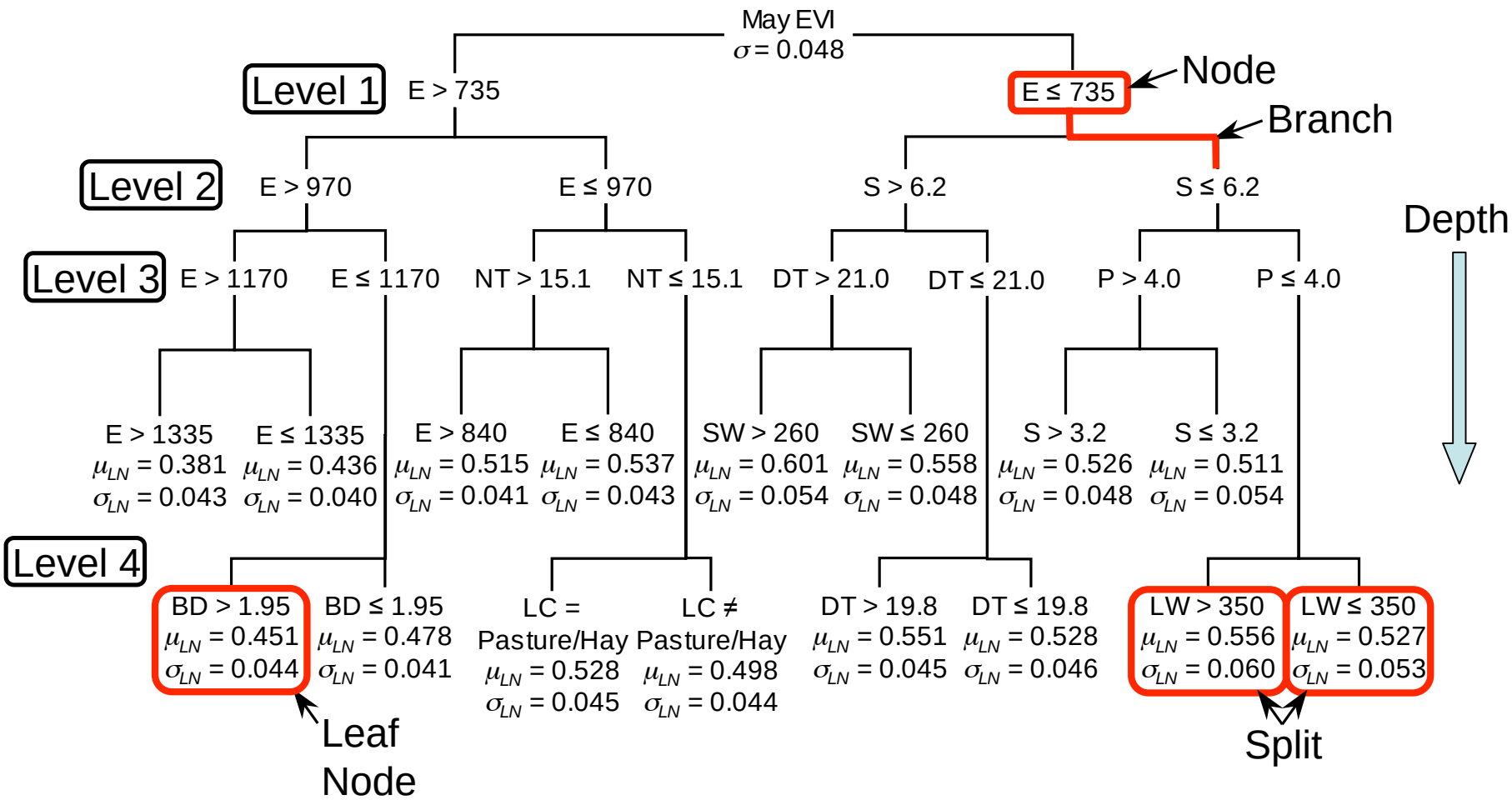
Land cover

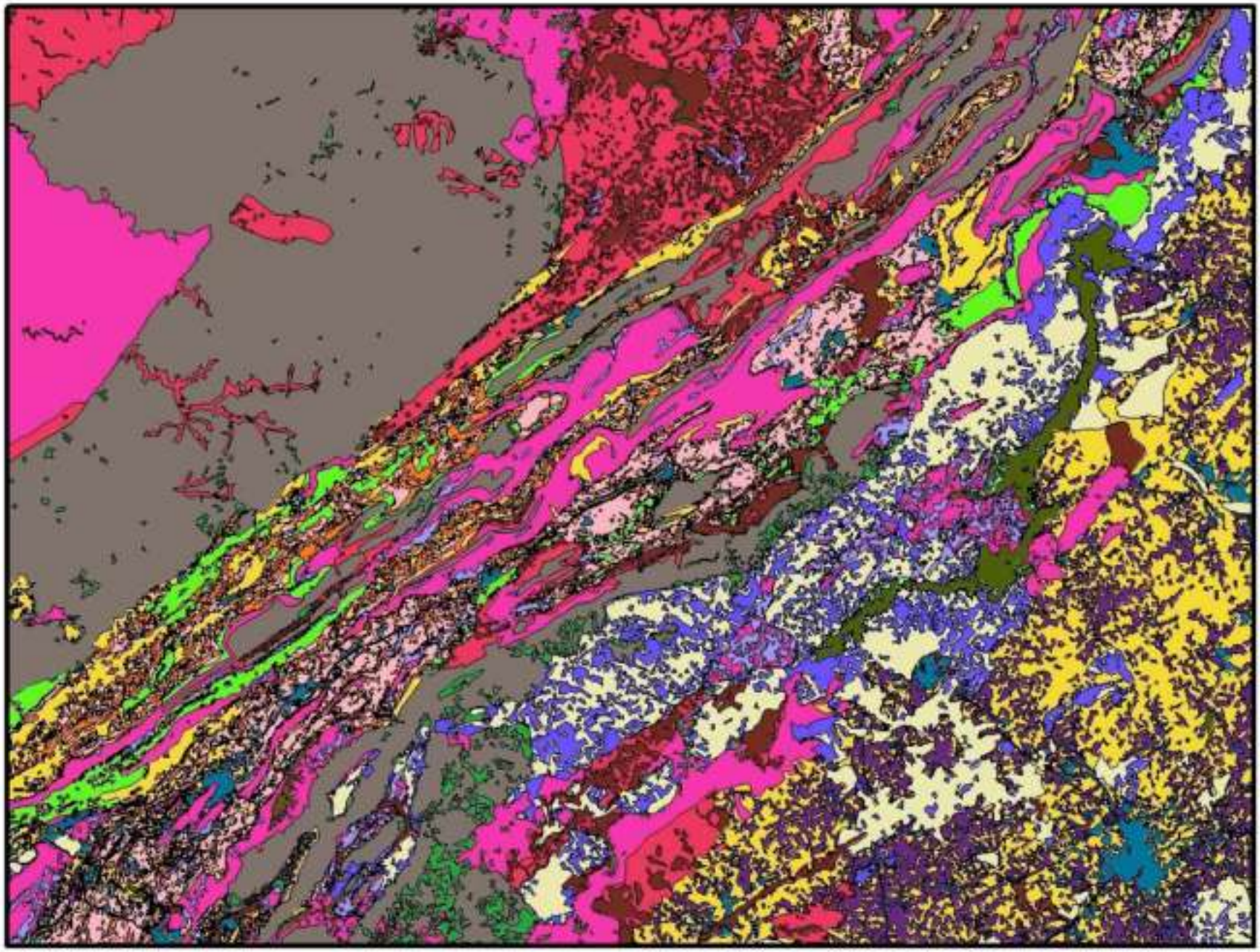
OUTPUT

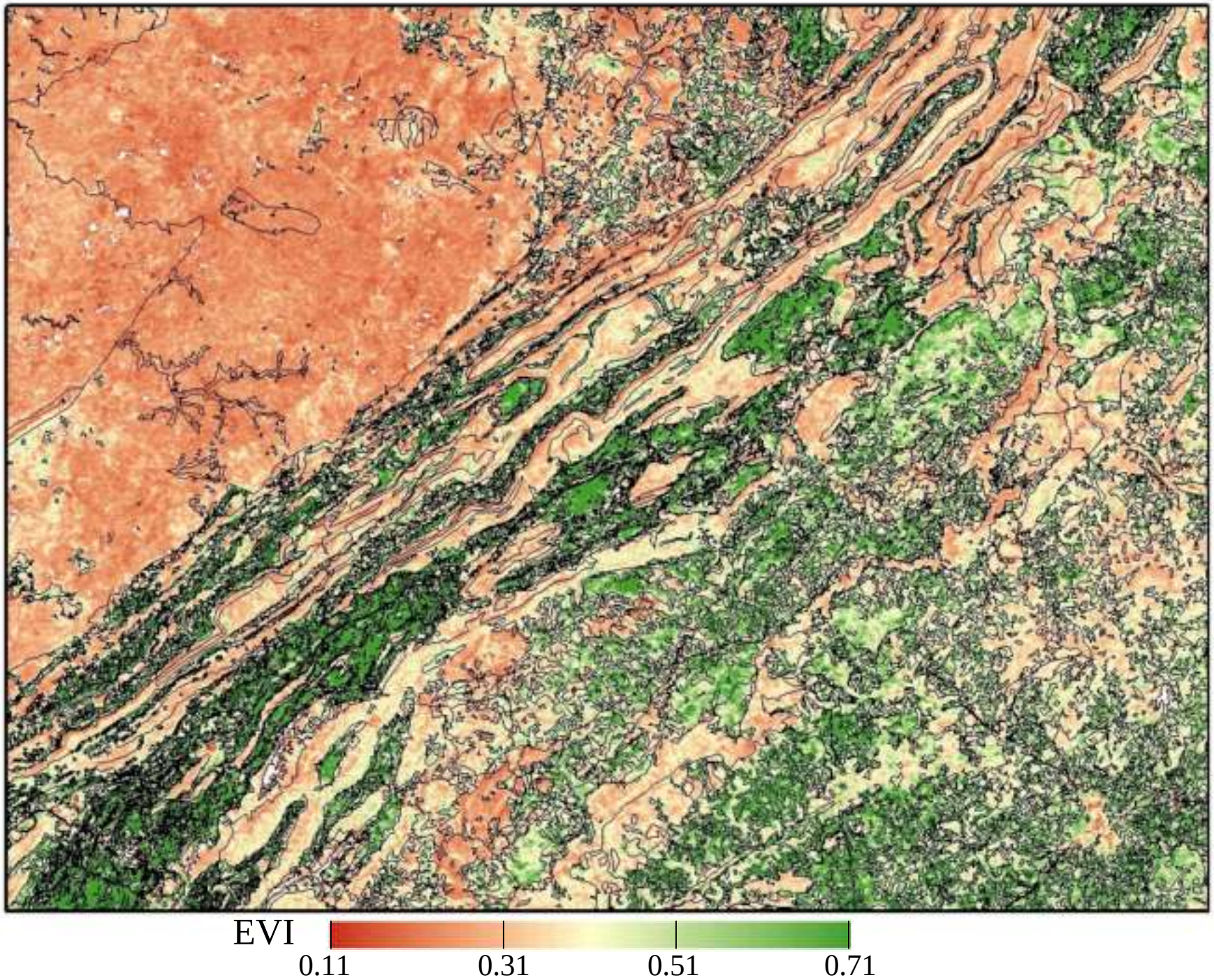
EVI

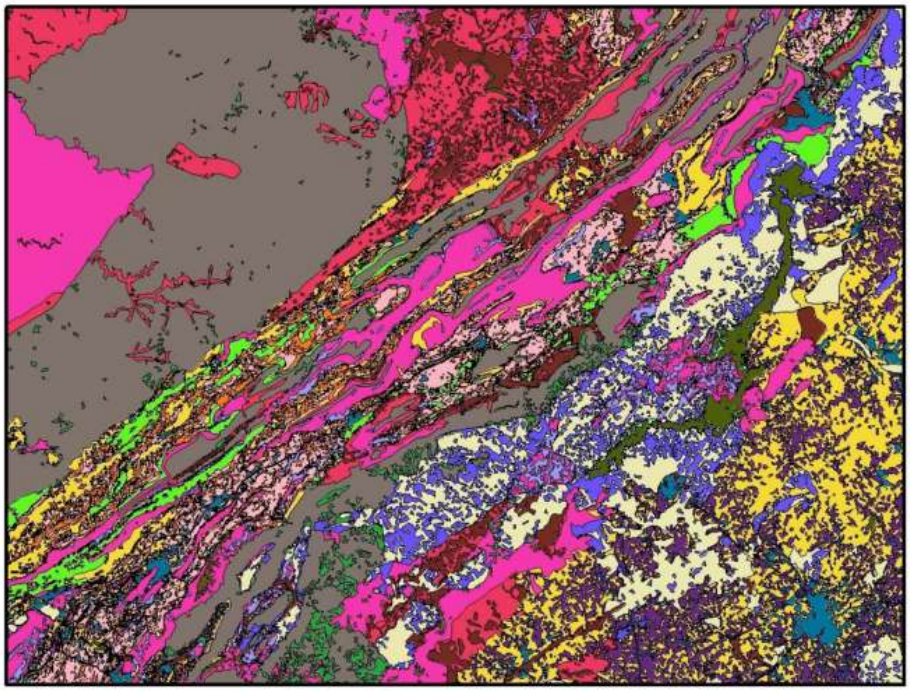


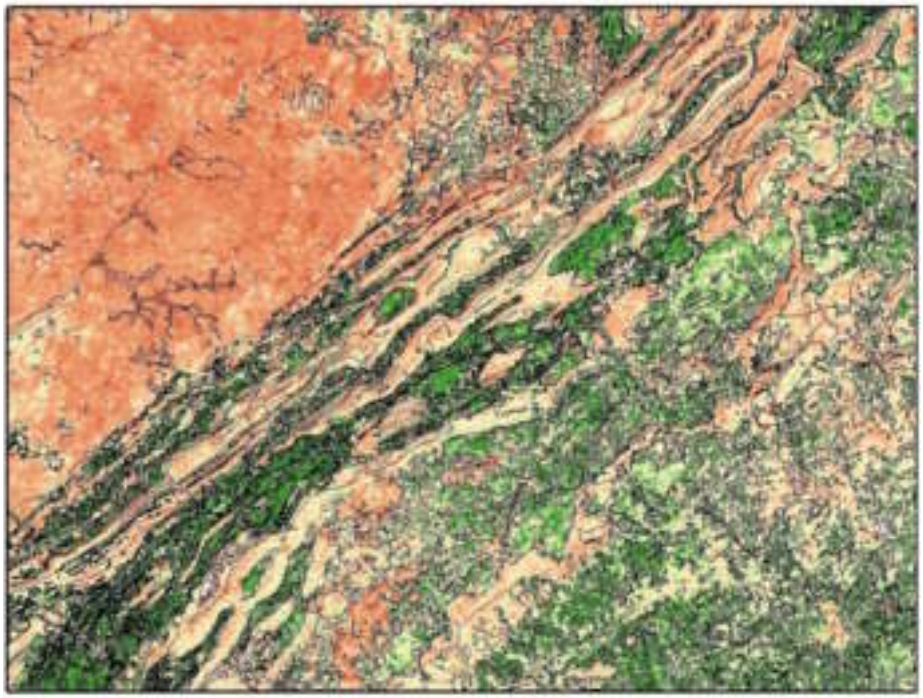
Regression tree terminology



