A Statistical Methodology for Detecting and Monitoring Change in Forest Ecosystems Using Remotely Sensed Phenology

Richard T. Mills[α,β], Jitendra Kumar[α], Forrest M. Hoffman[α], William W. Hargrove[γ], and Joseph Spruce[δ]

[α]Computational Earth Science Group, Oak Ridge National Laboratory
[β]Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville
[γ]Eastern Forest Environmental Threat Assessment Center, USDA Forest Service
[δ]NASA Stennis Space Center

rmills@ornl.gov,rtm@utk.edu

Wednesday December 7, 2011
AGU annual meeting
21st San Francisco, CA

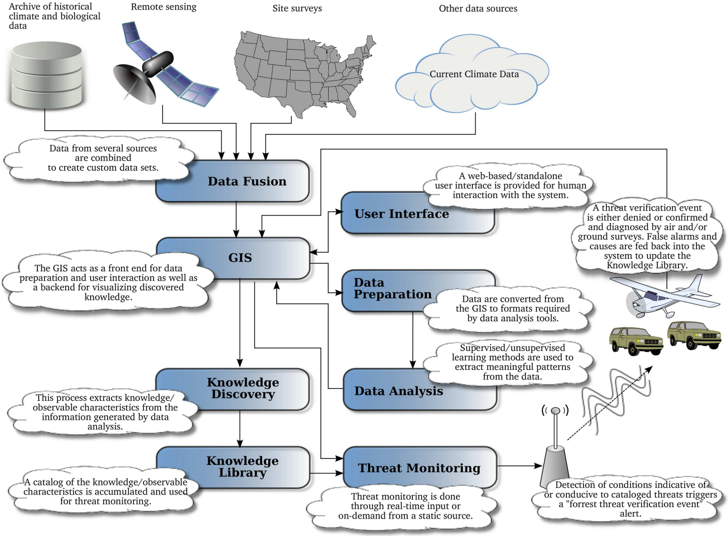
A Forest Early Warning System

The USDA Forest Service, NASA Stennis Space Center, and DOE Oak Ridge National Laboratory are creating a system to monitor threats to U.S. forests and wildlands at two different scales:

- Tier 1: Strategic — An Early Warning System (EWS) that routinely monitors wide areas at coarser resolution, repeated frequently — a *change detection system* to produce alerts or warnings for particular locations may be of interest
- Tier 2: Tactical — Finer resolution airborne overflights and ground inspections of areas of potential interest — *Aerial Detection Survey (ADS)* monitoring to determine if such warnings become alarms

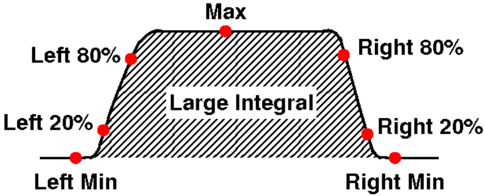
Tier 2 is largely in place, but Tier 1 is needed to optimally direct its labor-intensive efforts and discover new threats sooner.

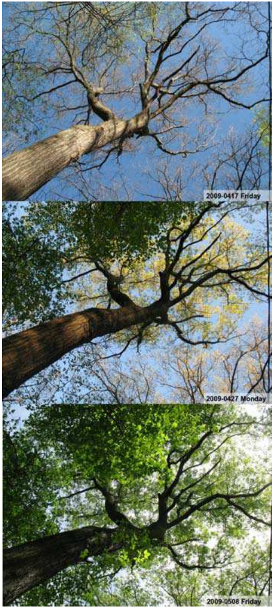
Overview of the Forest Incidence Recognition and State Tracking (FIRST) System



Phenology

- FIRST is interested in deviations from the “normal” seasonal cycle of vegetation growth and senescence.
- We utilize a set of National Phenology Datasets produced by NASA Stennis Space Center based on MODIS NDVI.
- Outlier/noise removal and temporal smoothing are performed, followed by curve-fitting and estimation of descriptive curve parameters.



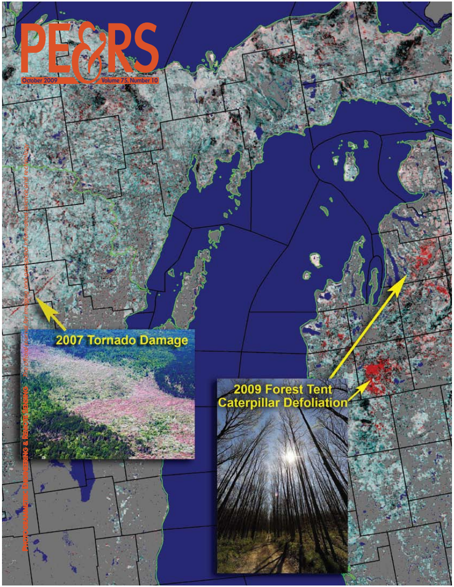


Up-looking photos of a scarlet oak showing the timing of leaf emergence in the spring (Hargrove et al., 2009).

Anomaly detection using map arithmetic

- To detect vegetation disturbances, the current NDVI measurement is compared with the normal, expected baseline for the same location.
- Substantial decreases from the baseline represent potential disturbances.
- Any increases over the baseline may represent vegetation recovery.
- Maximum, mean, or median NDVI may provide a suitable baseline value.

