An Early Warning System for Identification and Monitoring of Disturbances to Forest Ecosystems

Aaron A. Marshall[α,γ], Forrest M. Hoffman[α], Jitendra Kumar[α], William W. Hargrove[β], Joseph Spruce[ρ] Richard T. Mills[†]

[α]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States
[γ]Thomas Nelson Community College, Yorktown, VA, United States
[β]Eastern Forest Environmental Threat Assessment Center (EFETAC), USDA Forest Service, Asheville, NC, United States
[ρ]NASA Stennis Space Center, Bay St. Louis, MS, United States

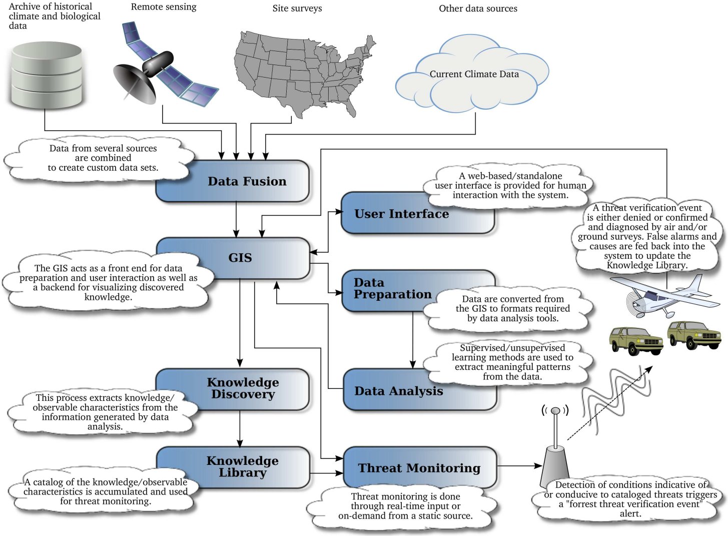
jkumar@climatemodeling.org

Friday December 09, 2011
AGU Fall Meeting 2011: B53D. Remote Sensing of Long-Term Ecological Trends I

Introduction

- Ecoregions are geographical regions of generally similar combination of ecologically relevant conditions like temperature, precipitation and soil characteristics.
- Understanding and delineation of ecoregions are useful for predicting suitable species range, stratification of ecological samples, and to help prioritize habitat preservation and remediation efforts.
- In the case of threatened or endangered species, a well-executed ecoregion classification can be used to identify and locate the extent of suitable habitat for the purposes of preserving or improving it.
- Large amount of data sets are available from satelite, airborne and ground based remote sensing; GCM model outputs
- Data mining tools can be used to extract knowledge from these data sets

Overview of the Forest Incidence Recognition and State Tracking (FIRST) System



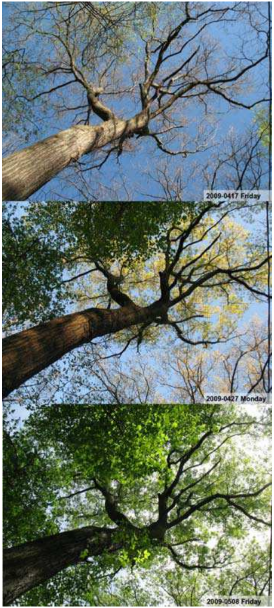
Normalized Difference Vegetation Index (NDVI)

- NDVI exploits the strong differences in plant reflectance between red and near-infrared wavelengths to provide a measure of "greenness" from remote sensing measurements.

$$\mathrm{NDVI} = \frac{(\sigma_{\mathrm{nir}} - \sigma_{\mathrm{red}})}{(\sigma_{\mathrm{nir}} + \sigma_{\mathrm{red}})} \tag{1}$$

- These spectral reflectances are ratios of reflected over incoming radiation, $\sigma = I_r/I_i$, hence they take on values between 0.0 and 1.0. As a result, NDVI varies between -1.0 and $+1.0$.
- Dense vegetation cover is 0.3–0.8, soils are about 0.1–0.2, surface water is near 0.0, and clouds and snow are negative.

- Phenology is the study of periodic plant and animal life cycle events and how these are influenced by seasonal and interannual variations in climate.
- FIRST is interested in deviations from the "normal" seasonal cycle of vegetation growth and senescence.
- NASA Stennis Space Center has developed a new set of National Phenology Datasets based on MODIS.
- Outlier/noise removal and temporal smoothing are performed, followed by curve-fitting and estimation of descriptive curve parameters.



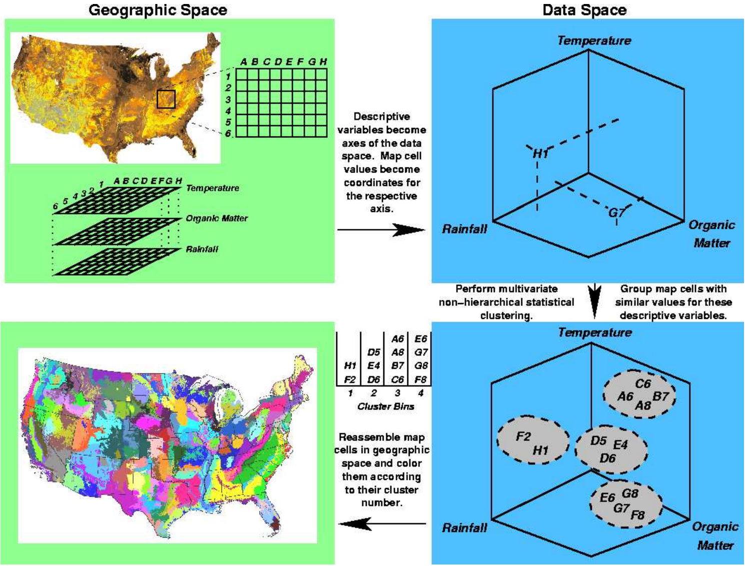
Up-looking photos of a scarlet oak showing the timing of leaf emergence in the spring (Hargrove et al. 2009)

Data Mining for Change Detection

- Changes in forest states are captured by the remote sensing.
- Difficult to use map arithmetic, since the appropriate choice of parameters may vary by region and/or type of forest disturbance.
- An automated, unsupervised change detection system is desired.
- We apply geospatiotemporal data mining techniques to perform unsupervised classification based on multiple years of NDVI history for the entire CONUS.
- Determine the "normal" seasonal and inter-seasonal variations expected at a geographic location.
- Further analysis of clustering outputs for change detection
- Identify unexpected changes in forest phenology states.