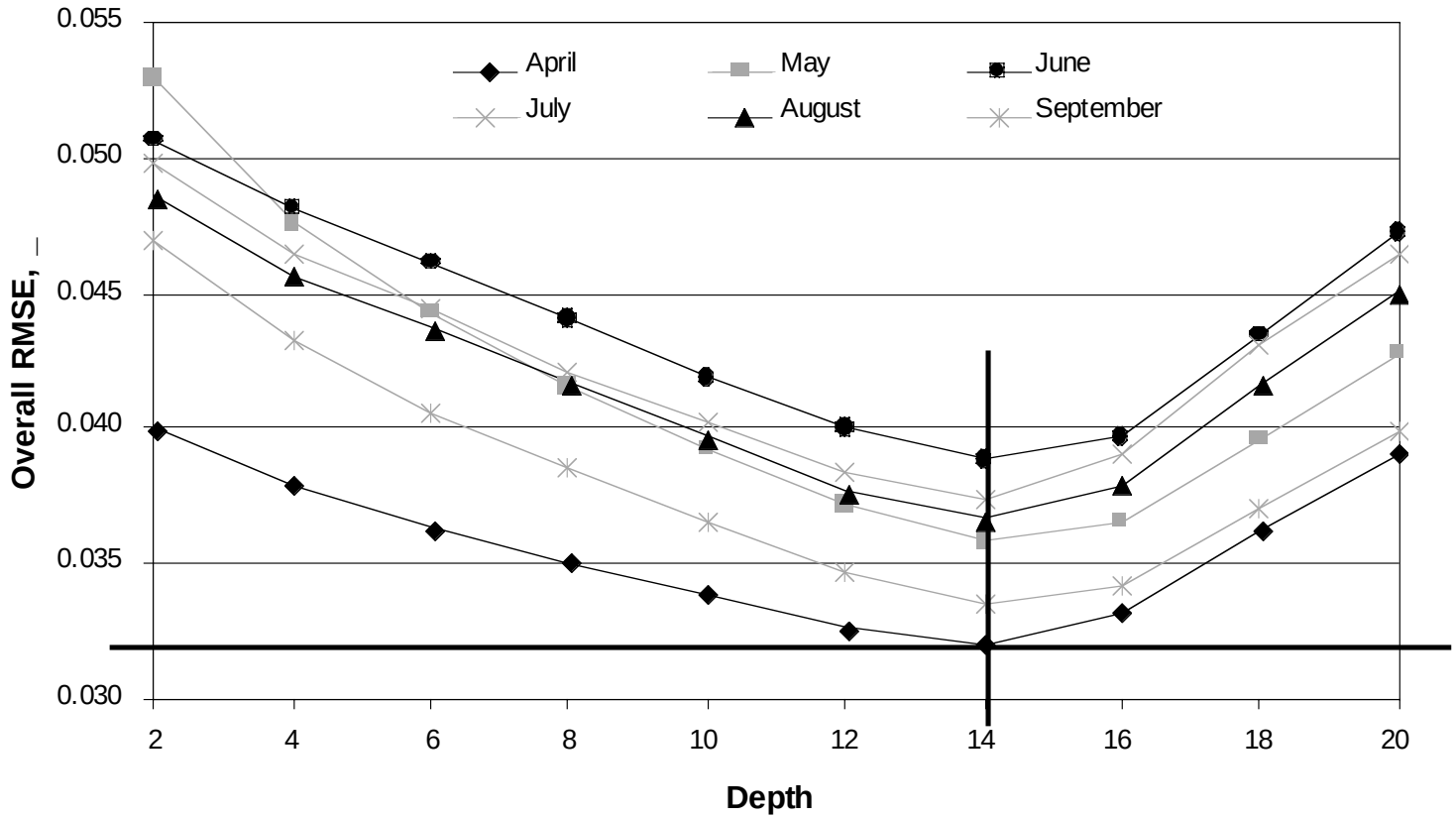



EVI
0.11
0.31
0.51
0.71

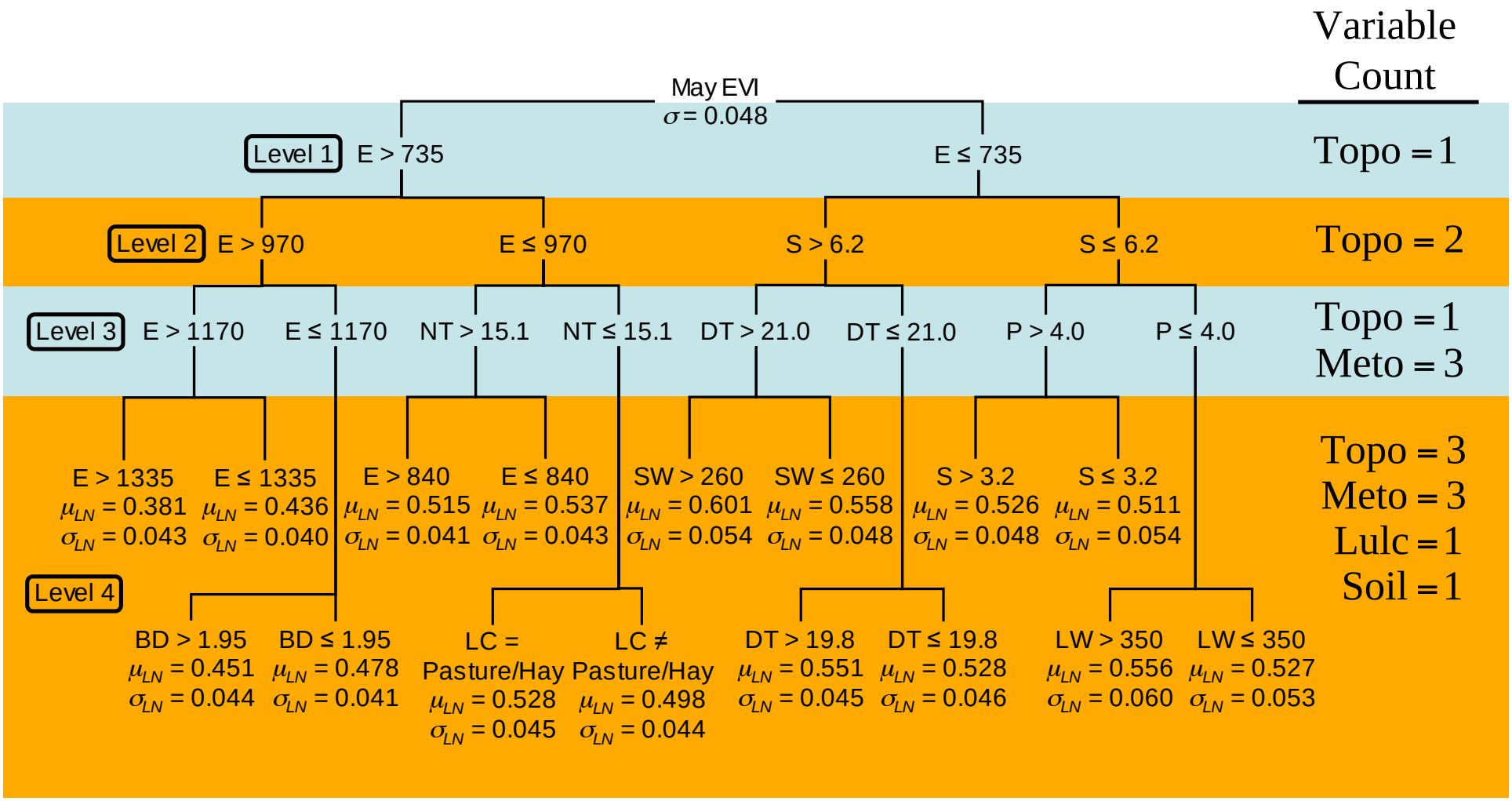




Overall tree error vs. depth



To compute relevance of an attribute…



$$R_v = \sum_{i=1}^{D} \frac{n}{2^{i-1} i} \quad \Rightarrow \quad RD_v = \frac{R_v}{\sum_V R_v}$$

