Detection of forest threats via unsupervised geospatiotemporal data mining of remotely sensed phenology data

Richard T. Mills[†], Forrest M. Hoffman[†], Jitendra Kumar[†], and William W. Hargrove[‡]

[†] Oak Ridge National Laboratory and
[‡] Eastern Forest Environmental Threat Assessment Center, USDA Forest Service

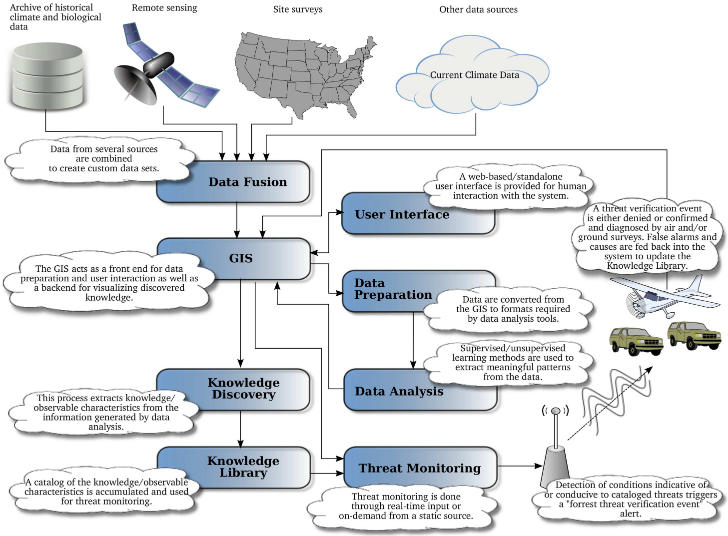
April 6, 2011
2011 US-IALE Symposium, Portland, OR

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



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.

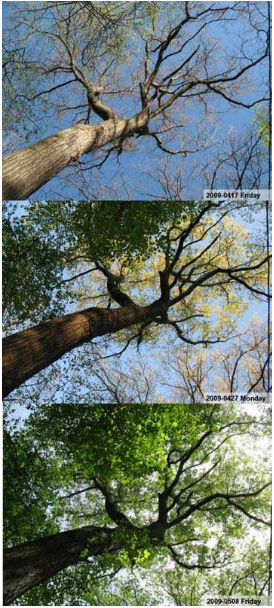
$$\mathrm{NDVI} = \frac{(\sigma_{\mathrm{nir}} - \sigma_{\mathrm{red}})}{(\sigma_{\mathrm{nir}} + \sigma_{\mathrm{red}})} \tag{1}$$

- 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.

MODIS MOD13 NDVI Product

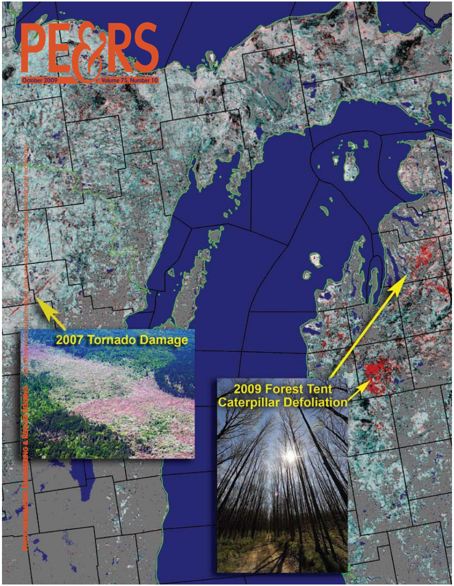
- The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard the Terra (EOS AM, N→S) and Aqua (EOS PM, S→N) satellites.
- Both view the entire surface of Earth every 1 to 2 days, acquiring data in 36 spectral bands.
- The MOD 13 product provides Gridded Vegetation Indices (NDVI and EVI) to characterize vegetated surfaces.
- Available are 6 produces at varying spatial (250 m, 1 km, 0.05°) and temporal (16-day, monthly) resolutions.
- The Terra and Aqua products are staggered in time so that a new product is available every 8 days.
- Results shown here are derived from the 16-day Terra MODIS product at 250 m resolution, processed by NASA Stennis Space Center.

- 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 (**?**).

- Map arithmetic approaches provide one straightforward approach.
- 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.