Mills, Richard Tran, Forrest M. Hoffman, Jitendra Kumar, and William W. Hargrove. Cluster Analysis-based Approaches for Geospatiotemporal Data Mining of Massive Data Sets for Identification of Forest Threats. Proceedings of the International Conference on Computational Science (ICCS 2011)

Geospatiotemporal Data Mining

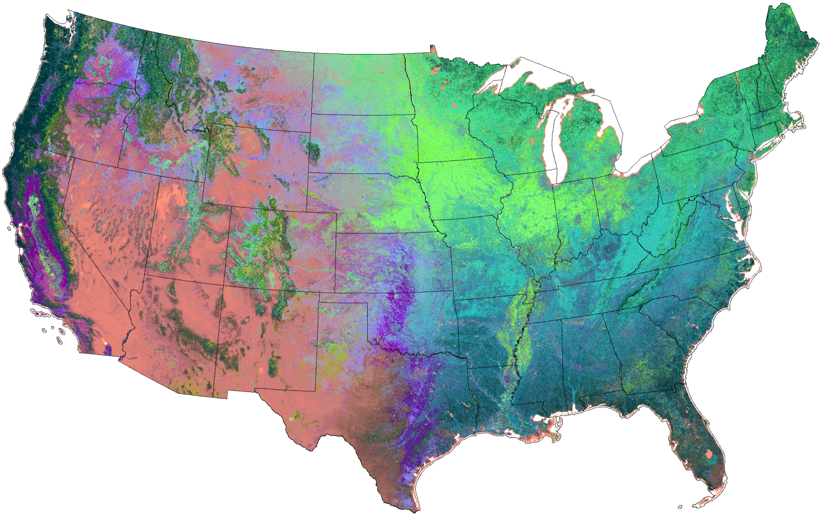


Clustering the MODIS NDVI data

- Data from MODIS: Continental US at 250m resolutions, every 8 days
- The ~67B NDVI values in the data set are arranged as annual NDVI traces of 46 values, for each grid cell (~146.4M cells) in each of the 10 yearly maps (2000-2009),
- The entire set of NDVI traces for all years and map cells is combined into one 251 GB (single precision binary) data set of 46-dimensional "observation" vectors that are analyzed via the *k*-means algorithm.
- After applying *k*-means, cluster assignments are mapped back to the map cell and year from which each observation came, yielding ten maps in which each cell is classified into one of *k* phenoclasses
- The phenoclasses form a "dictionary" of representative or prototype annual NDVI traces (the cluster centroids) derived from the full spatiotemporal extent of the observations in the input data set.

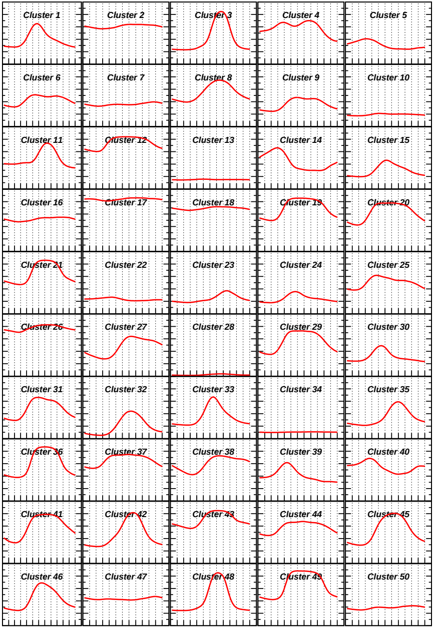
Kumar, Jitendra, Mills, R. T., Hoffman, F. M., Hargrove, W. W. (2011) "Parallel k-means clustering for quantitative ecoregion delineation using large data sets." In International Conference on Computational Science (ICCS 2011), Singapore, June 1-3, 2011.

Cluster based ecoregions: K=50



Similarity colors assigns similar RGB colors to the clusters which are similar in the data space

Representative profiles:K=50



Comparison with available land cover maps

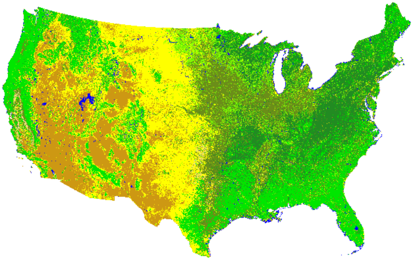
Cluster	IGBP Land cover	Olesen's Global Ecoregions
1	Grasslands	cool_grasses_and_shrubs
2	Evergreen_Needleleaf_Forest	cool_conifer_forest
3	Croplands	corn_and_beans_cropland
4	Cropland/Natural_Vegetation_Mosaic	cool_forest_and_field
5	Open_Shrublands	semi_desert_sage
6	Grasslands	cool_conifer_forest
7	Grasslands	hot_and_mild_grasses_and_shrubs
8	Cropland/Natural_Vegetation_Mosaic	cool_forest_and_field
9	Grasslands	hot_and_mild_grasses_and_shrubs
10	Open_Shrublands	semi_desert_shrubs
11	Croplands	corn_and_beans_cropland
12	Evergreen_Needleleaf_Forest	conifer_forest
13	Open_Shrublands	semi_desert_shrubs
14	Savannas	savanna_(woods)
15	Grasslands	hot_and_mild_grasses_and_shrubs
16	Evergreen_Needleleaf_Forest	cool_conifer_forest
17	Evergreen_Needleleaf_Forest	cool_conifer_forest
18	Evergreen_Needleleaf_Forest	cool_conifer_forest
19	Deciduous_Broadleaf_Forest	deciduous_broadleaf_forest
20	Deciduous_Broadleaf_Forest	deciduous_broadleaf_forest
21	Deciduous_Broadleaf_Forest	cool_broadleaf_forest
22	Open_Shrublands	semi_desert_sage
23	Grasslands	cool_grasses_and_shrubs
24	Grasslands	semi_desert_sage
25	Croplands	woody_savanna

Hargrove, William W., Forrest M. Hoffman, and Paul F. Hessburg. May 12, 2006. Mapcurves: A Quantitative Method for Comparing Categorical Maps. J. Geograph. Syst., 8(2):187208.