June 10–23, 2009, NDVI is loaded into blue and green; maximum NDVI from 2001–2006 is loaded into red (Hargrove et al., 2009).



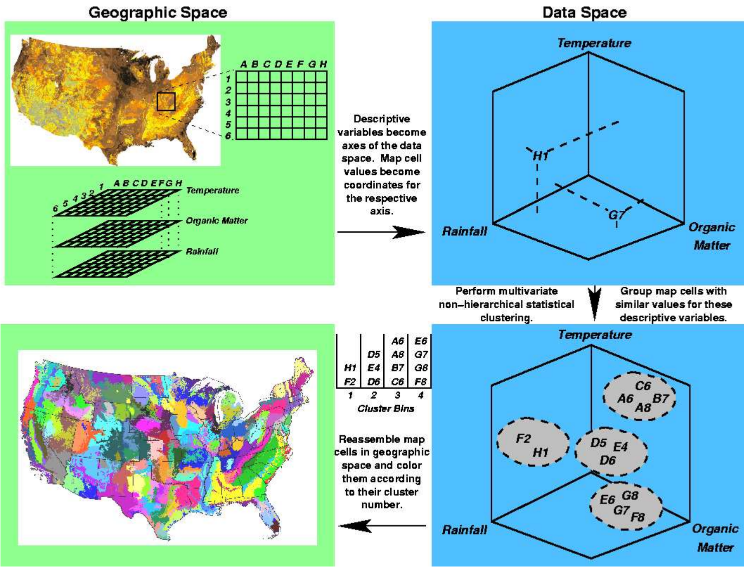
Data Mining for Change Detection

- A difficulty with map arithmetic is identification of appropriate parameters (maximum NDVI, 20% "spring" NDVI, etc.) to use, since the appropriate choice of parameters may vary by region and/or type of forest disturbance.
- To complement such approaches, we desire an automated, unsupervised change detection system.
- Using high-performance computing, we apply geospatiotemporal data mining techniques to perform unsupervised classification based on multiple years (2000–2010) of NDVI history for the entire CONUS.
- These classifications use the full volume of available NDVI data (297GB here) to construct a potential basis for determining the "normal" seasonal and inter-seasonal variation expected at a geographic location.

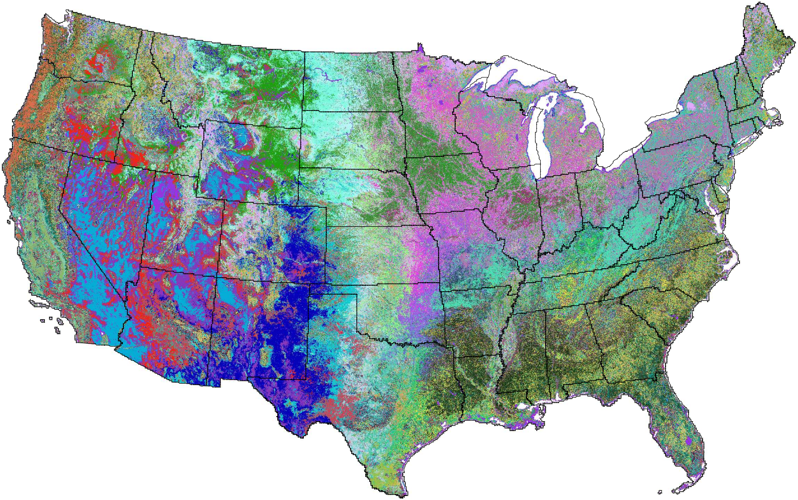
Clustering the MODIS NDVI data

- All 71B NDVI values in the data set are arranged as annual NDVI traces of 46 values, for each grid cell (146.4M records) in each of the 11 yearly maps.
- The entire set of NDVI traces for all years and map cells is combined into one 297 GB (single precision) data set of 1606M 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 11 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.