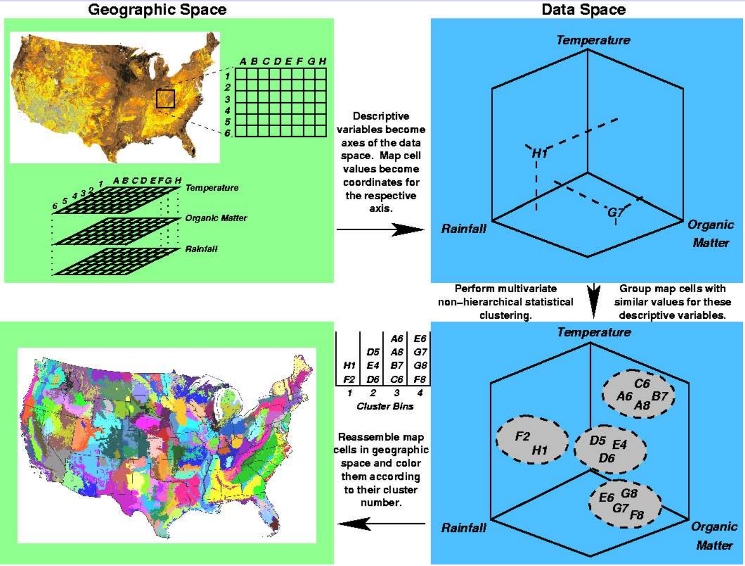
Data Mining for Change Detection

- 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 of NDVI history for the entire CONUS.
- These classifications use the full volume of available NDVI data (80GB here) to construct a potential basis for determining the "normal" seasonal and inter-seasonal variation expected at a geographic location.

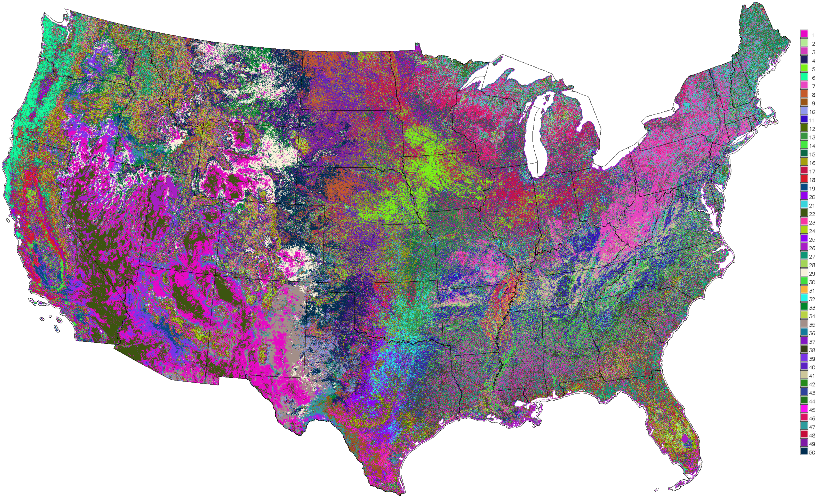
Clustering the MODIS NDVI data

- The 19B NDVI values in the data set are arranged as annual NDVI traces of 22 values, for each grid cell (146.4M records) in each of the six yearly maps,
- The entire set of NDVI traces for all years and map cells is combined into one 77 GB (single precision) data set of 878 22-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 six 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.

