Integrating Statistical and Expert Knowledge to Develop Phenoregions for the Continental United States

Forrest M. Hoffman[1], Jitendra Kumar[1], William W. Hargrove[2], Steven P. Norman[2], and Bjørn-Gustaf J. Brooks[2]

[1]Oak Ridge National Laboratory and [2]USDA Forest Service
Eastern Forest Environmental Threat Assessment Center (EFETAC)

May 22, 2017

JpGU–AGU Joint Meeting 2017

May 20–25, 2017 – Makuhari Messe International Conference Hall, Chiba, Japan

OAK RIDGE National Laboratory

CLIMATE CHANGE SCIENCE INSTITUTE
OAK RIDGE NATIONAL LABORATORY

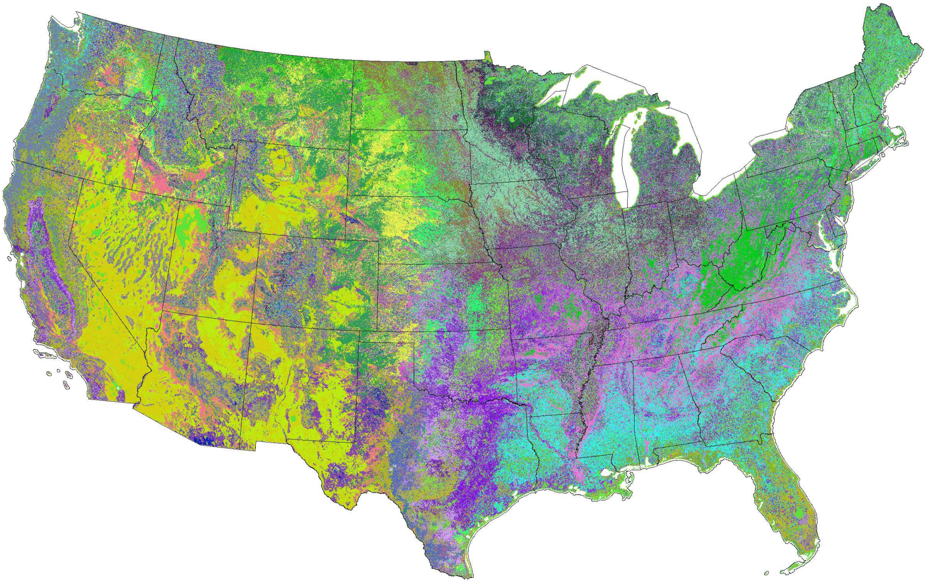
FOREST SERVICE U S DEPARTMENT OF AGRICULTURE

ForWarn
Satellite-Based Change Recognition and Tracking

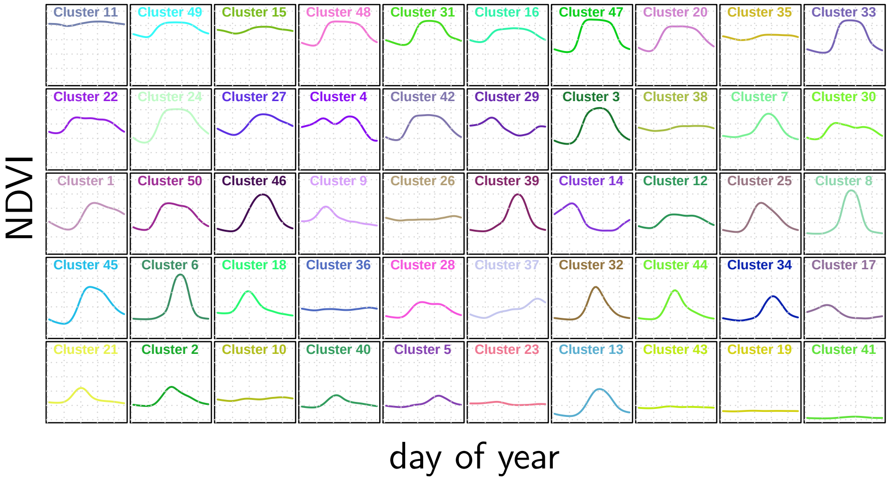
Clustering MODIS NDVI to Produce Phenoregions