$$R_{\text{topo}} = \frac{1}{(2^0)(1)} + \frac{2}{(2^1)(2)} + \frac{1}{(2^2)(3)} + \frac{3}{(2^3)(4)} = 1.677$$

$$R_{\text{soil}} = \frac{1}{(2^3)(4)} = 0.031$$

$$R_{\text{lulc}} = \frac{1}{(2^3)(4)} = 0.031$$

$$R_{\text{mete}} = \frac{3}{(2^2)(3)} + \frac{3}{(2^3)(4)} = 0.344$$

$$\sum = 2.083$$

$$RD_{\text{topo}} = \frac{1.677}{2.083} \Rightarrow 80.5\%$$

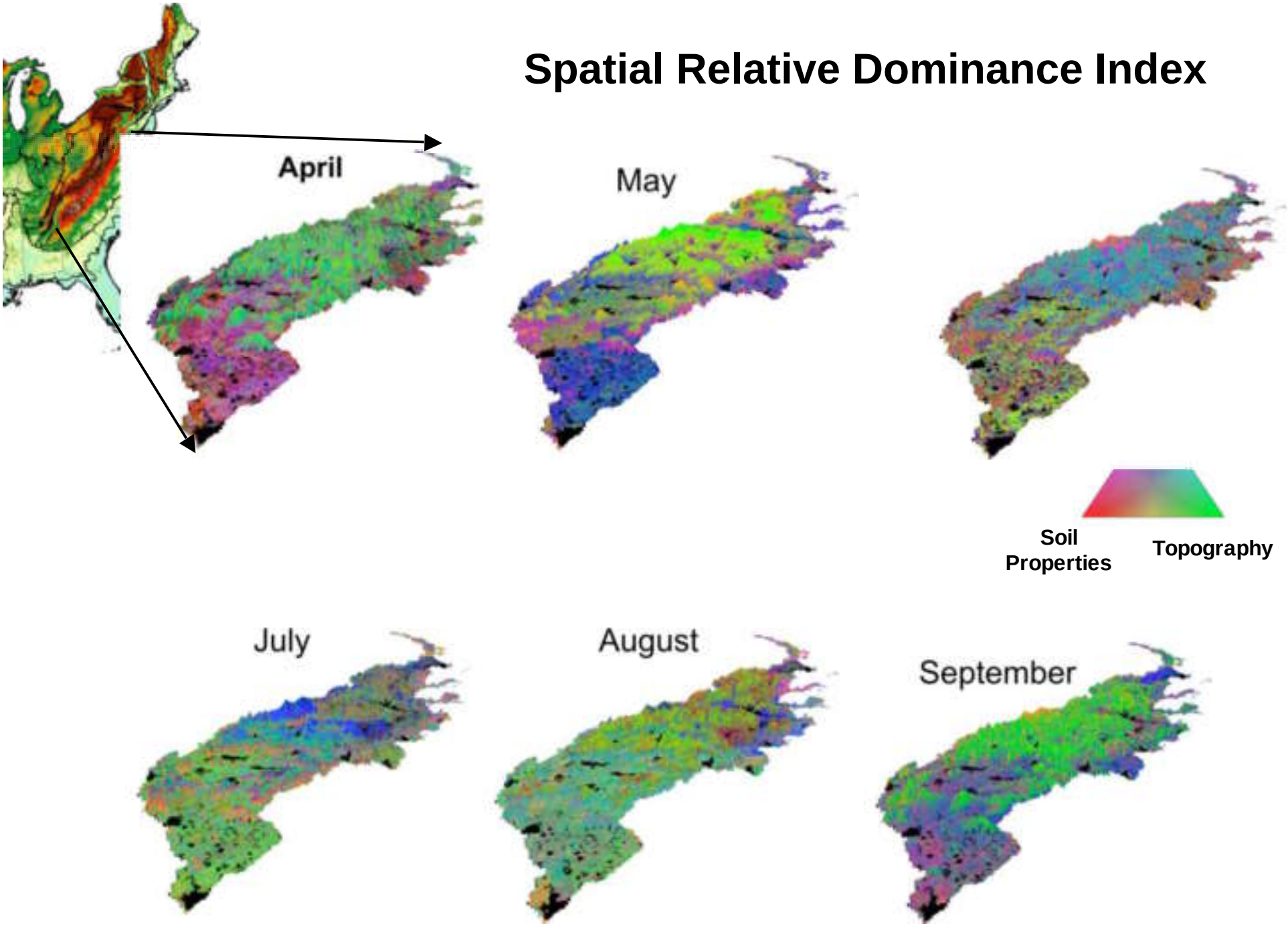
$$RD_{\text{soil}} = \frac{0.031}{2.083} \Rightarrow 1.5\%$$

$$RD_{\text{lulc}} = \frac{0.031}{2.083} \Rightarrow 1.5\%$$

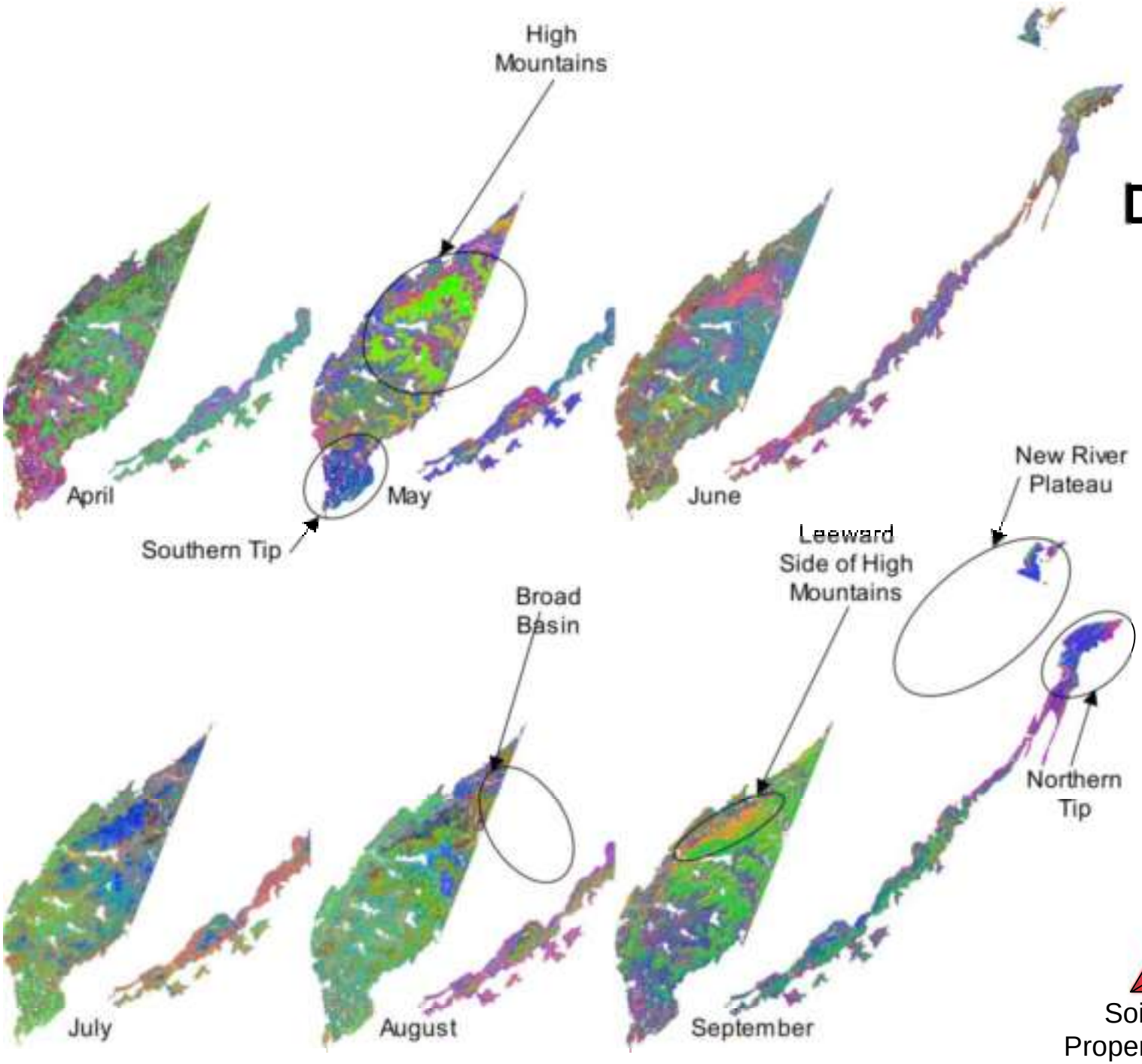
$$RD_{\text{mete}} = \frac{0.344}{2.083} \Rightarrow 16.5\%$$