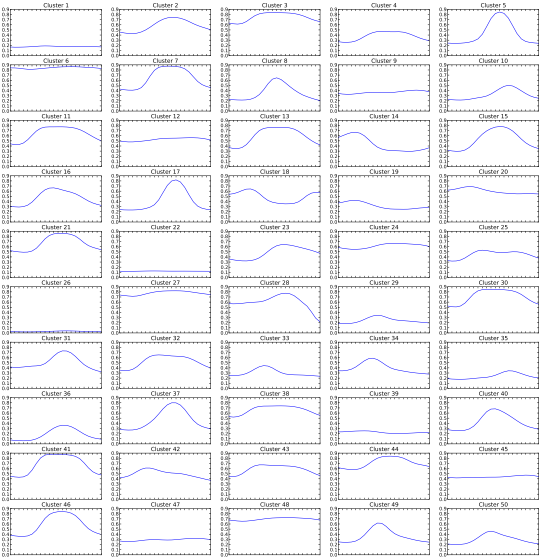
Geospatiotemporal Clustering



50 Phenoregions for Year 2008 (Clustering 2003–2008)



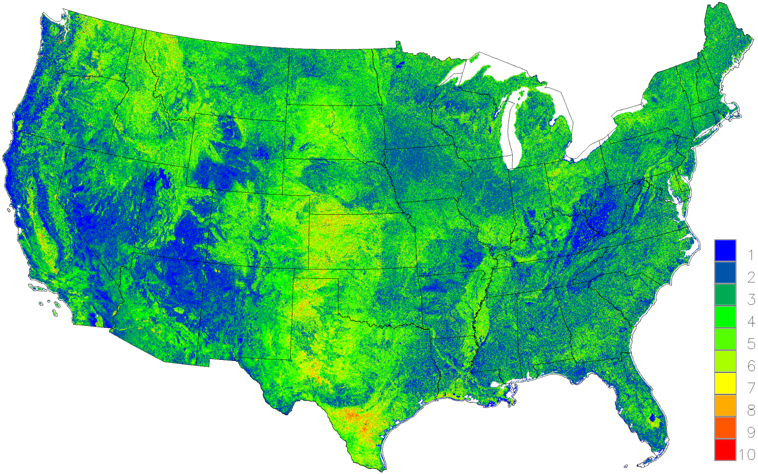
50 Phenoregion Prototypes



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.

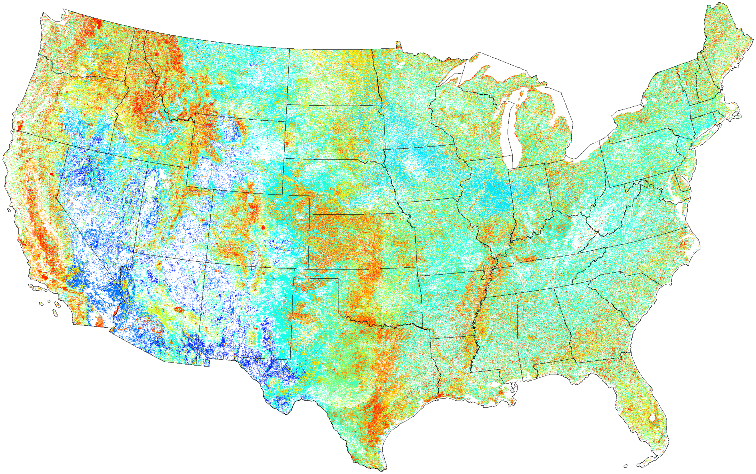
Cluster Persistence Map (2000–2009)



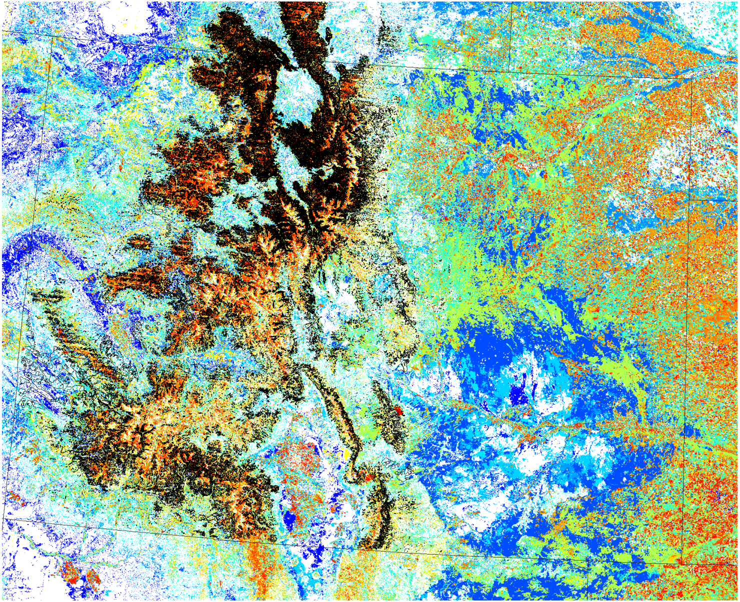
Euclidean Transition Distance as an Indicator of Change

- 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.
- One alternative: use a larger value of k and create maps of Euclidean "transition" distance between phenostate assignments in data space.
- This provides a relative measure of the strength of the observed change in phenological behavior between any two years.
- A large transition distance at any location indicates a significant change in the annual phenological cycle between the initial and final year.

Cluster Transition Distances, 2000–2009



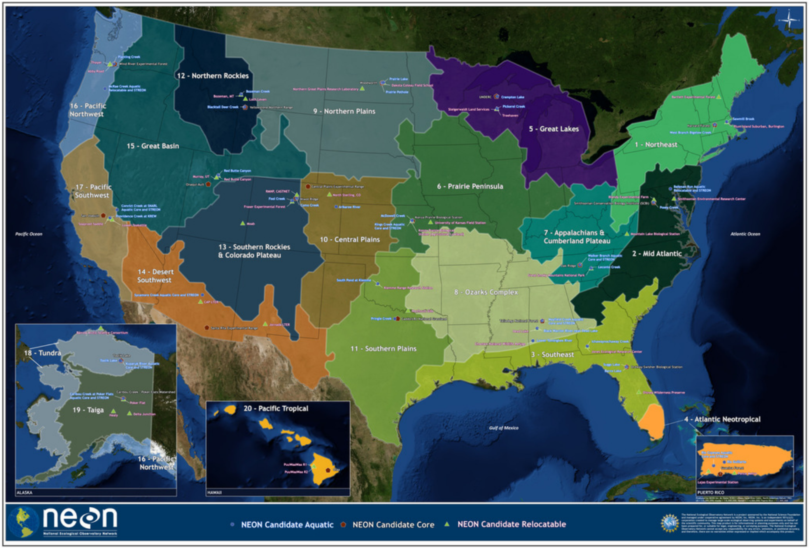
Mountain Pine Beetle in Colorado for (2008 – 2003)