- Hoffman and Hargrove previously used *k*-means clustering to detect brine scars from hyperspectral data (Hoffman, 2004) and to classify phenologies from monthly climatology and 17 years of 8 km NDVI from AVHRR (White et al., 2005).
- This data mining approach requires high performance computing to analyze the entire body of the high resolution MODIS NDVI record for the continental U.S.
- **$>$107B NDVI values**, consisting of **$\sim$146.4M cells** for the CONUS at 250 m resolution with **46 maps per year** for **16 years** (2000–2015), analyzed using *k*-means clustering.
- The annual traces of NDVI for every year and map cell are combined into one **431 GB single-precision binary** data set of 46-dimensional observation vectors.
- Clustering yields 16 phenoregion maps in which each cell is classified into one of *k* phenoclasses that represent prototype annual NDVI traces.

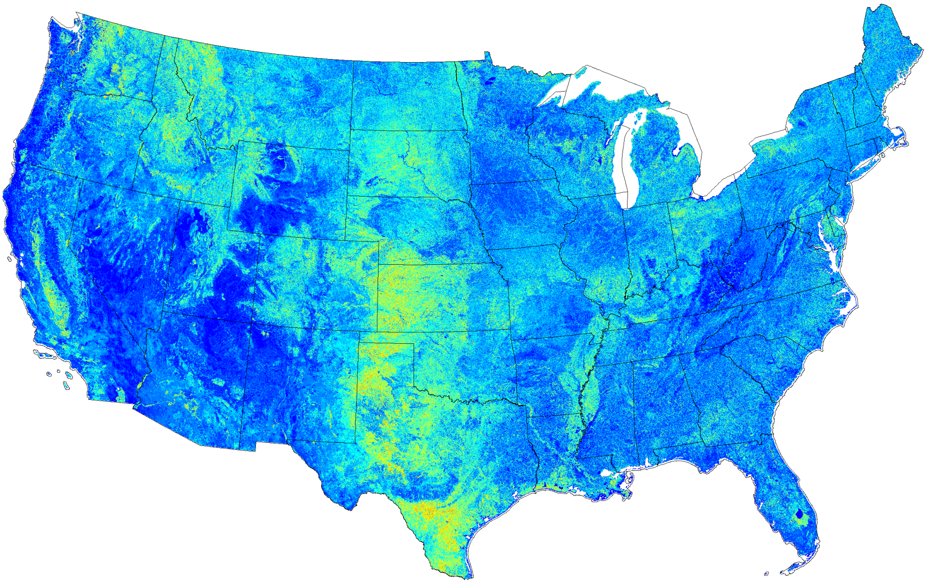
50 Phenoregions for year 2012 (Random Colors)



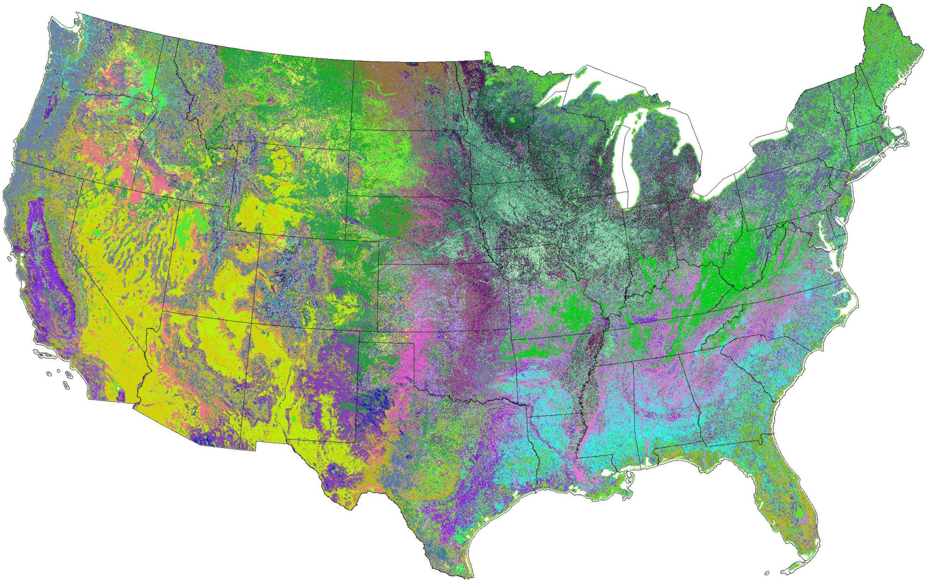
50 Phenoregion Prototypes (Random Colors)