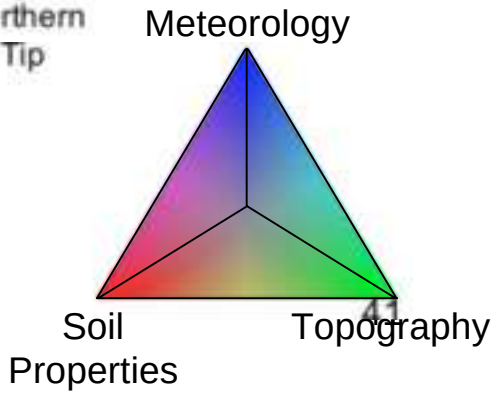
$$\sum = 100.0\%$$

Spatial Relative Dominance Index

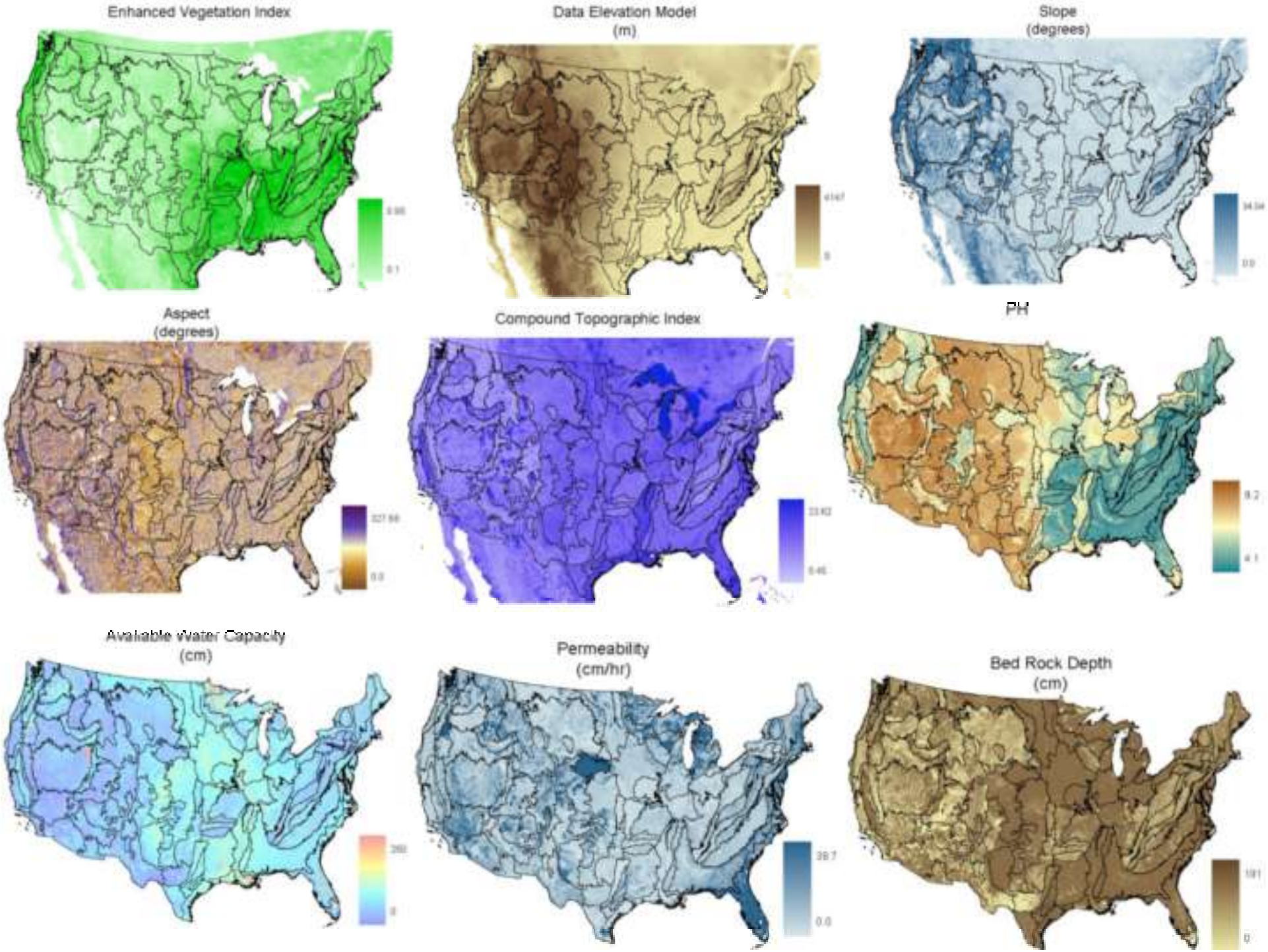


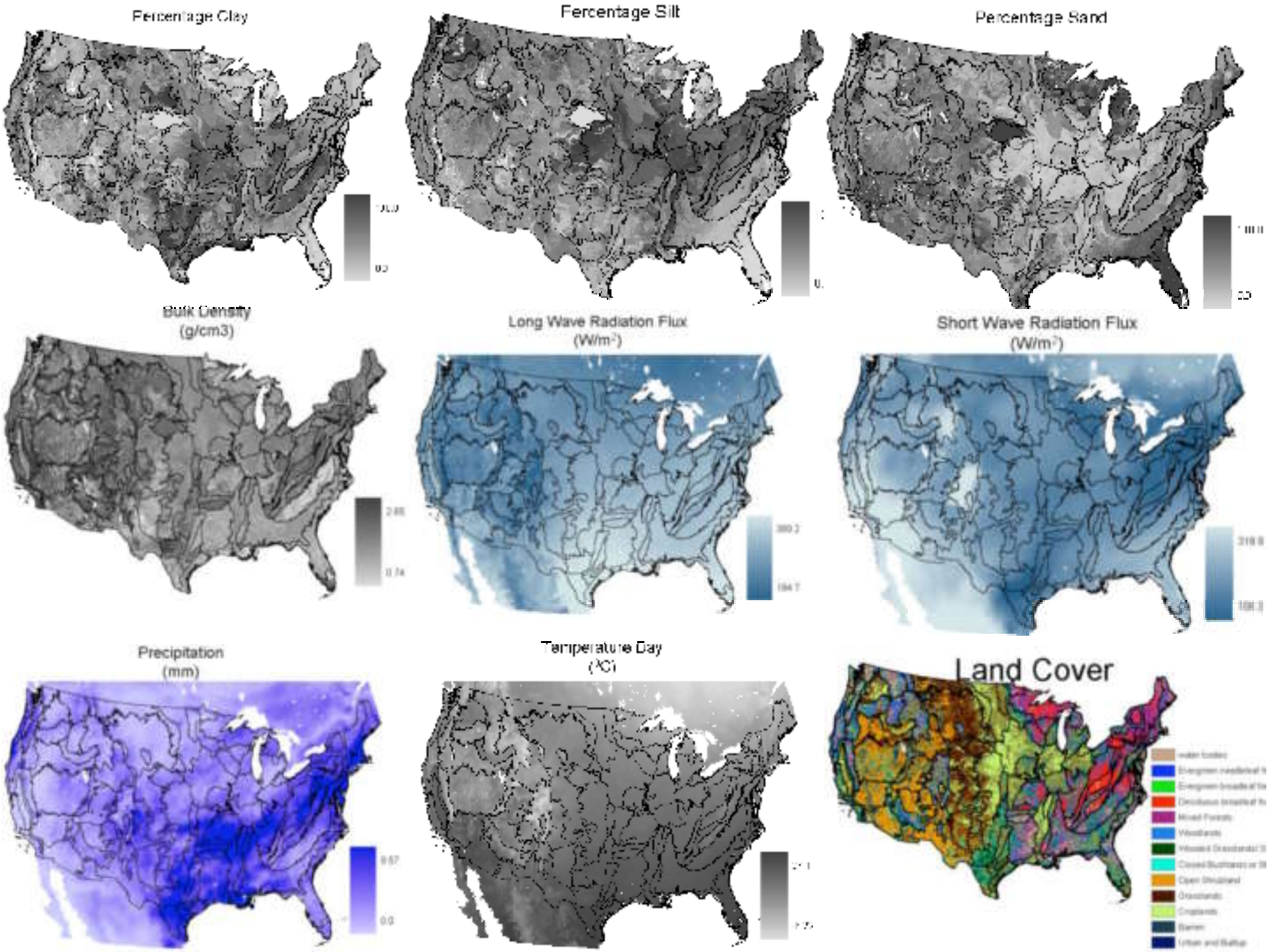
Spatial Relative Dominance Index



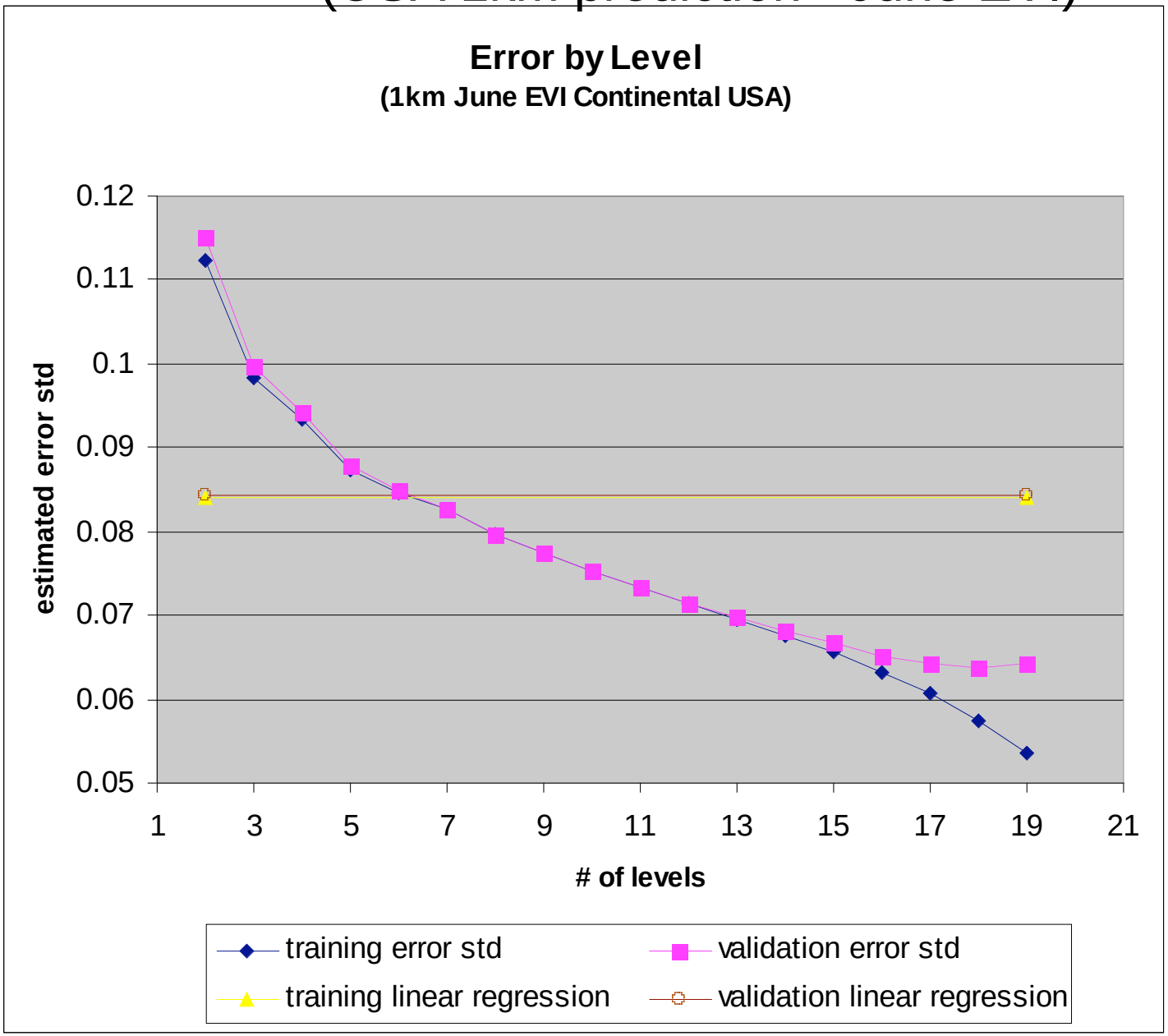


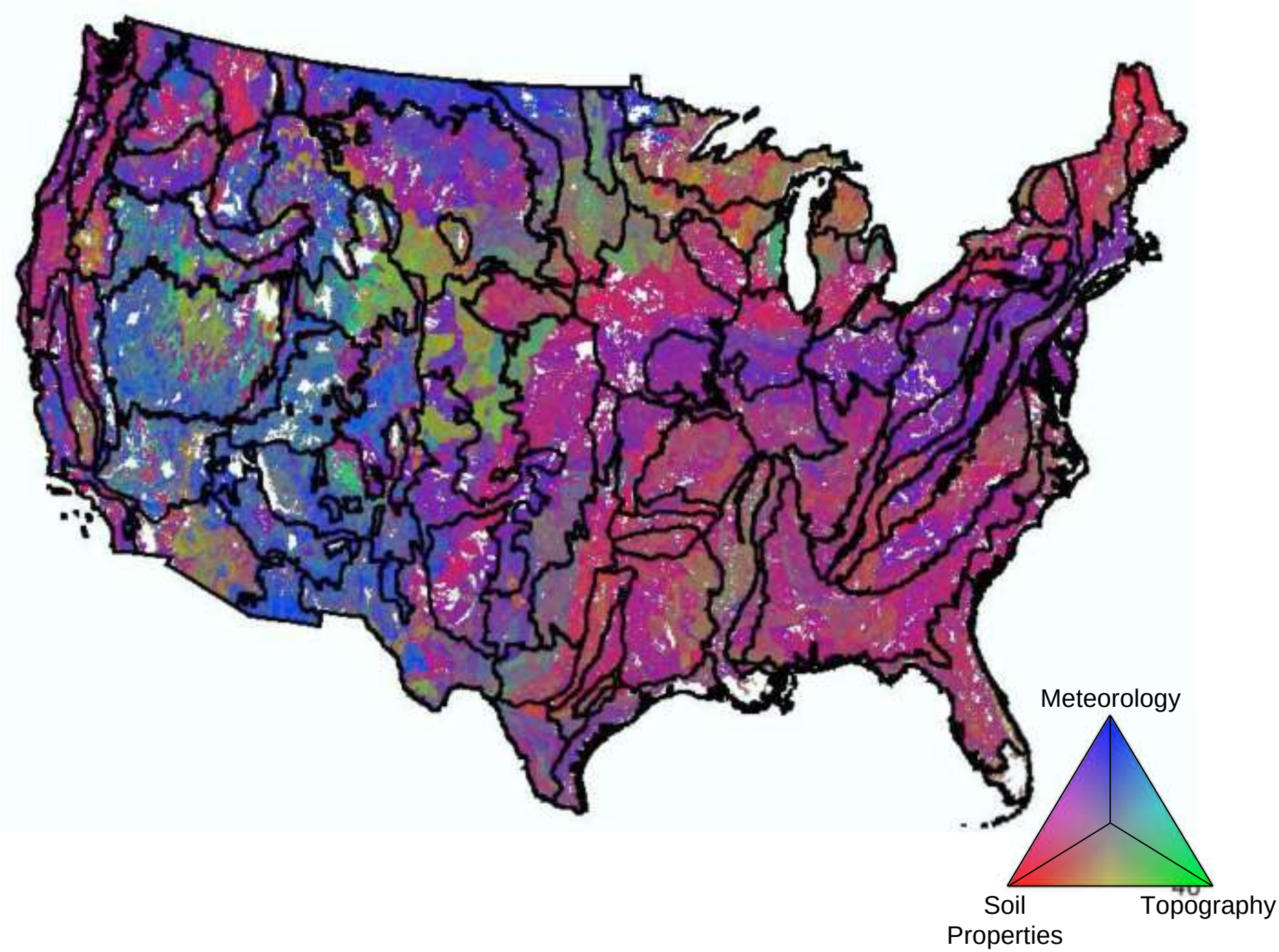
Data for Continental US Analysis


Enhanced Vegetation Index
Data Elevation Model
(m)
Slope
(degrees)
Aspect
(degrees)
Compound Topographic Index
PH
Available Water Capacity
(cm)
Permeability