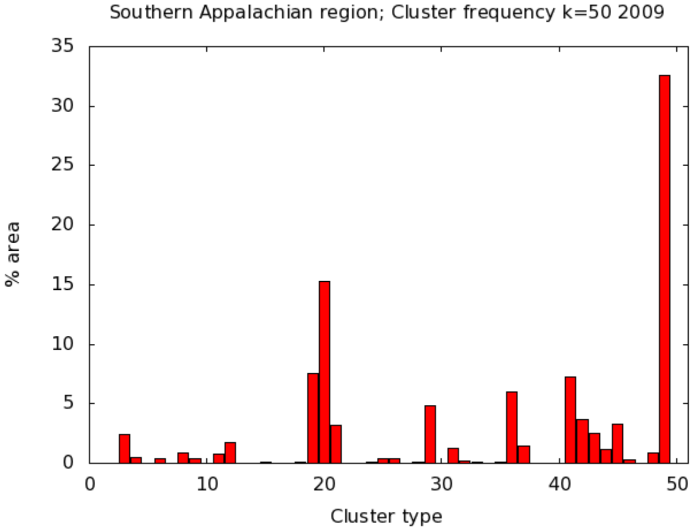
Detecting intra-year anomalies

- Thus far, we have focused on time evolution of phenostates between years.
- We also look for anomalies within single years.
- Looking within regions of homogeneous ecoclimatological characteristics (NEON domains):
  - Identify clusters that appear infrequently.
  - Use principal components classification to identify observations that do not fit the correlation structure of the data.

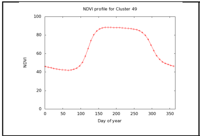
NEON Domains



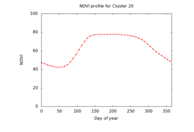
Cluster frequencies: k=50, 2009



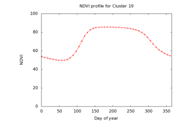
NDVI profiles of most abundant phenoclasses (k=50, 2009)



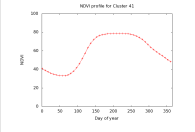
32.61%



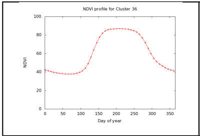
15.23%



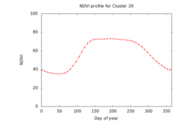
7.57%



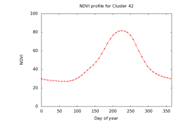
7.20%



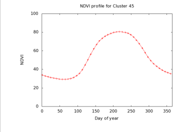
6.03%



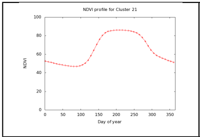
4.81%



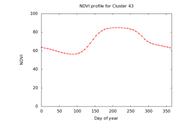
3.69%



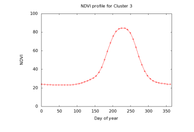
3.32%



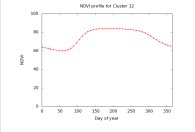
3.23%



2.48%

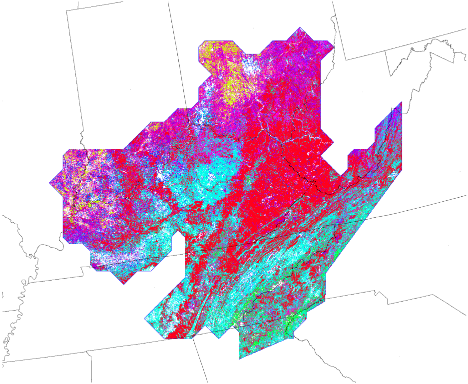


2.37%



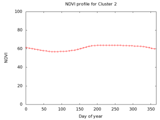
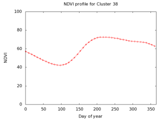
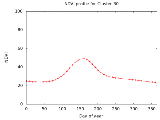
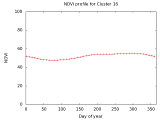
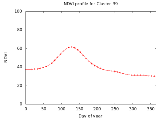
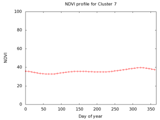
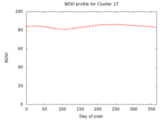
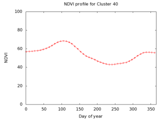
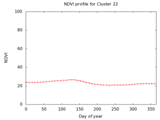
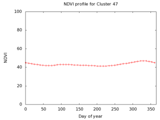
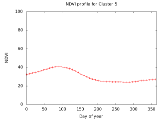
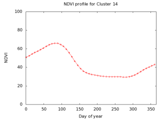
1.69%

Distribution of most abundant phenoclasses: 90% area (k=50, 2009)