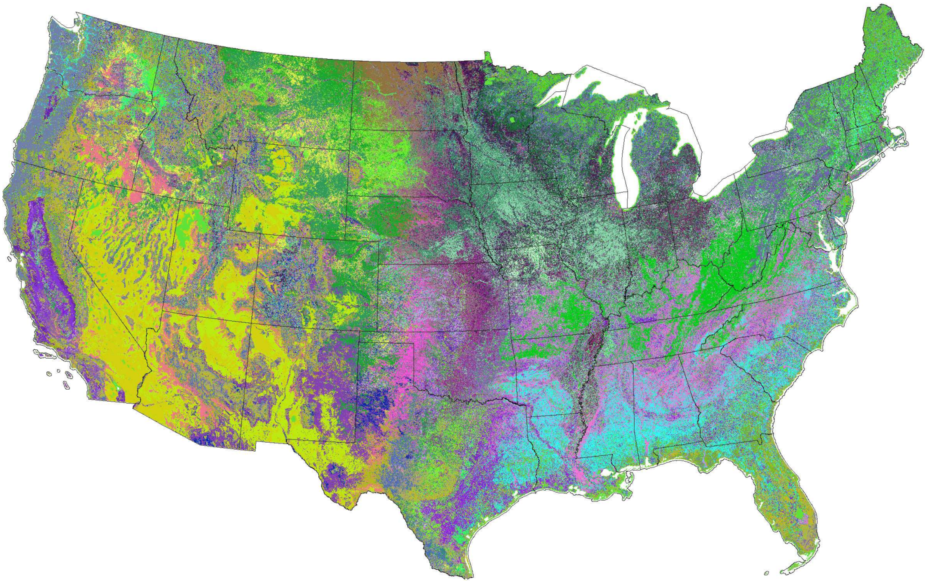
50 Phenoregions Persistence



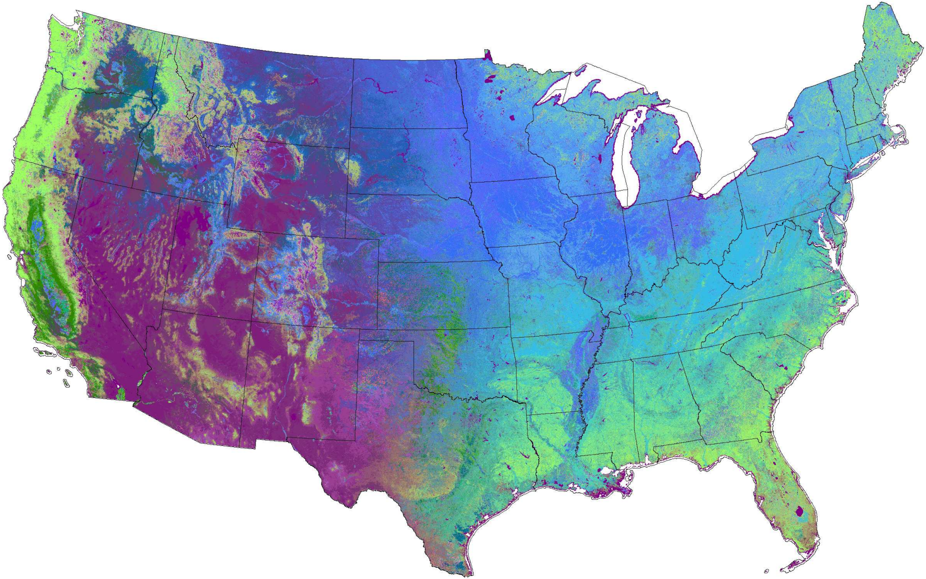
50 Phenoregions Mode (Random Colors)