(cm/hr)
Bed Rock Depth
(cm)


Percentage Clay
Percentage Silt
Percentage Sand
Bulk Density
(g/cm3)
Long Wave Radiation Flux
(W/m²)
Short Wave Radiation Flux
(W/m²)
Precipitation
(mm)
Temperature Day
(°C)
Land Cover

Scalable Regression Trees: Summary Performance Statistics (USA 1km prediction - June EVI)



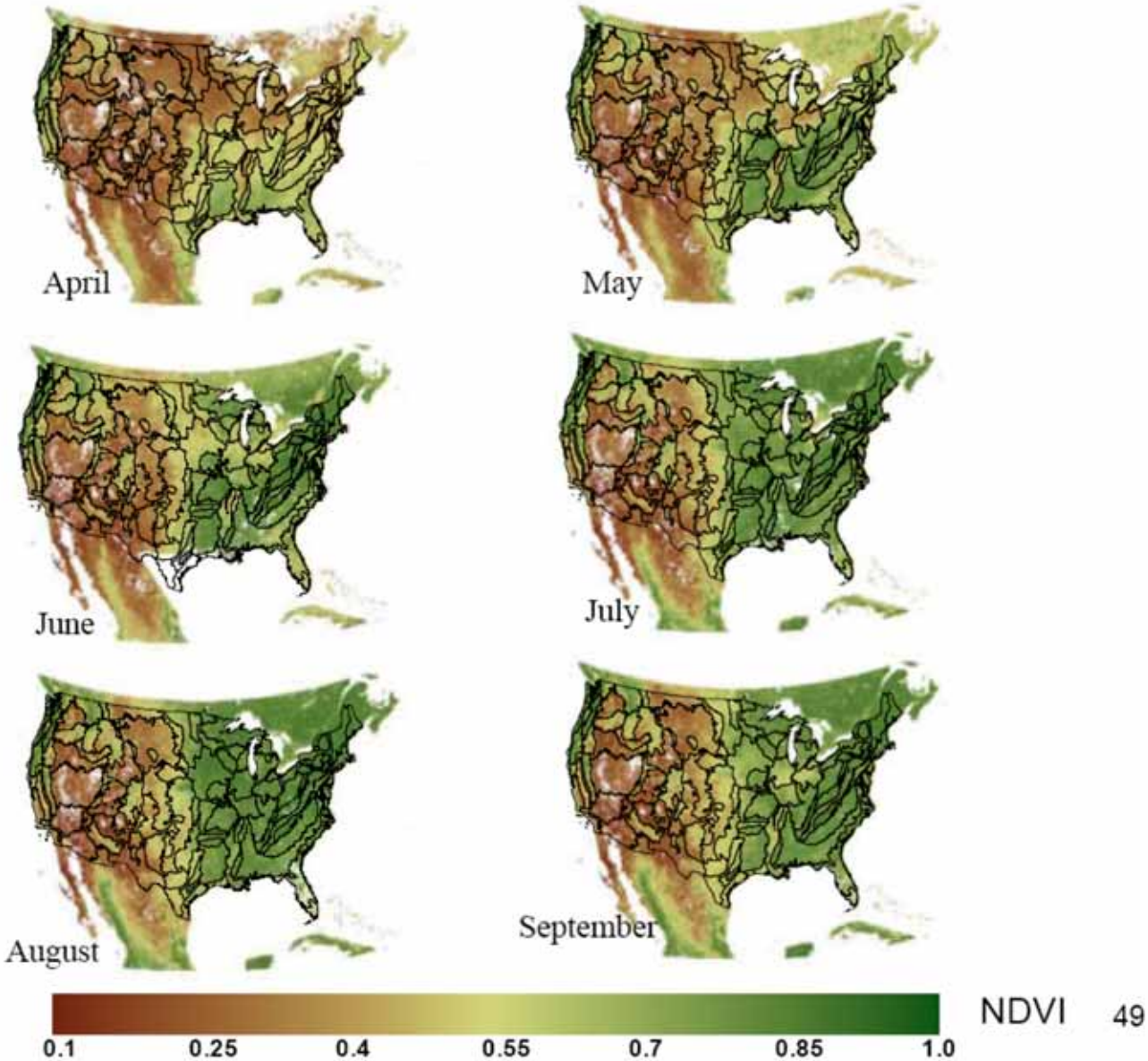

Meteorology
Soil Properties
Topography

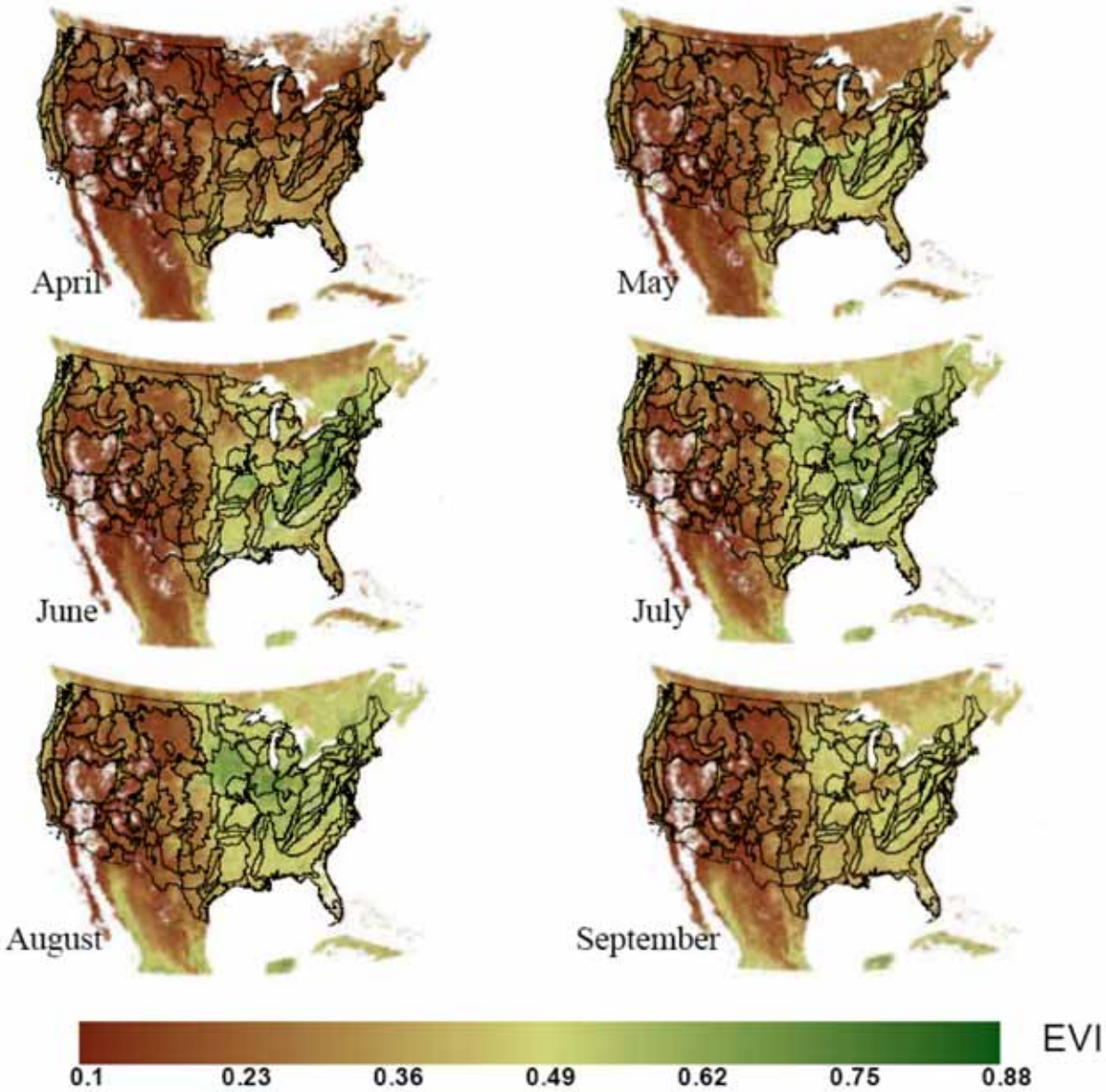
Clustering Analysis over Continental US (1 KM resolution)