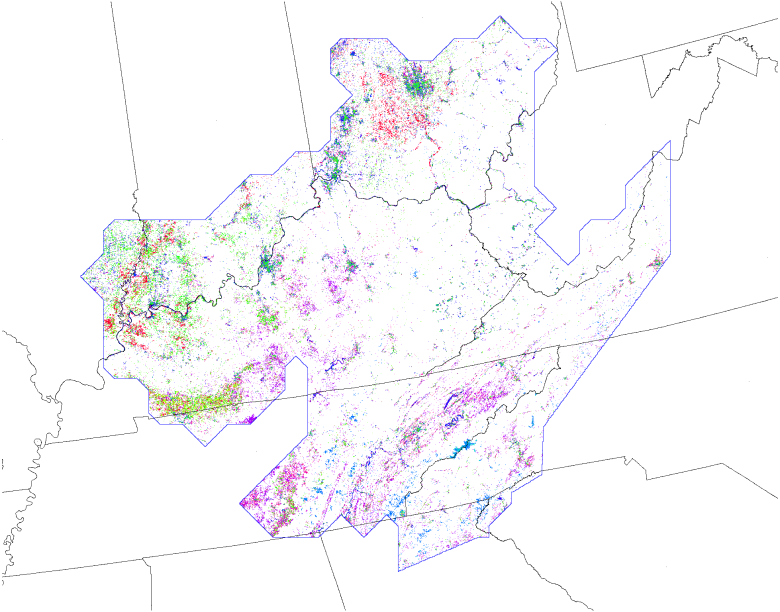


Cluster 49	32.61 %	Deciduous Broadleaf Forest
Cluster 20	15.23%	Deciduous Broadleaf Forest
Cluster 19	7.57%	Deciduous Broadleaf Forest
Cluster 41	7.20%	Cropland/Natural veg
Cluster 36	6.03%	Deciduous Broadleaf Forest
Cluster 29	4.81%	Croplands
Cluster 42	3.69%	Croplands
Cluster 45	3.32%	Croplands/Natural veg
Cluster 21	3.23%	Deciduous Broadleaf Forest
Cluster 43	2.48%	Mixed forests
Cluster 3	2.37%	Croplands
Cluster 12	1.69%	Evergreen needleleaf Forest

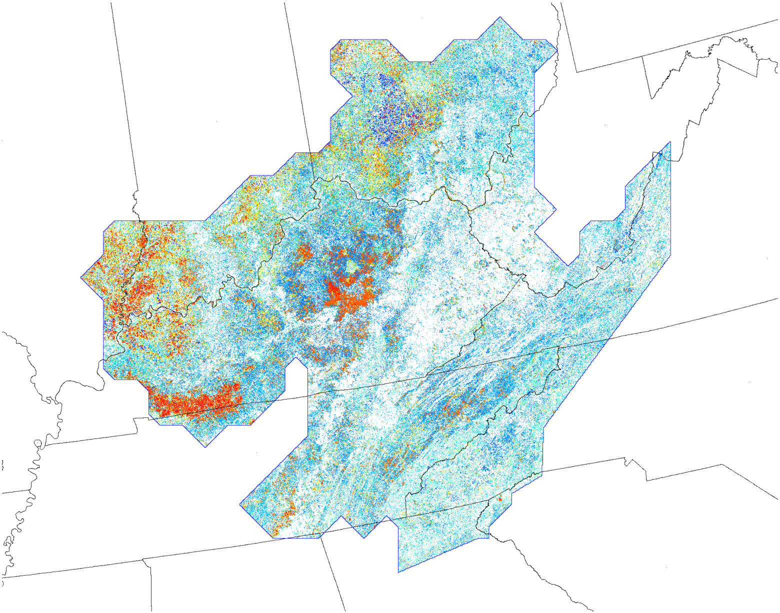
NDVI profiles of least abundant phenoclasses (k=50, 2009)

			
0.0155%	0.0113%	0.0087%	0.0089%
			
0.0071%	0.0042%	0.0027%	0.0019%
			
0.0016%	0.0014%	0.0007%	0.00014%

Distribution of least abundant phenoclasses 10% area (k=50, 2009)

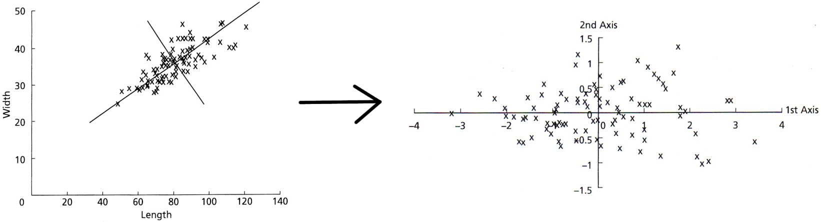


Transition Distance Map: Year 2008-2009, K=50



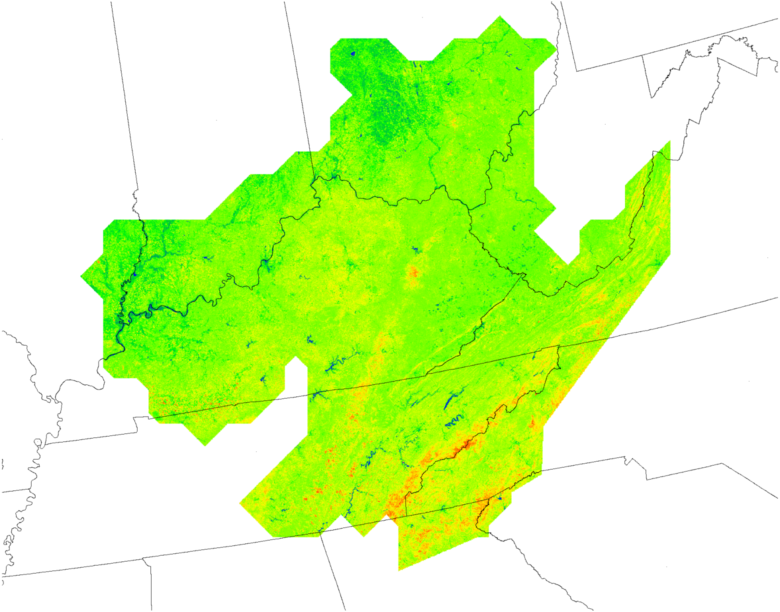
A complementary approach: Principal component analysis