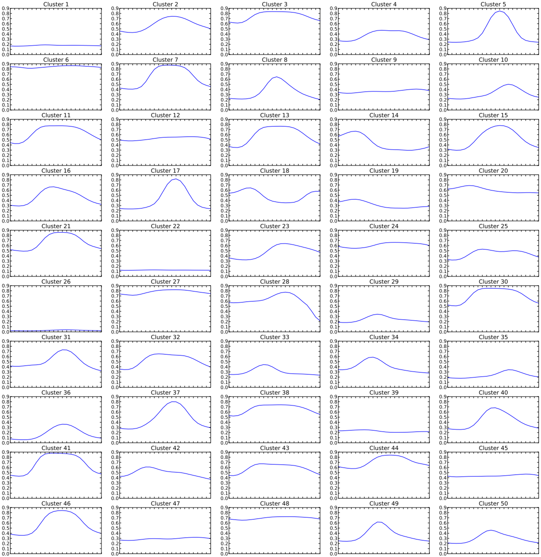
Geospatiotemporal Data Mining



50 Phenoregions for Year 2009

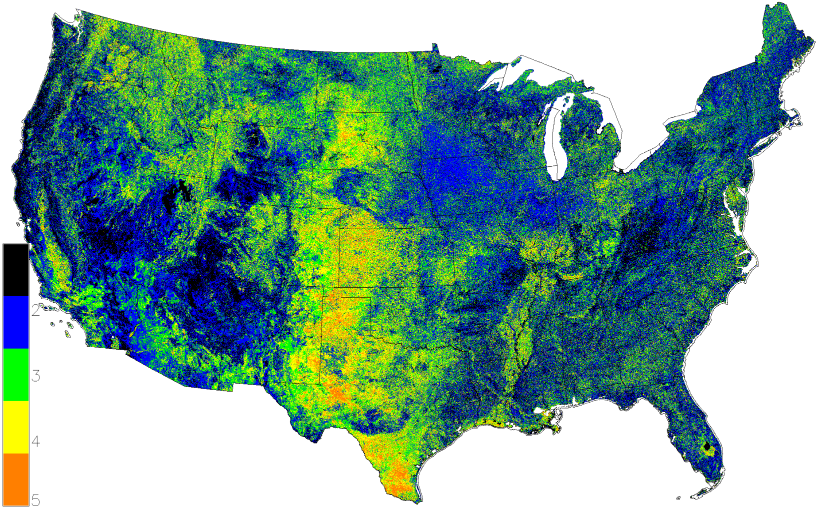


50 phenoclass prototypes (cluster centroids)



- The cluster analysis yields six maps, one for each year, that classify each cell as belonging to one of k phenoclasses (its cluster membership for that year).
- The time evolution of phenoclass assignment, or phenostate, for every cell in the map indicates a change in the phenological behavior and ecosystem productivity observed at that location due to natural or anthropogenic disturbance, forest regrowth, or ecosystem responses to interannual changes in climate.
- Comparison of the current phenostate with the nominal behavior of "healthy" vegetation indicated by the historical phenoclass assignment at every location in the CONUS forms the basis for an early warning system.
- One straightforward approach: examine the current phenostate compared to historical phenostates at a given map cell, and then flag the present state of a cell as "abnormal" if the cell has very infrequently or never occupied this state in the past.

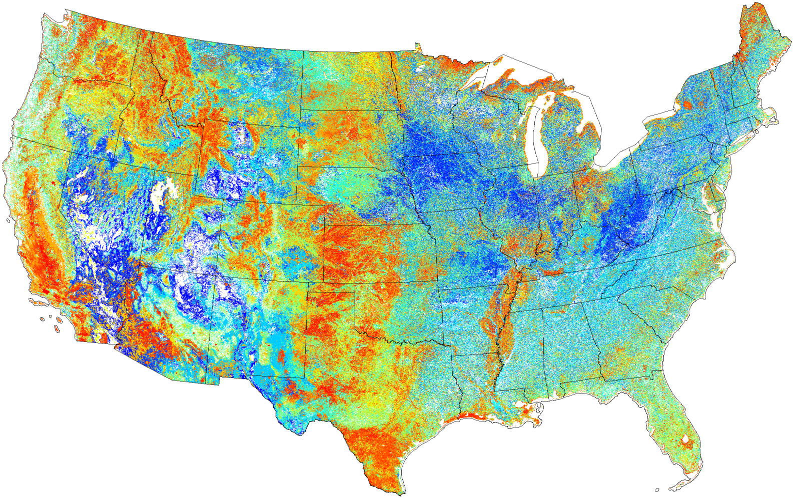
Cluster Persistence Map (2003–2008)