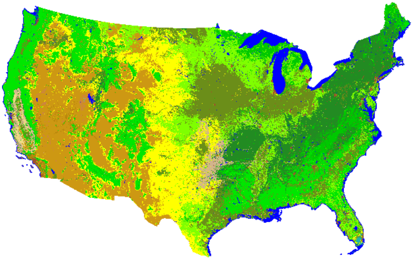
Comparison with available land cover maps

Cluster	IGBP Land cover	Olesen's Global Ecoregions
26	Evergreen_Needleleaf_Forest	conifer_forest
27	Evergreen_Needleleaf_Forest	cool_conifer_forest
28	Water	inland_water
29	Croplands	woody_savanna
30	Grasslands	cool_grasses_and_shrubs
31	Croplands	cool_crops_and_towns
32	Water	inland_water
33	Grasslands	cool_grasses_and_shrubs
34	Open_Shrublands	semi_desert_shrubs
35	Grasslands	hot_and_mild_grasses_and_shrubs
36	Deciduous_Broadleaf_Forest	cool_broadleaf_forest
37	Evergreen_Needleleaf_Forest	deciduous_broadleaf_forest
38	Evergreen_Needleleaf_Forest	cool_conifer_forest
39	Grasslands	hot_and_mild_grasses_and_shrubs
40	Croplands	broadleaf_crops
41	Cropland/Natural_Vegetation_Mosaic	cool_fields_and_woods
42	Croplands	corn_and_beans_cropland
43	Mixed_Forests	cool_broadleaf_forest
44	Croplands	deciduous_broadleaf_forest
45	Cropland/Natural_Vegetation_Mosaic	cool_forest_and_field
46	Cropland/Natural_Vegetation_Mosaic	crops,_grass,_shrubs
47	Evergreen_Needleleaf_Forest	crops,_grass,_shrubs
48	Croplands	corn_and_beans_cropland
49	Deciduous_Broadleaf_Forest	cool_broadleaf_forest
50	Grasslands	cool_grasses_and_shrubs

Cluster maps reclassed based on IGBP land cover types

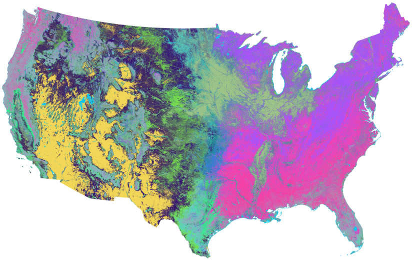


(a) Cluster based map (reclassed)

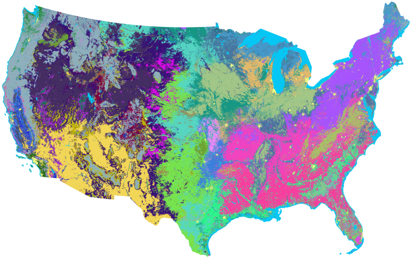


(b) IGBP Land cover map

Cluster maps reclassed based on Olesen's global ecoregions

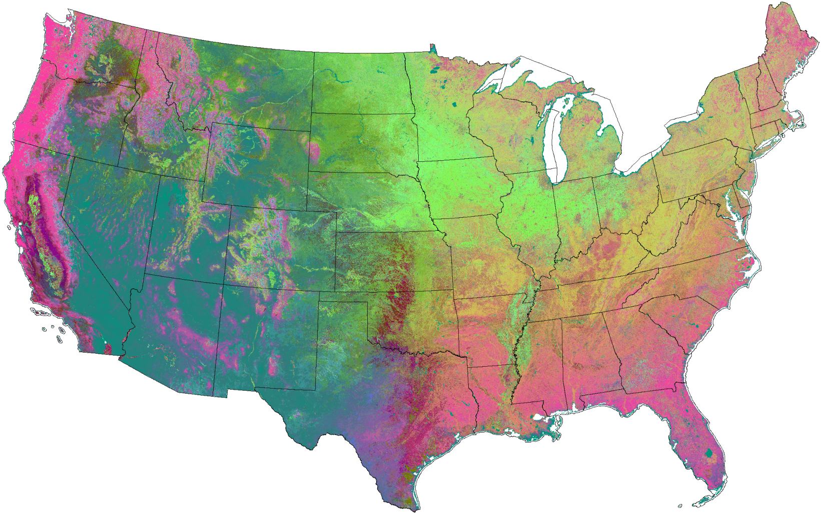


(c) Cluster based map (reclassed)