Principal Components Analysis (PCA) determines, for a p-dimensional data set, an orthogonal set of p new axes (linear combinations of the original p variables) such that the first axis explains the greatest variance, the second explains the next most variance, and so on.

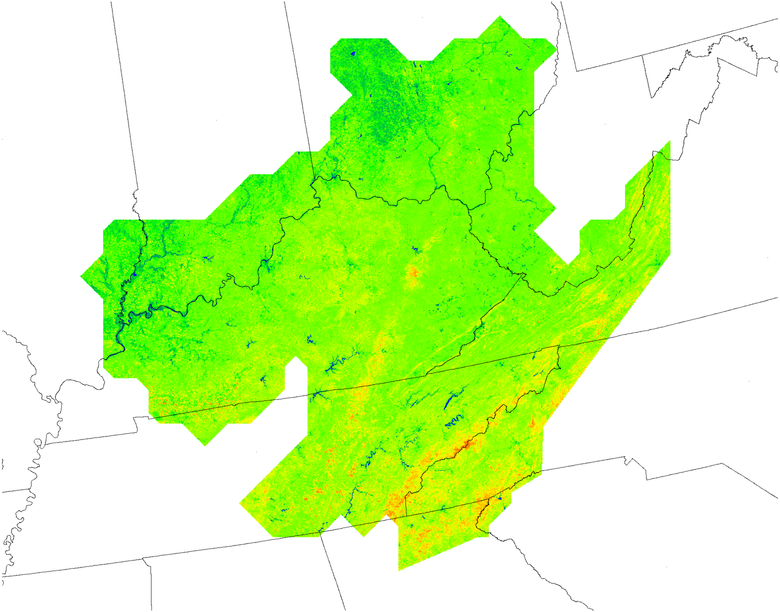


- Computed by finding eigenpairs of the covariance matrix
- Commonly used to determine dominant patterns in data
- But can also be used to determine the anomalous patterns: Observations that score strongly on low order components do not follow the correlation structure of the data.

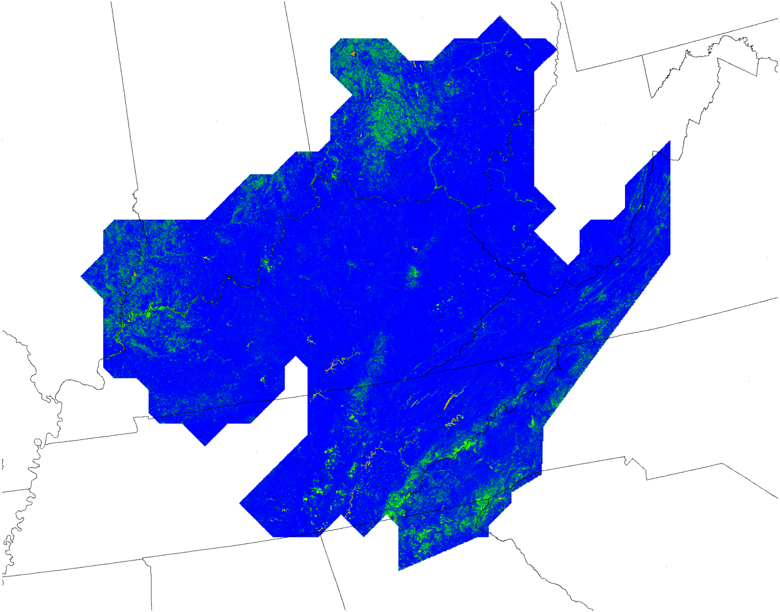
Scores Along Principal Component 1: Year 2008



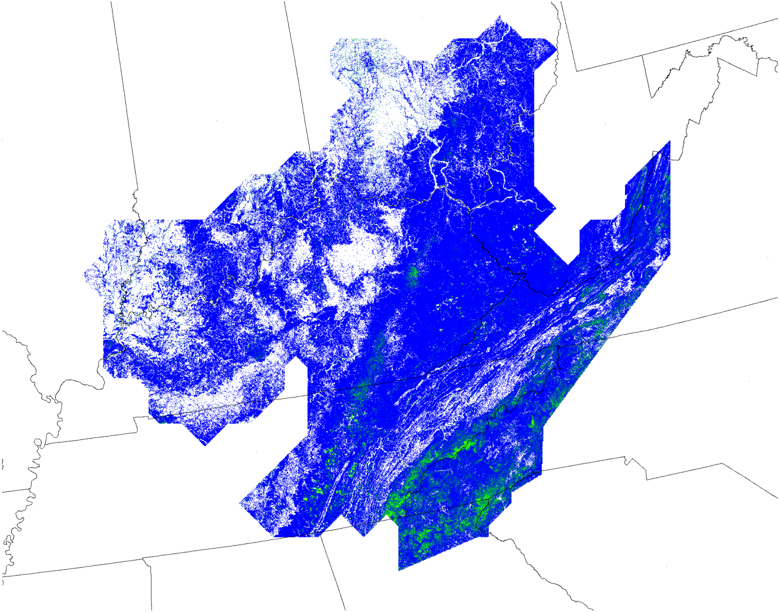
Scores Along Principal Component 2: Year 2008