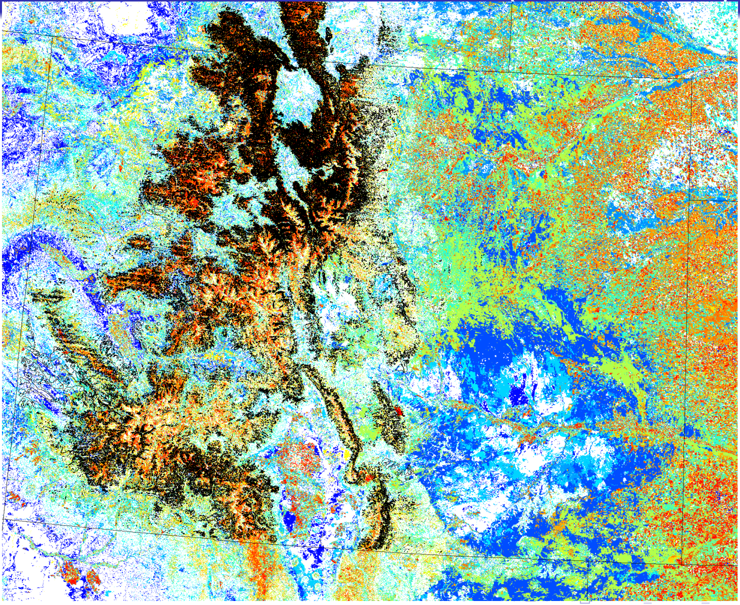
Euclidean “transition distance” between phenostates

- If k is too large, then the normal seasonal variation in NDVI will likely result in a different phenostate assignment each year, leading to many “false positive” commission errors, even though the different phenostates may, in fact, be very similar.
- An alternative approach is to create maps of the “transition distance” between years, plotting at each map cell the Euclidean distance between the new and old phenostate centroids; this distance gives a relative multivariate measure of how different the observed phenology is between the two years.

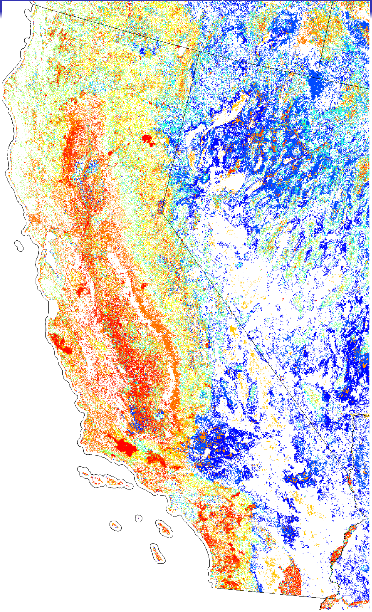
Maximum Transition Distances for (2003 – 2009)

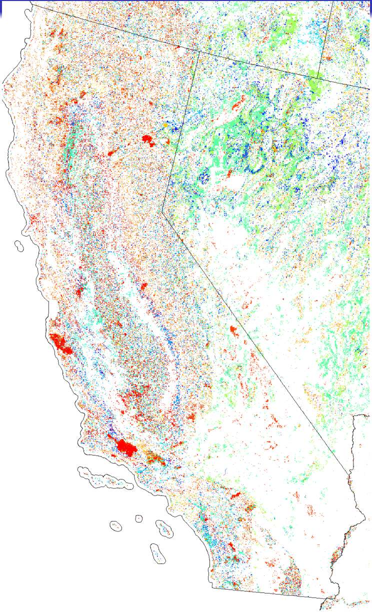


Mountain Pine Beetle in Colorado for (2008 – 2003)