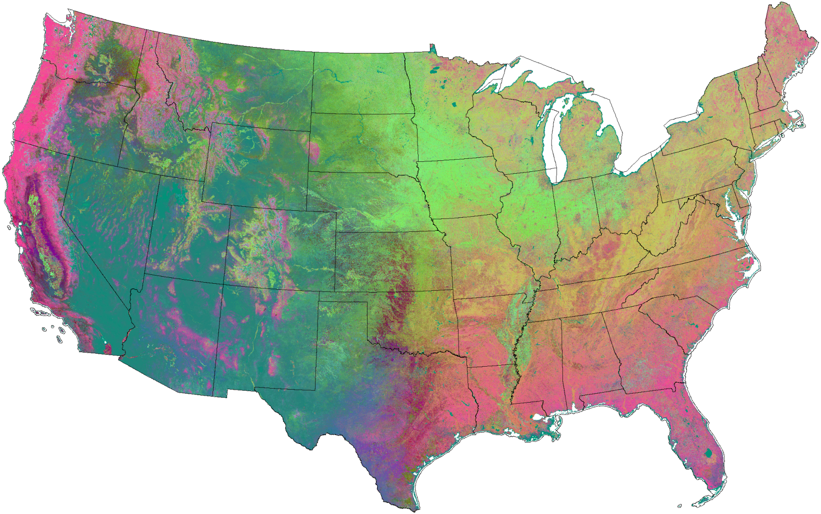


(d) Olesen's global ecoregions

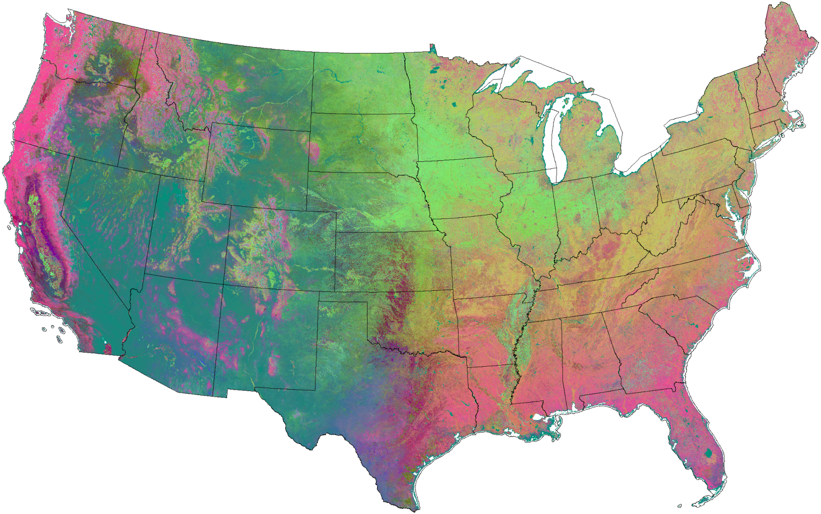
Cluster based ecoregions: K=100



Cluster based ecoregions: K=500



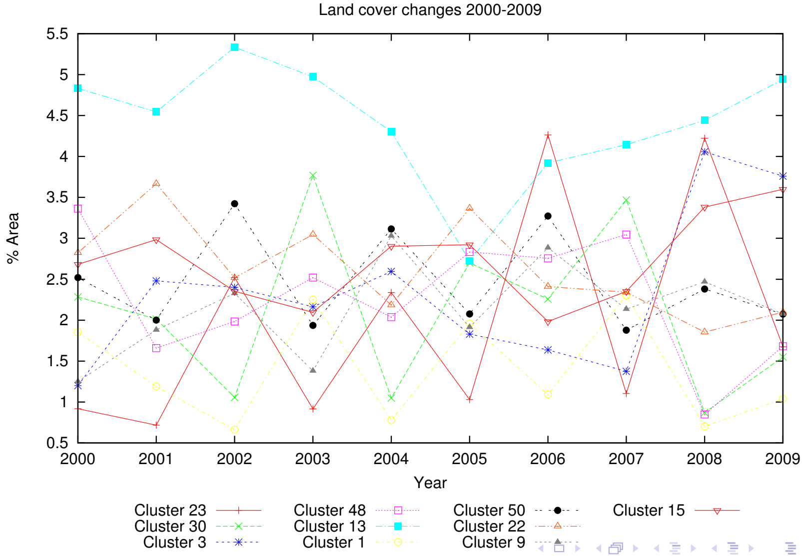
Cluster based ecoregions: K=1000



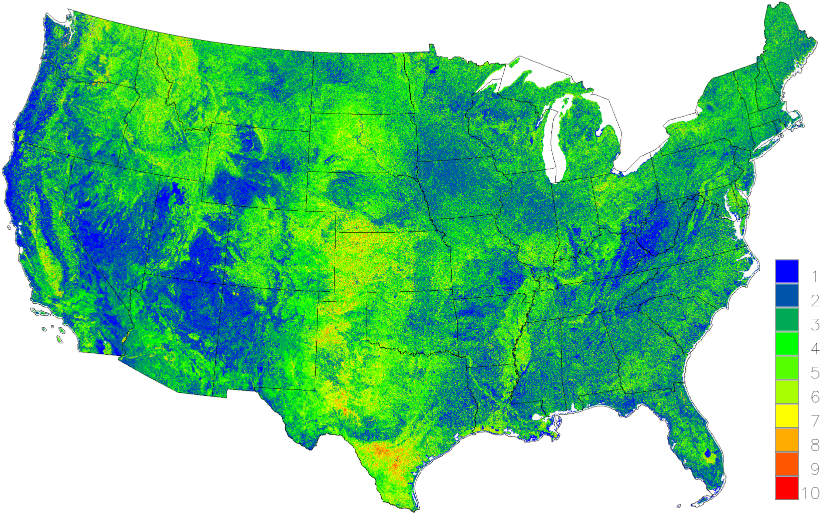
Time Evolution of Cluster Assigments

- Cluster analysis yields 11 maps, one for each year, that classify each cell into one of the k phenoclasses. Here $k = 50$.
- The time evolution of phenoclass assignment, or phenostate, of each cell indicates a trajectory of change in the phenological behavior observed at that location due to natural or anthropogenic disturbance and ecosystem responses to interannual climate variability and long term climate trends.
- Comparison of the current phenostate with the nominal historical phenostate for each cell forms the basis for an early warning system for forest threats.
- Frequency of phenostate occupation for each map cell across all years provides insights into the phenological persistence or variability at every location in the CONUS.