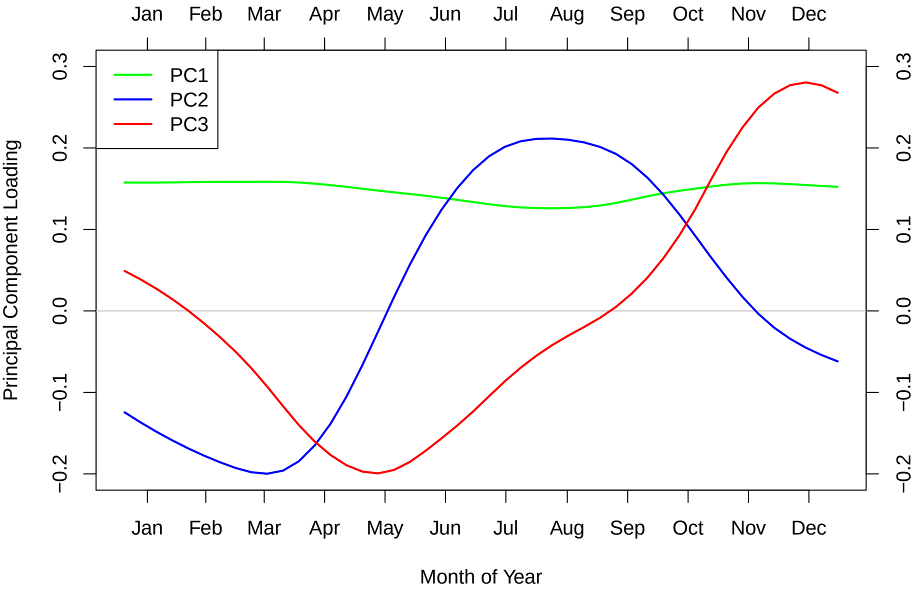
50 Phenoregions Max Mode (Random Colors)



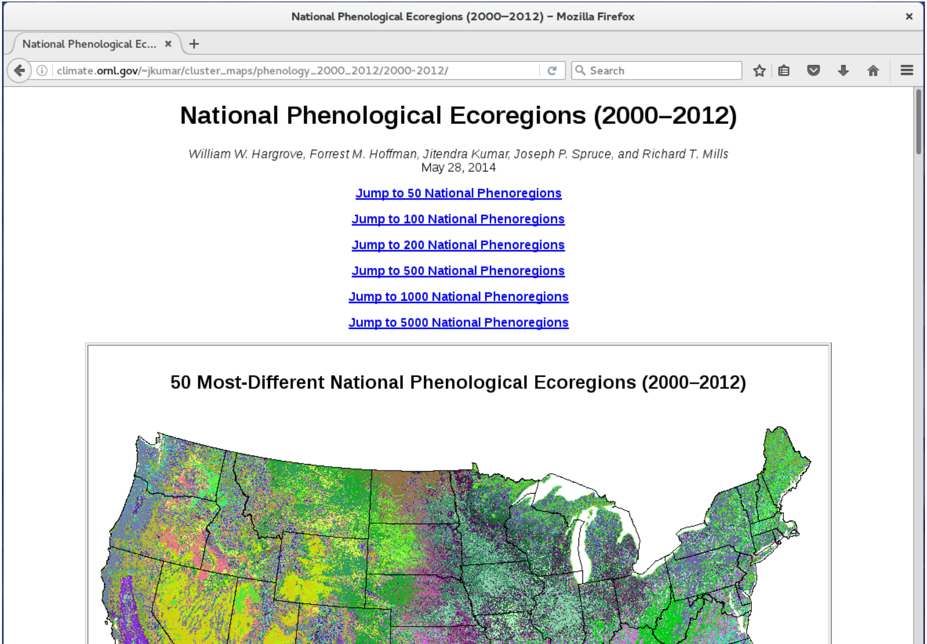
50 Phenoregions Max Mode (Similarity Colors)



50 Phenoregions Max Mode (Similarity Colors Legend)

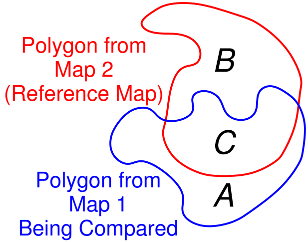


Phenoregions Clearinghouse



Mapcurves: A Method for Comparing Categorical Maps

- Hargrove et al. (2006) developed a method for quantitatively comparing categorical maps that is
  - independent of differences in resolution,
  - independent of the number of categories in maps, and
  - independent of the directionality of comparison.