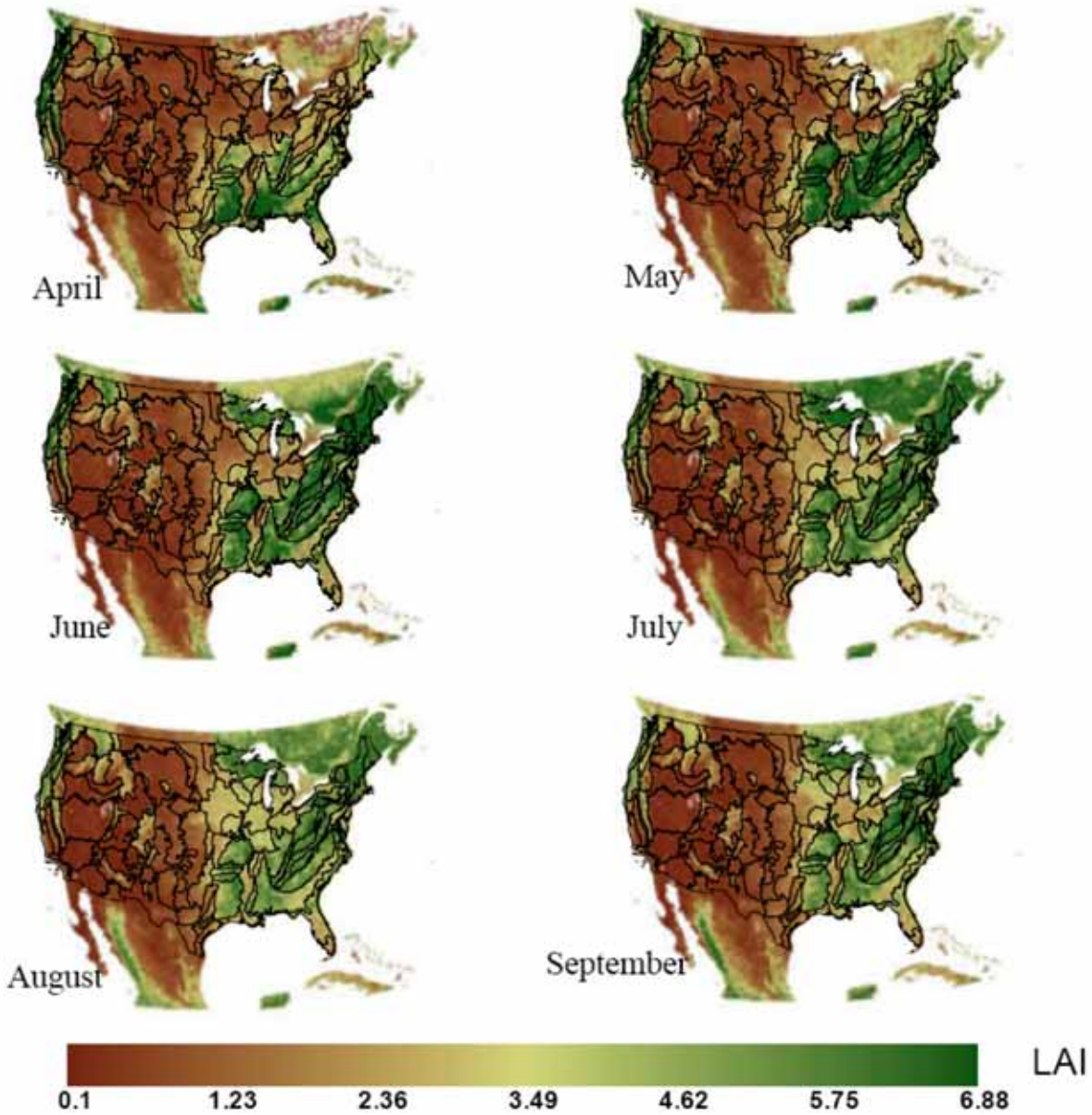
MODIS Data Spatial and Temporal Resolution

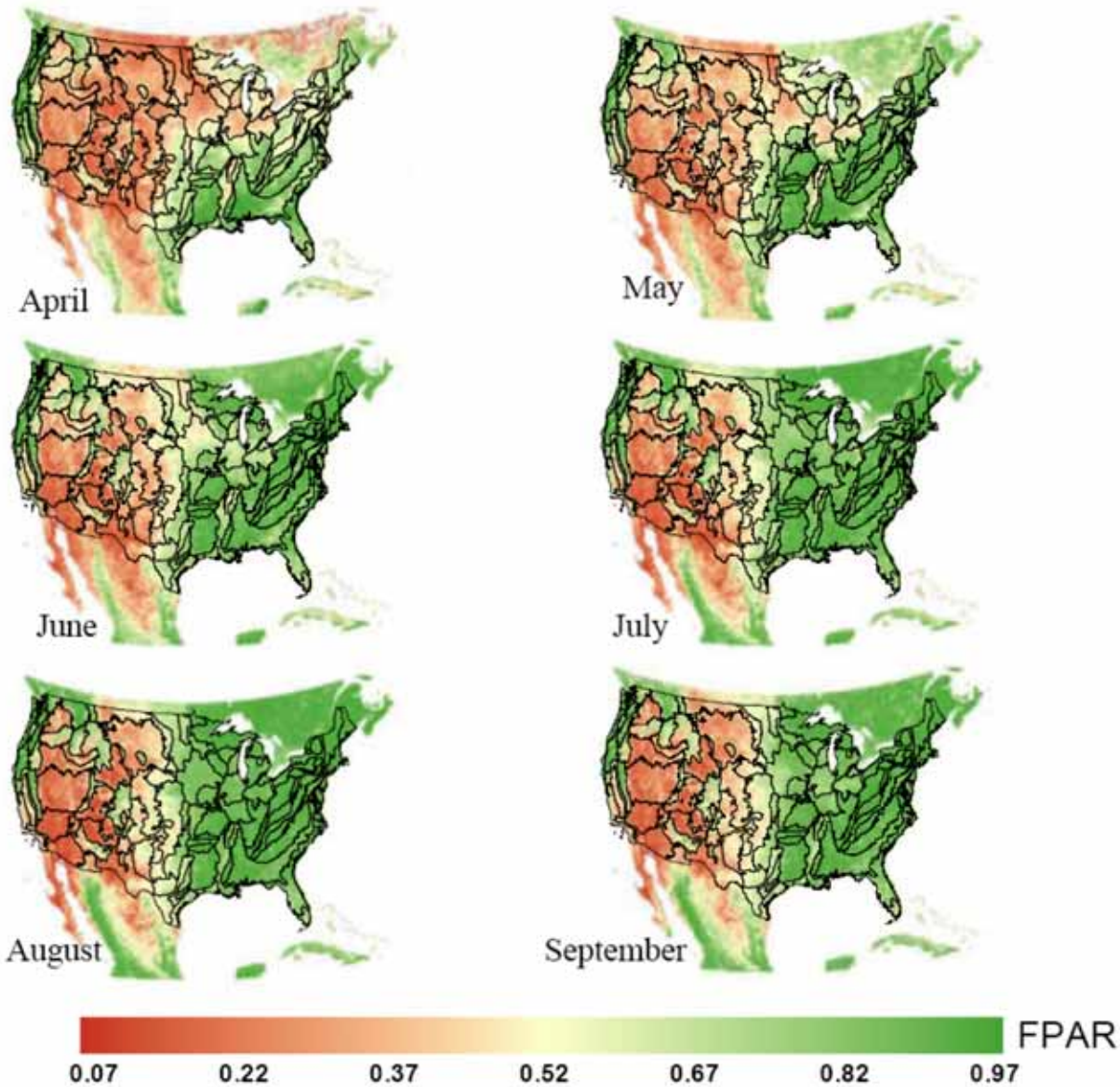
Feature Name	Projection	Resolution (meters)	Geospatial data type	Time Period used (days)
EVI / NDVI	Sinusoidal	500	Grid	16
LAI / FPAR	Sinusoidal	1000	Grid	8
LST / EMISSIVITY	Sinusoidal	1000	Grid	8
SNOW COVER	Sinusoidal	500	Grid	8
ALBEDO	Sinusoidal	1000	Grid	16