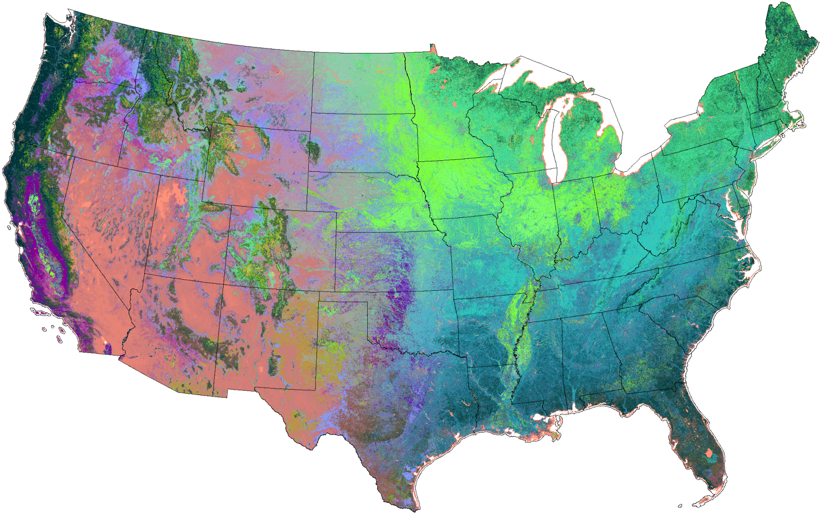
Long term trends in the land cover: K=50



Land cover change/diversity: K=50



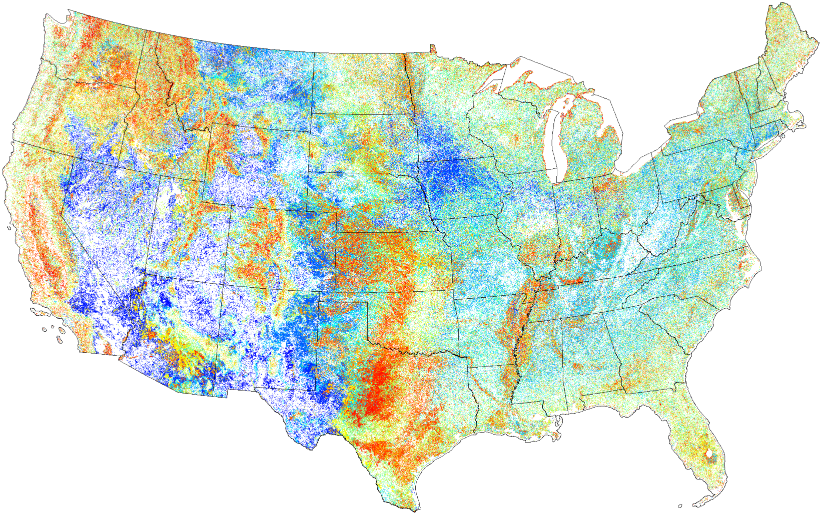
Land cover mode: K=50



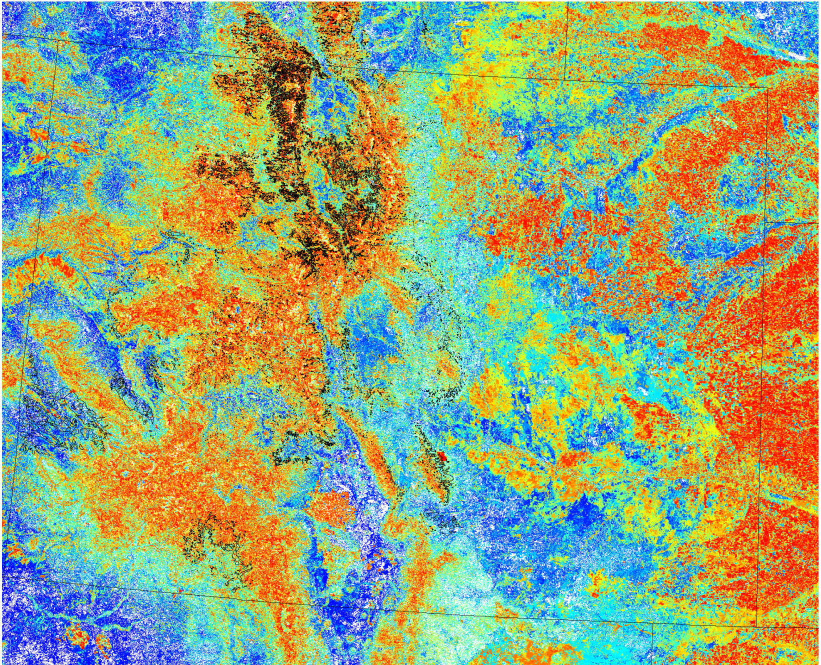
Indicator of change/Disturbance index

- Cluster persistence is strongly dependent on the choice of k.
  - k too large: normal interannual variability results in different phenostate assignment each year.
  - k too small: important phenological change may be missed.
- At a large value of k, difference between the phenostates (transition distance) would provide a relative measure of the strength of the observed change in phenological behavior
  - Euclidean distance between phenostates (in 'n' dimensional space)
  - Integrated area under the NDVI curve
- This provides a relative measure of the strength of the observed change in phenological behavior between any two years.
- Large transition distance would indicate a significant change in the annual phenology over the years