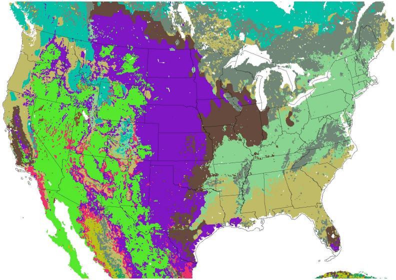


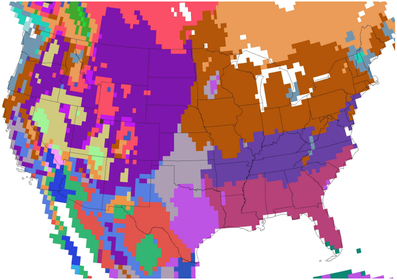
Goodness of Fit (GOF) is a unitless measure of spatial overlap between map categories:

$$\mathrm{GOF} = \sum_{\text{polygons}} \frac{C}{B + C} \times \frac{C}{A + C}$$

- GOF provides "credit" for the area of overlap, but also "debit" for the area of non-overlap.
- Mapcurves comparisons allow us to reclassify any map in terms of any other map (*i.e.*, color Map 2 like Map 1).
- A greyscale GOF map shows the degree of correspondence between two maps based on the highest GOF score.

Expert-Derived Land Cover/Vegetation Type Maps

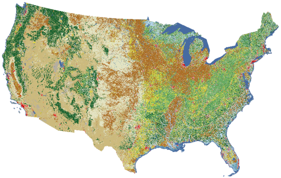

Foley Land Cover


Holdridge Life Zones

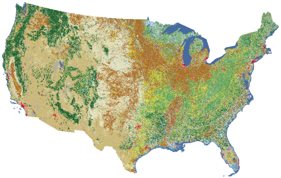
	Expert Map	# Cats
1.	DeFries UMd Vegetation	12
2.	Foley Land Cover	14
3.	Fedorova, Volkova, and Varlyguin World Vegetation Cover	31
4.	GAP National Land Cover	578
5.	Holdridge Life Zones	25
6.	Küchler Types	117
7.	BATS Land Cover	17
8.	IGBP Land Cover	16
9.	Olson Global Ecoregions	49
10.	Seasonal Land Cover Regions	194
11.	USGS Land Cover	24
12.	Leemans-Holdridge Life Zones	26
13.	Matthews Vegetation Types	19
14.	Major Land Resource Areas	197
15.	National Land Cover Database 2006	16
16.	Wilson, Henderson, & Sellers Primary Vegetation Types	23
17.	Landfire Vegetation Types	443

Label Stealing: Having your cake and eating it too!

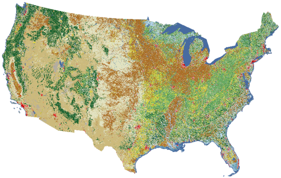
- Clustering is an unsupervised classification technique, so phenoregions have no descriptive labels like **Deciduous Forest**.
- **Label stealing** allows us to perform automated "supervision" to "steal" the best human-created descriptive labels to assign to phenoregions.
- We employ the **Mapcurves GOF** to select the best ecoregion labels from land cover maps constructed by human experts.