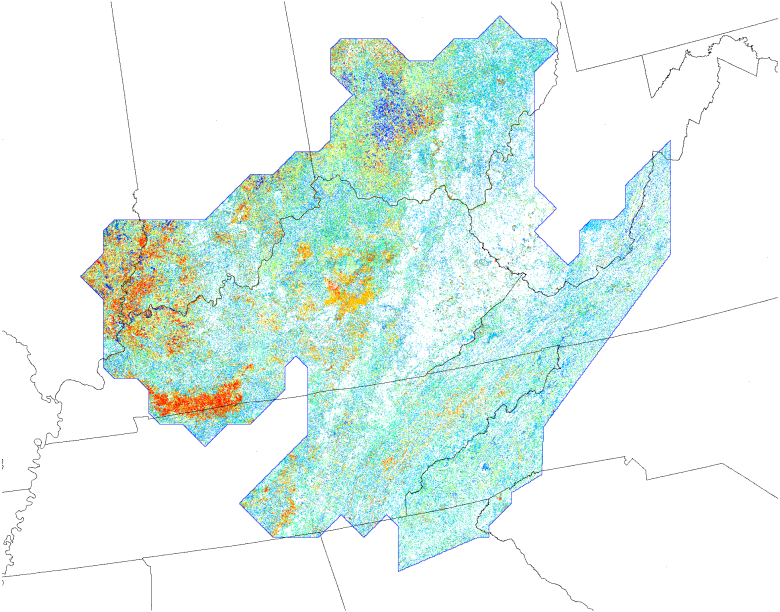
Squared Scores, Principal Component 22: Year 2008



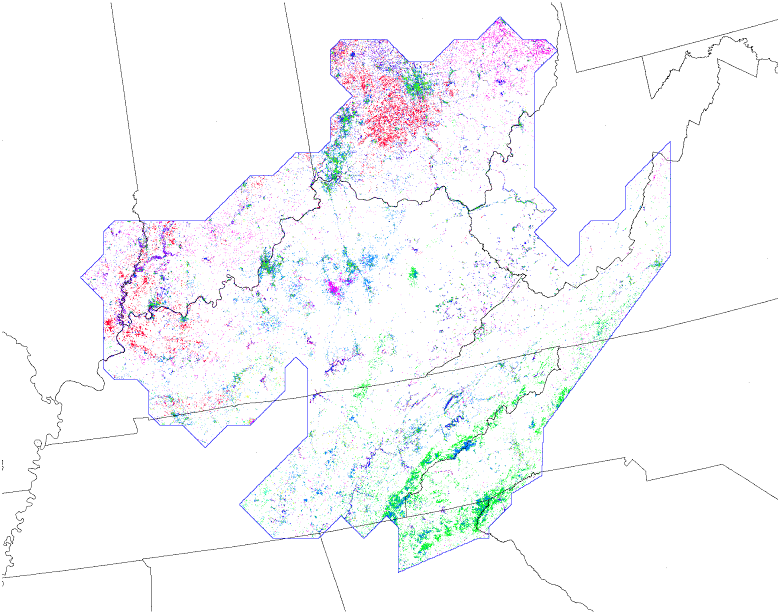
Squared Scores, PC 22: Year 2008 (forests only)



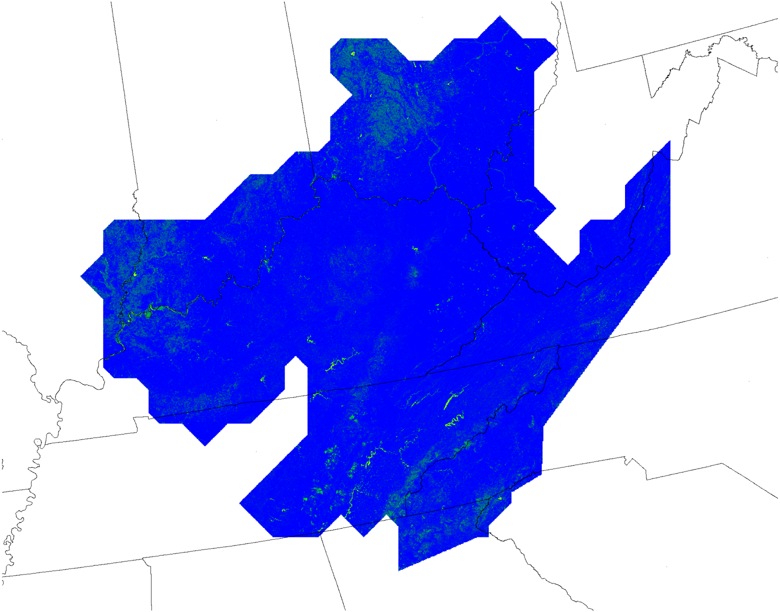
Transition Distance Map: Year 2007-2008, K=50