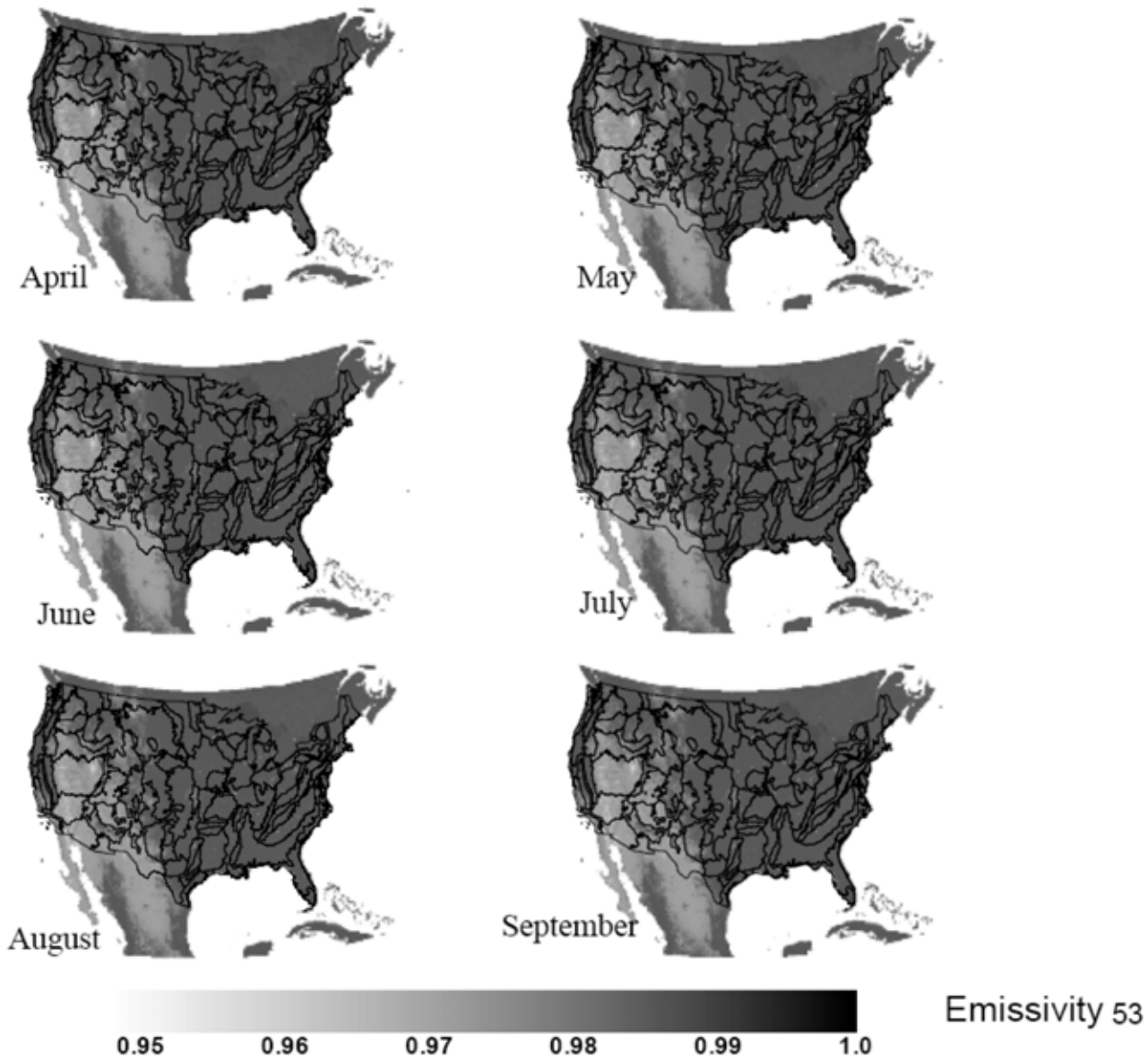
Final Spatial Resolution: 1000m

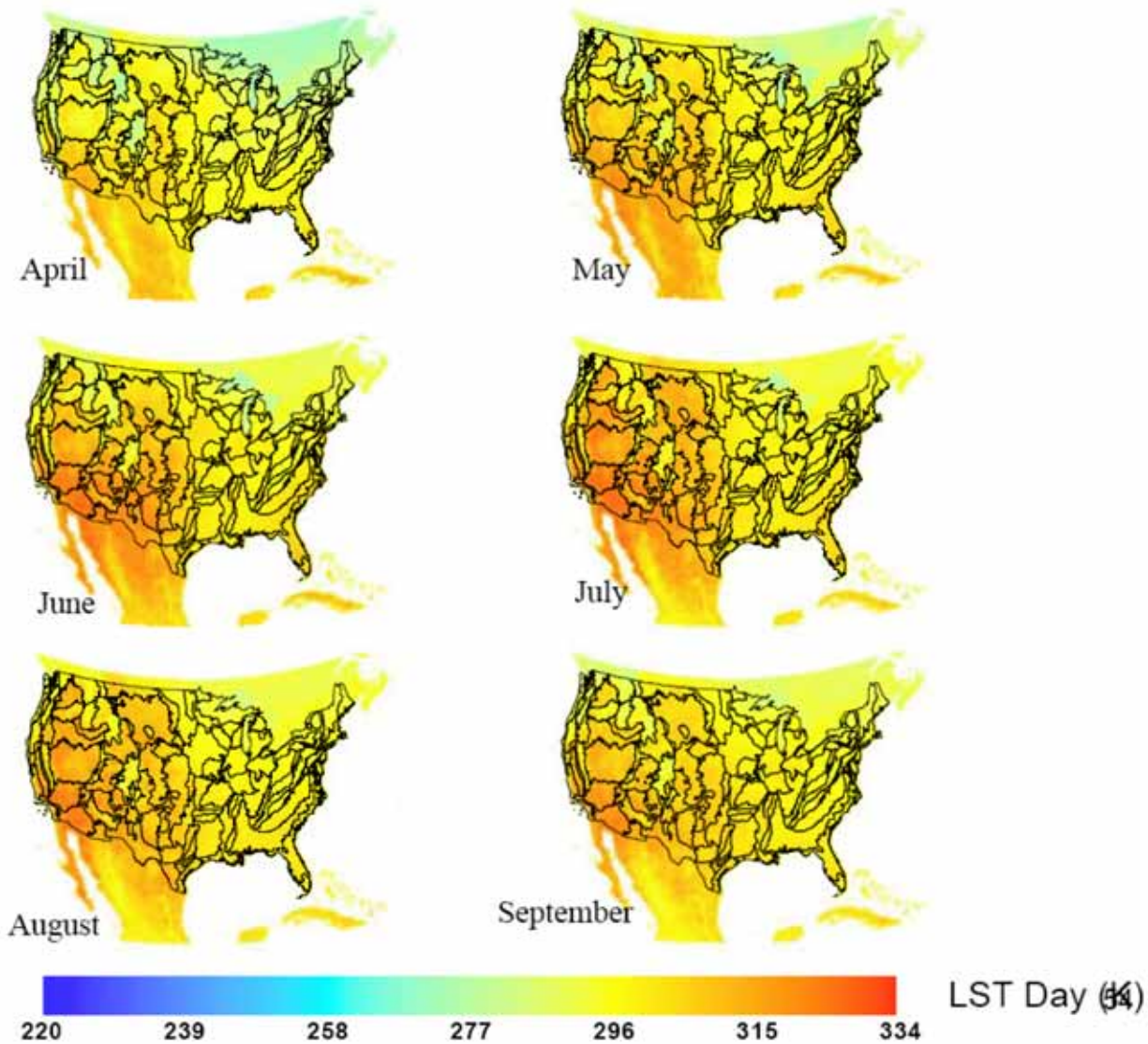

April
May
June
July
August
September
0.1
0.25
0.4
0.55
0.7
0.85
1.0
NDVI

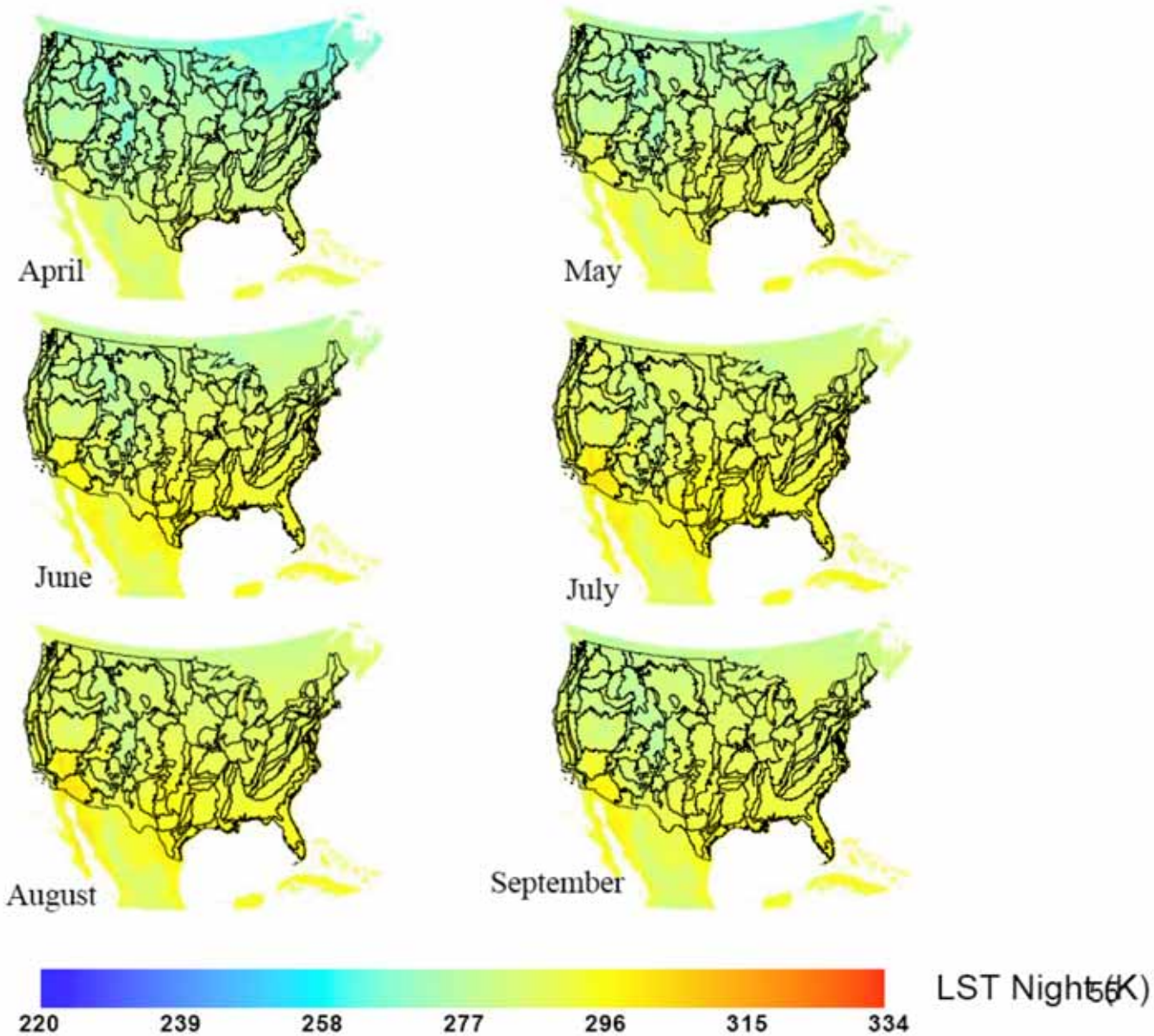

April
May
June
July
August
September
EVI
0.1
0.23
0.36
0.49
0.62
0.75
0.88


April
May
June
July
August
September
LAI
0.1
1.23
2.36
3.49
4.62
5.75
6.88

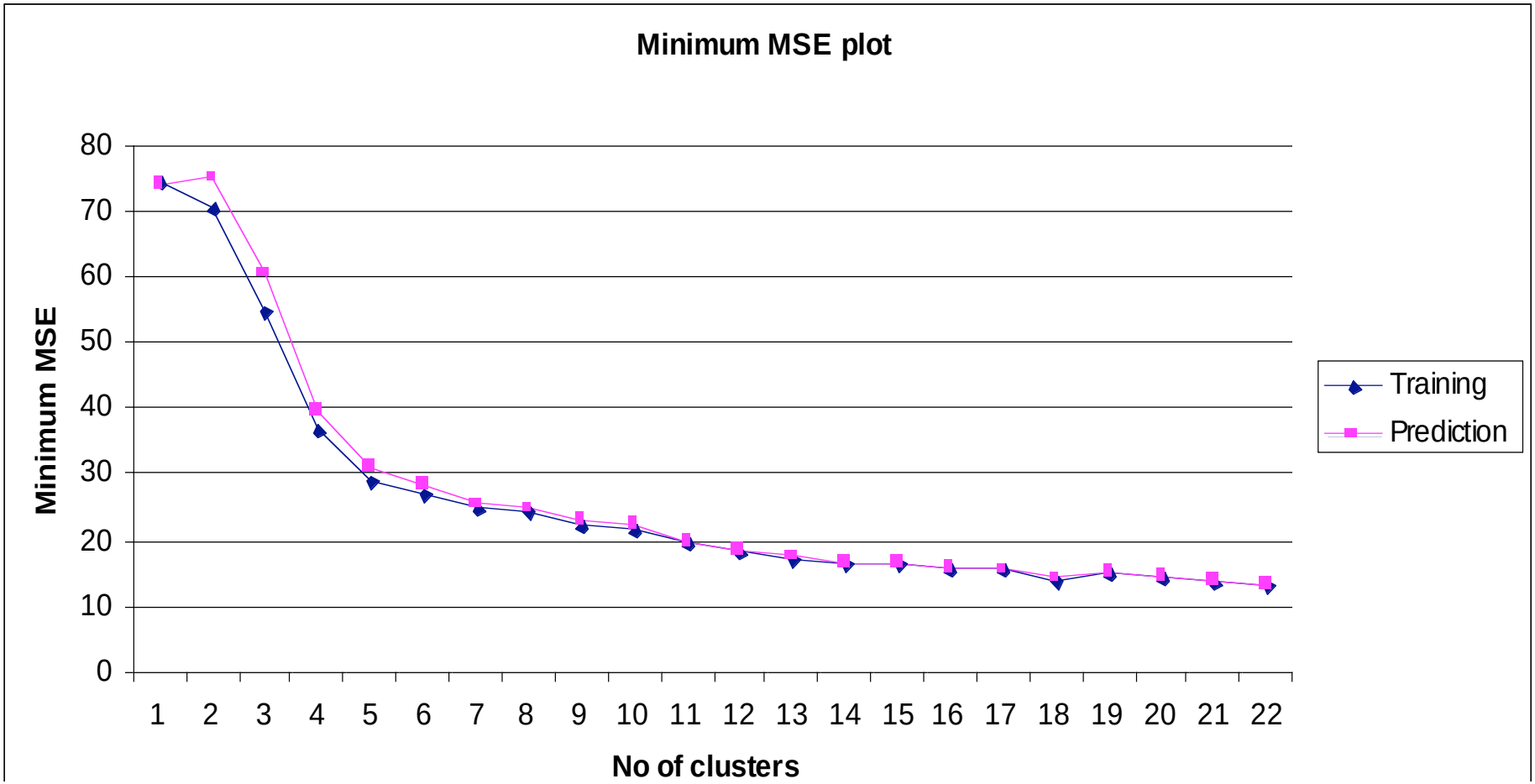

April
May
June
July
August
September
FPAR
0.07
0.22
0.37
0.52
0.67
0.82
0.97


April
May
June
July
August
September
0.95
0.96
0.97
0.98
0.99
1.0
Emissivity

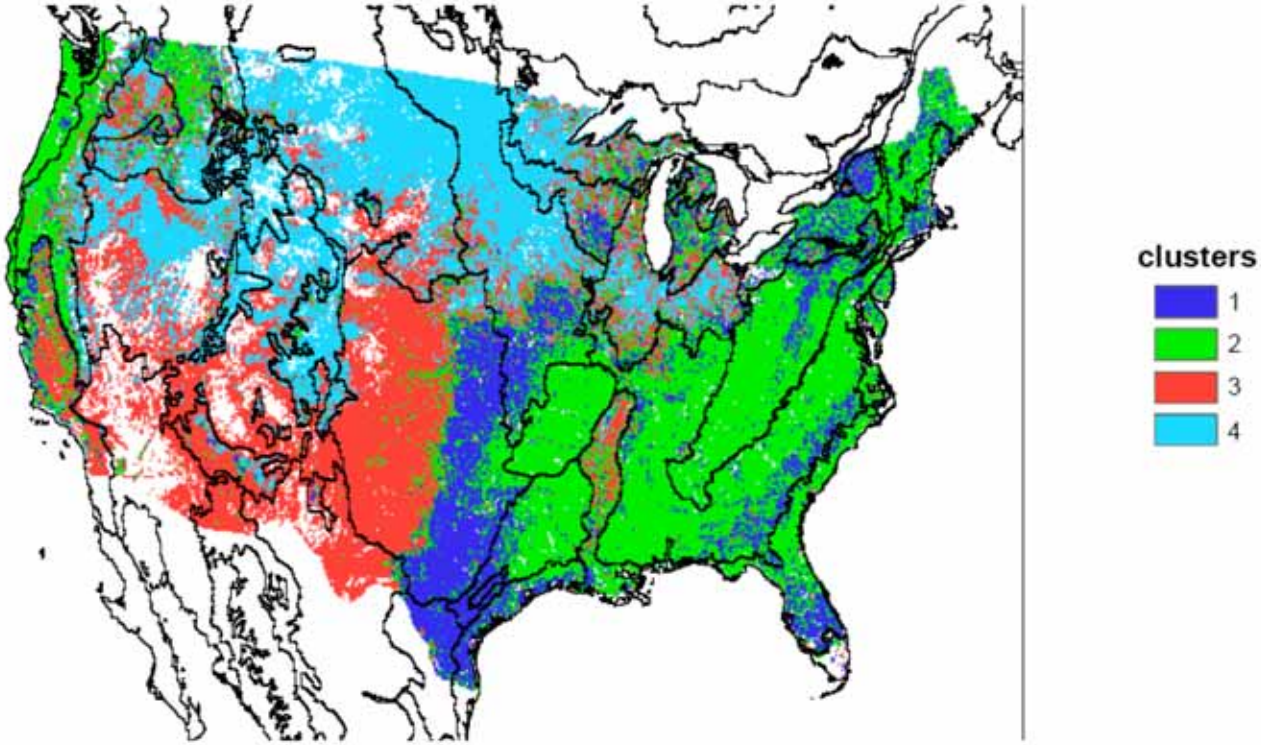

April
May
June
July
August
September
LST Day (K)
220
239
258
277
296
315
334


April
May