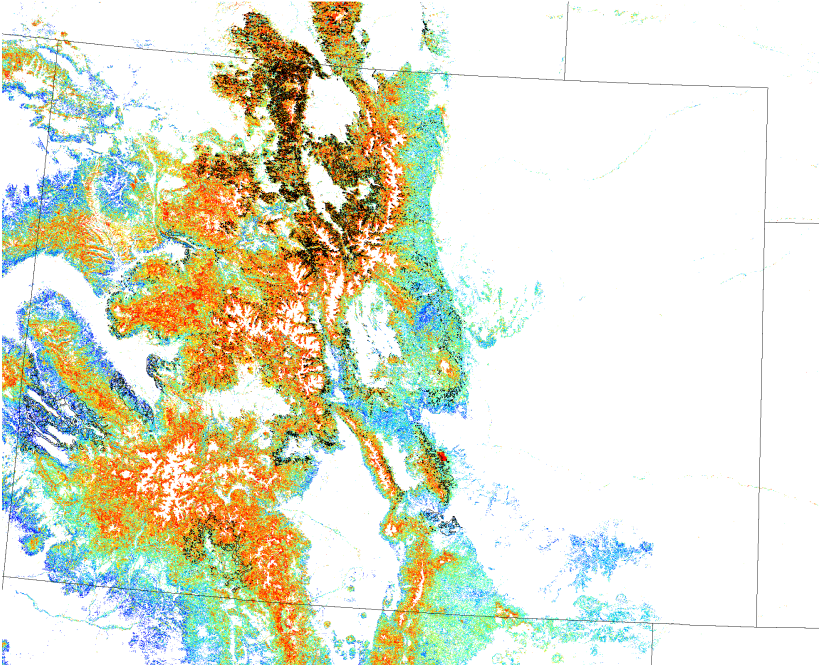
Cluster transition distances 2000-2001: K=50