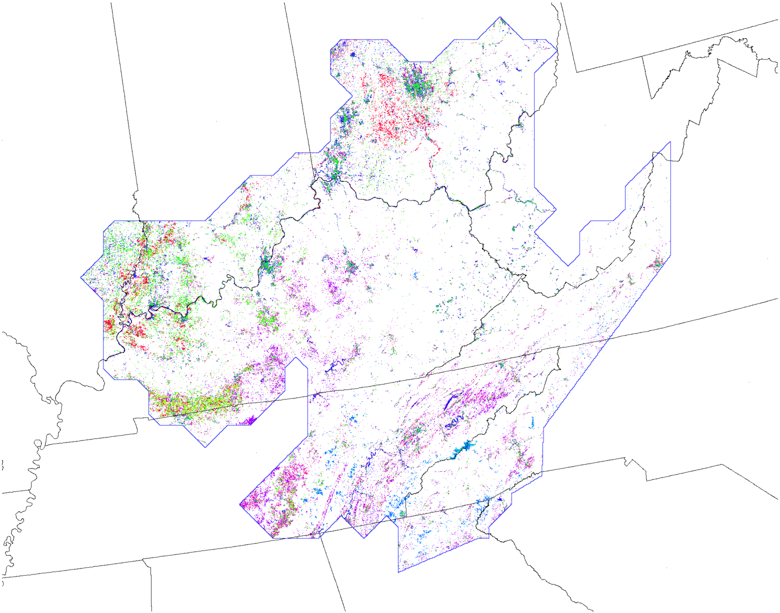
Distribution of least abundant phenoclasses 10% area (k=50, 2008)



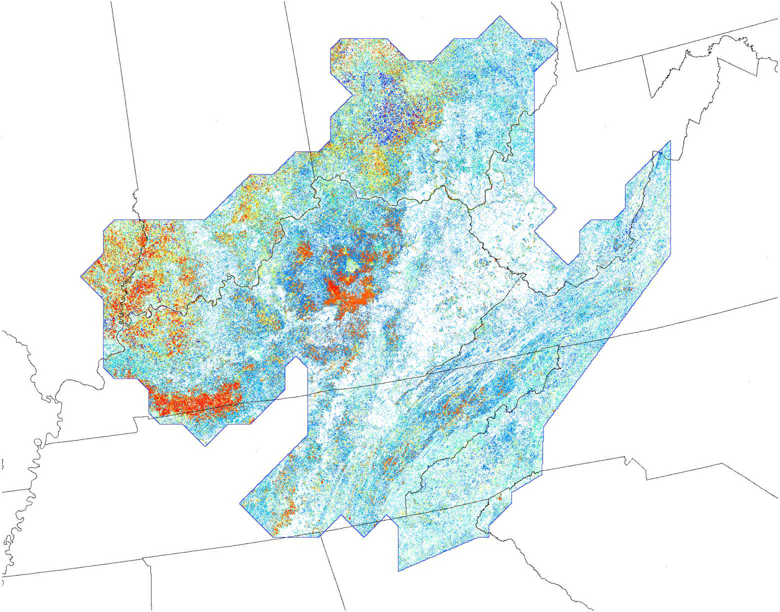
Squared Scores, Principal Component 22: Year 2009



Distribution of least abundant phenoclasses 10% area (k=50, 2009)



Transition Distance Map: Year 2008-2009, K=50



Conclusions and Future Work

- The combination of geospatiotemporal clustering and principal components analysis of NDVI time-series data appears promising:
  - Large transition distances indicate large departure from previous phenology.
  - Infrequent clusters and strong scores on lower-order principal components identify statistically unusual phenology.
- Transition distances excel at identifying high-mortality events (e.g. fires, storms)
- Cluster frequency and PCA techniques can identify less dramatic declines
- Many tasks to pursue in the future:
  - More validation with ground and aerial surveys
  - Establish biome-specific thresholds for transition distances, etc.
  - Build a library of declines attributed specific agents for use in complementary, supervised classification

Also see...

- Friday, 9:30 AM, Moscone West 2008
  ABSTRACT FINAL ID: B51P-07
  TITLE: Using Land Surface Phenology as the Basis for a National Early Warning System for Forest Disturbances
  SESSION TITLE: B51P. Beyond Earlier Spring: Diverse Phenological Responses to Climate Across Species and Ecosystems II
  AUTHORS: William Walter Hargrove, Joseph Spruce, Steven P. Norman, Forrest M Hoffman
- Friday, 2:40 PM, Moscone West 2008
  ABSTRACT FINAL ID: B53D-05
  TITLE: An Early Warning System for Identification and Monitoring of Disturbances to Forest Ecosystems
  SESSION TITLE: B53D. Remote Sensing of Long-Term Ecological Trends
  AUTHORS: Aaron A Marshall, Forrest M Hoffman, Jitendra Kumar, William W Hargrove, Joseph Spruce, Richard T Mills

References

William W. Hargrove, Joseph P. Spruce, Gerald E. Gasser, and Forrest M. Hoffman. Toward a national early warning system for forest disturbances using remotely sensed phenology. *Photogramm. Eng. Rem. Sens.*, 75(10):1150–1156, October 2009.