June
July
August
September
220
239
258
277
296
315
334
LST Night (K)

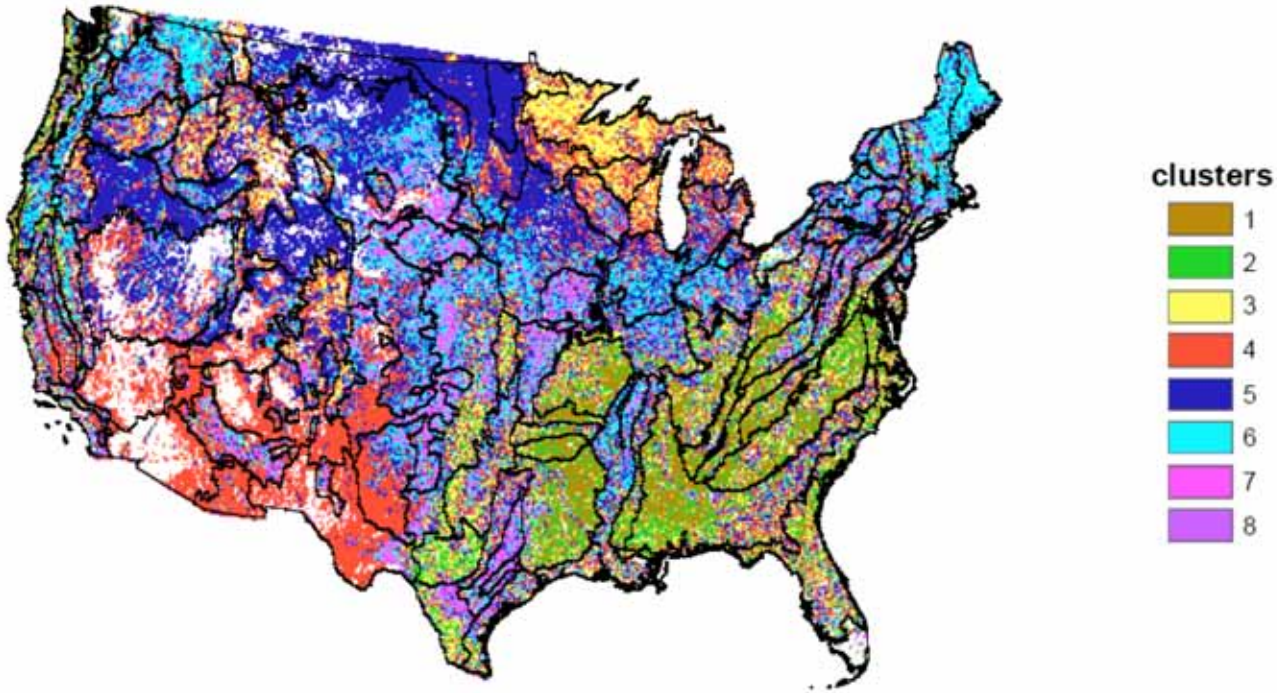
Minimum Mean Square Error Vs. Number of Clusters (US Continent, May 2004)



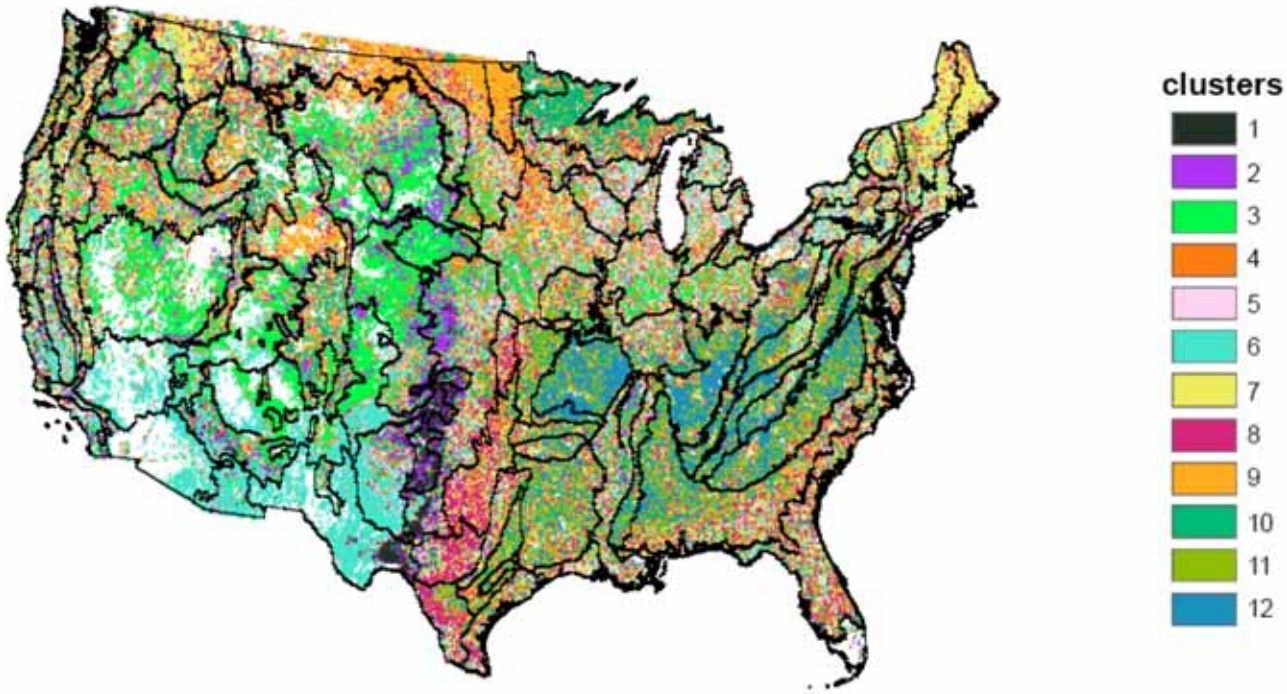
Eco-region level II



Eco-region level III



Eco-region level III



Summary

Current TRL 5, Target TRL 6 by Sep'06

- Architecture for integration remote sensing and GIS data for data mining applications
- Ability to perform detailed (pixel-by-pixel) investigations in space-time domain
  - Dependencies over multitude of variables can be analyzed
- Ability to analyze very large domains with using very fine resolution data
- http://cee.uiuc.edu/research/hydrology/hydroinf_Intro.html