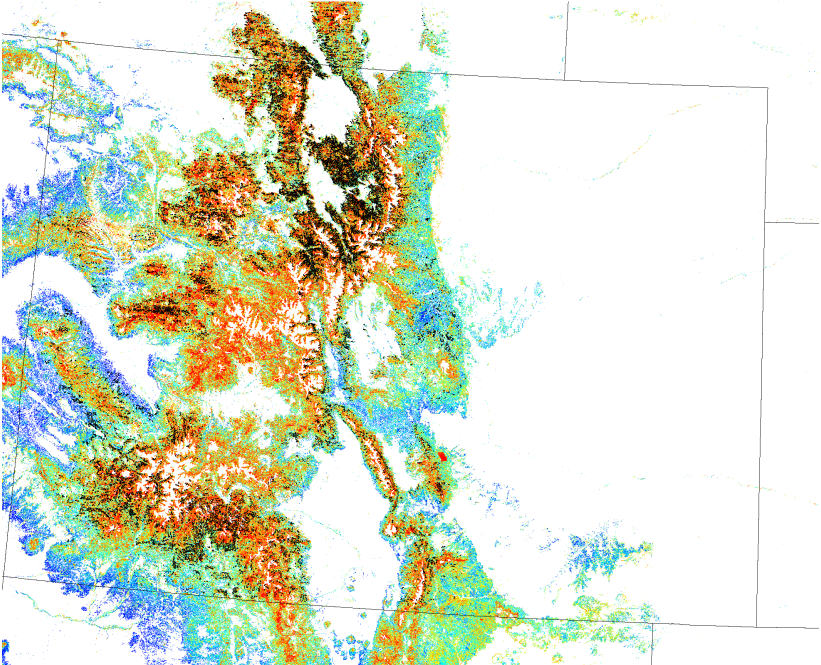
Colorado Mountain Pine Beetle: 2004



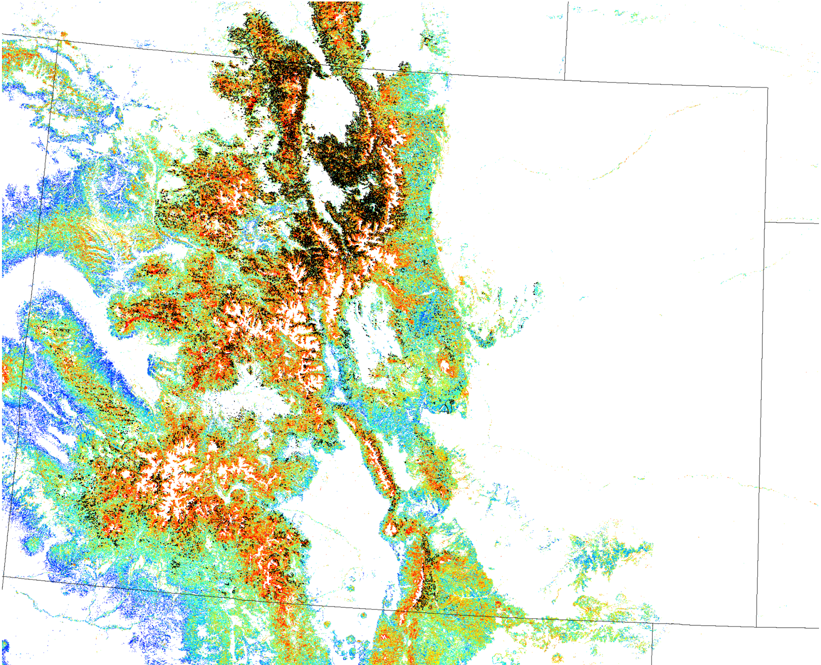
Colorado Mountain Pine Beetle: 2004



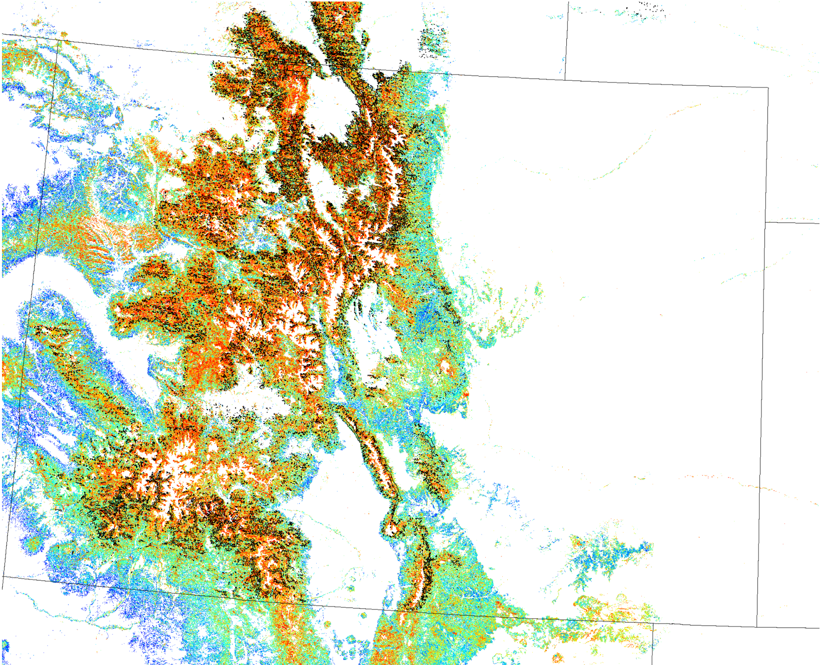
Colorado Mountain Pine Beetle: 2005



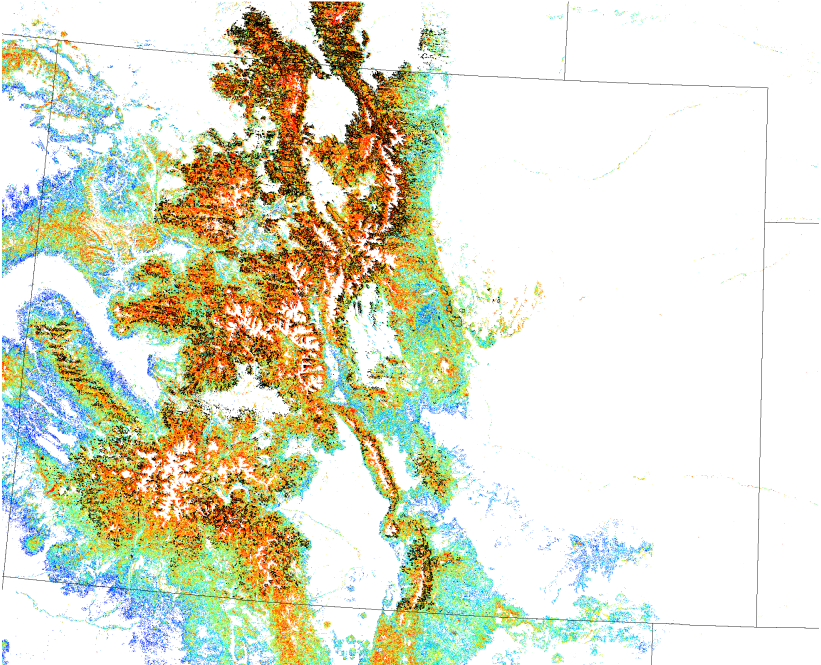
Colorado Mountain Pine Beetle: 2006



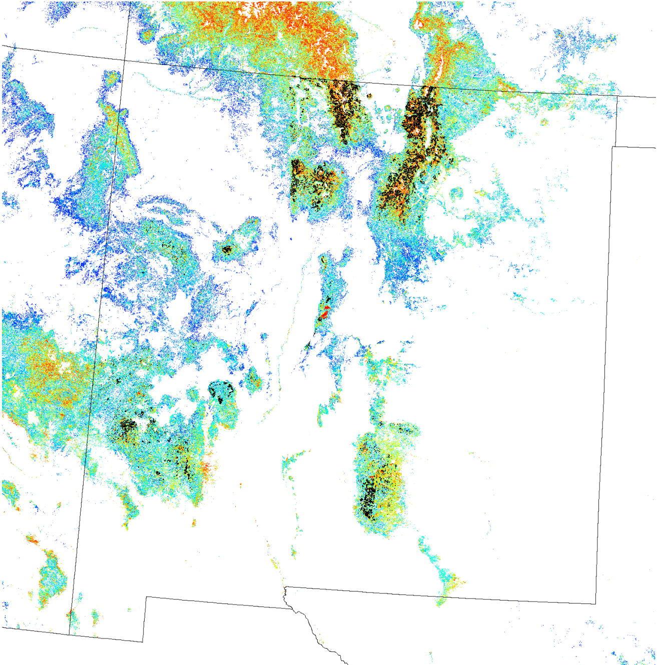
Colorado Mountain Pine Beetle: 2007



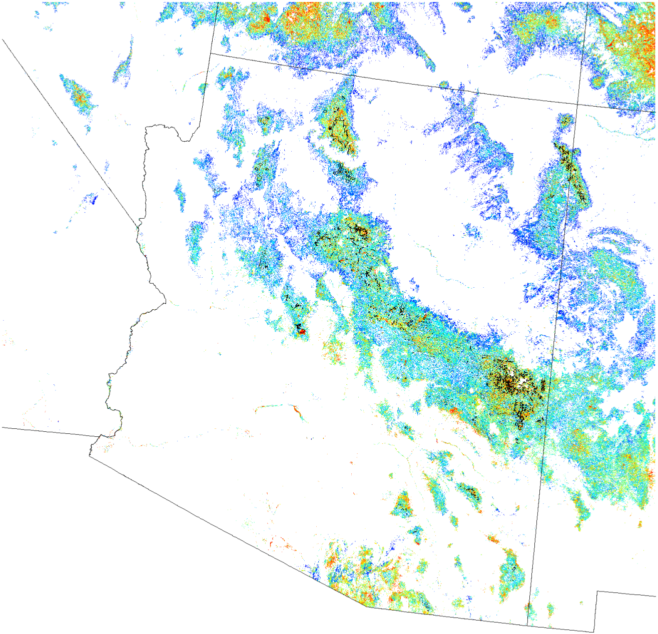
Colorado Mountain Pine Beetle: 2008



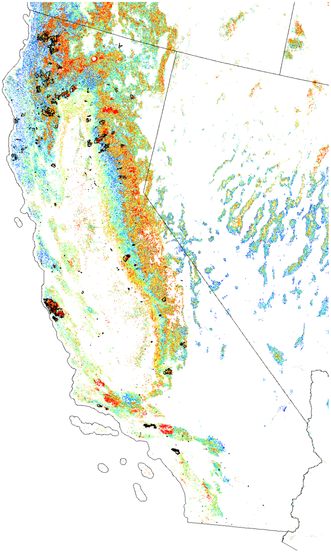
New Mexico Forest Health Decline: 2008



Arizona Forest Health Decline: 2008



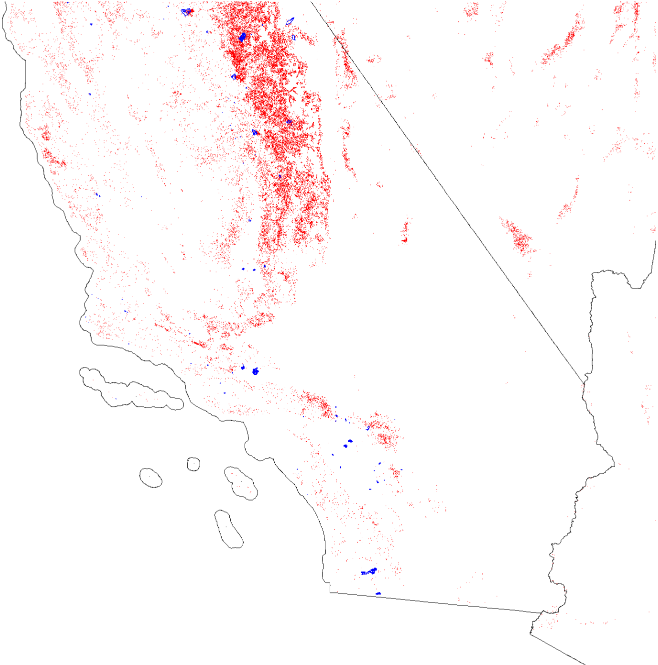
California wildfire: 2008



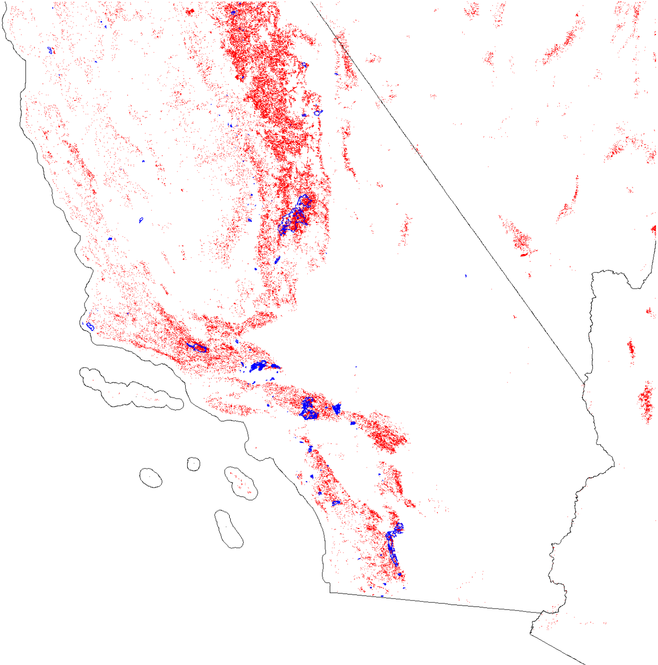
Automated identification of disturbances

- Can we identify disturbances in forest ecosystems using a automated filter??
  - Filter out areas with small changes in phenology: natural variability
  - Identify area of significant changes in phenology: indicative of some disturbance events
- Filters can be developed using observed data for known disturbances disturbances