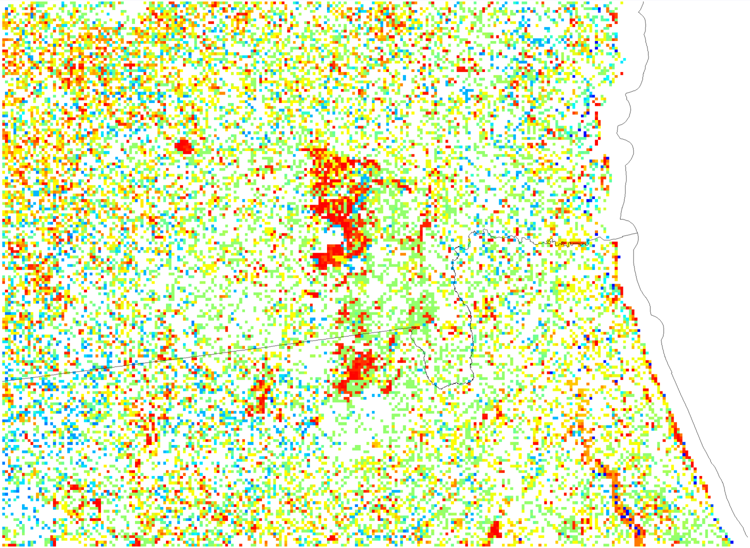


California Wildfires, 2006–2008

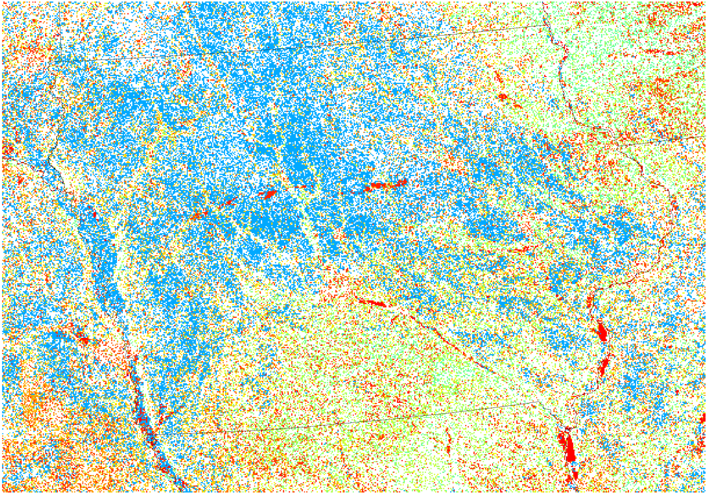




Okefenokee Bugaboo Scrub Fire, 2007



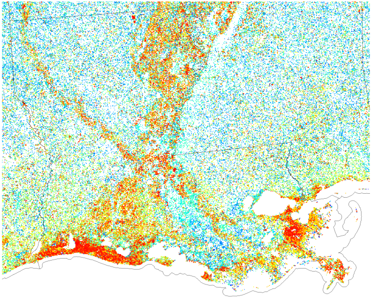
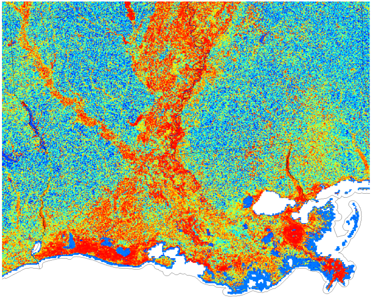
Iowa hailstorm damage, 2008–2009



Other clustering-based approaches

- Besides the year to year transition maps, other clustering-based approaches are possible.
- If the level of division, k, is large enough to allow it, very unusual profiles, which may indicate adverse events, will tend towards falling into clusters with very few members.
- (This, in fact, is how the original data quality issues were uncovered in the 2007 NDVI product, when with $k = 50$ we found clusters that had very few members, all from year 2007.)
- Another approach is to look for NDVI traces that are poor fits to their assigned cluster, as it seems reasonable to expect that adverse events will have anomalous traces that are far from the cluster centroid.

Hurricane damage, 2004–2005

Conclusions and Future Work

- Initial results of geospatiotemporal cluster analysis of phenology from MODIS NDVI are promising.
- The enhanced, accelerated k-means clustering algorithm enables the analysis of very large, high resolution remote sensing data.
- Other, complementary detection algorithms based on singular value decomposition will also be explored.
- Determining "normal" phenological patterns is difficult—due to interannual climate variability, spatially variable climate change trend, and relatively short satellite record—mortality events, like progressive Mountain Pine Beetle damage, are easily detected.
- The next step is to establish biome-specific thresholds based on interannual variability, obtain validation from ADS and ground surveys, and track and accumulate both loss and new growth for carbon accounting.
- Future work will build a library of phenostate transitions attributed to pests or pathogens for individual biomes, allowing the system to hypothesize about causes of future disturbances detected.

Acknowledgments

The authors wish to thank Joseph P. Spruce at the NASA Stennis Space Center for providing quality controlled NDVI maps generated from the MODIS MOD13 Collection 5 product.

This research was sponsored by the U.S. Department of Agriculture Forest Service, Eastern Forest Environmental Threat Assessment Center. This research used resources of the National Center for Computational Science at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

References