- Due to vast spatial variability, these automated filters needs to developed regionally
- Indicates drastic change in phenology, not the specific disturbance event that may have caused it

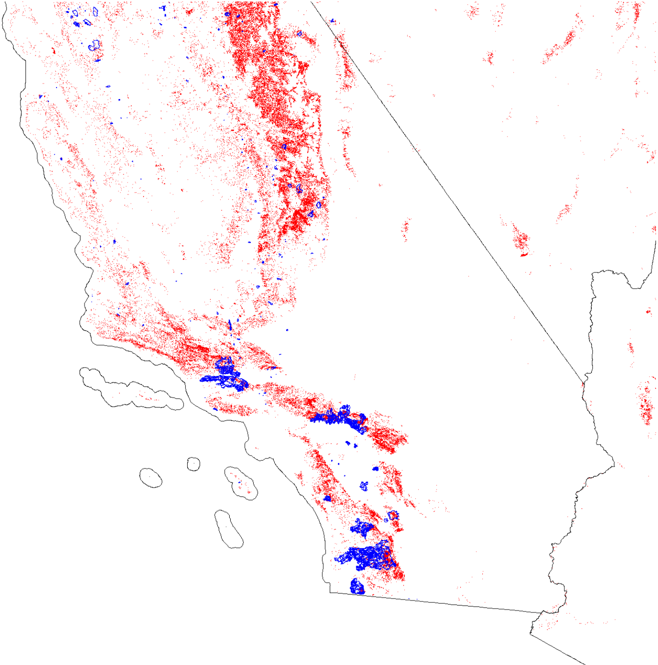
Southern California wildfire: 2001



Southern California wildfire: 2002



Southern California wildfire: 2003



Summary

- Quantitative techniques and tool for analysis of large remote sensing data sets
- Identification of disturbances in the forest ecosystems
- Statistics based objective and consistent method for change detection
- Successfully applied it for a range of know disturbances across CONUS
- Validated using available observed aerial and ground survey datasets

Future/Ongoing directions

- Quantification of identification skill using aerial detection survey datasets
- Establish biome-specific thresholds/filters for transition distances, etc.
- Build a library of declines attributed specific agents for use in complementary, supervised classification