(a) 2001



(b) 2006



(c) 2011

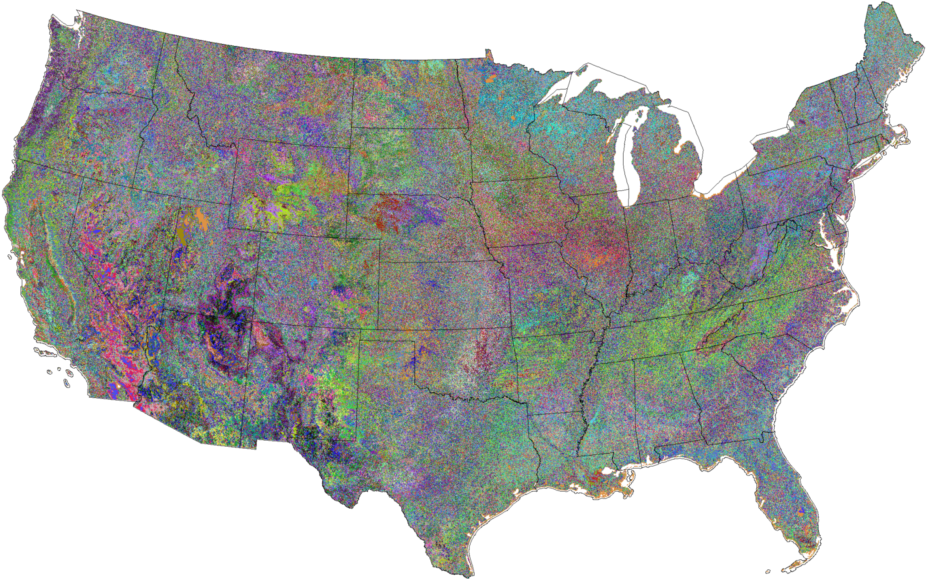
Figure: The National Land Cover Database (NLCD) provides land cover maps at 30 m resolution updated every 5 years.

Areas from the National Land Cover Database (NLCD)

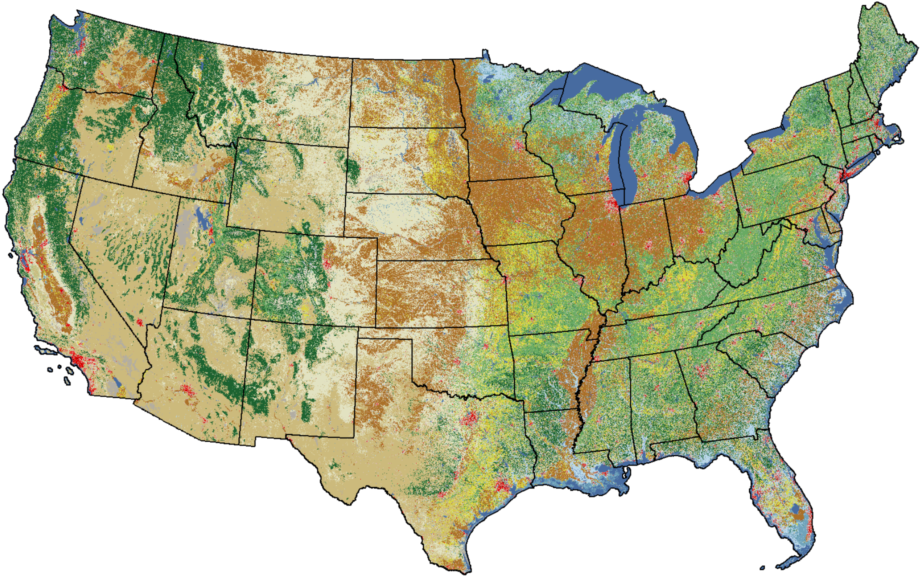
Land Cover Type	2001	2006	2011
11-Open Water	103.770	103.210	103.210
12-Perennial Ice Snow	0.355	0.355	0.355
21-Developed Open Space	64.117	64.750	64.750
22-Developed Low Intensity	28.162	28.826	28.826
23-Developed Medium Intensity	10.980	12.324	12.324
24-Developed High Intensity	3.916	4.404	4.404
31-Barren Land	23.891	24.419	24.419
41-Deciduous Forest	220.317	218.936	218.936
42-Evergreen Forest	240.573	235.658	235.658
43-Mixed Forest	42.449	41.302	41.302
52-Shrub/Scrub	423.604	426.936	426.936
71-Grassland Herbaceous	284.981	287.860	287.860
81-Pasture Hay	135.133	133.563	133.563
82-Cultivated Crops	309.648	309.078	309.078
90-Woody Wetlands	77.921	77.811	77.811
95-Emergent Herbaceous Wetlands	25.062	25.448	25.448

Units of millions of acres

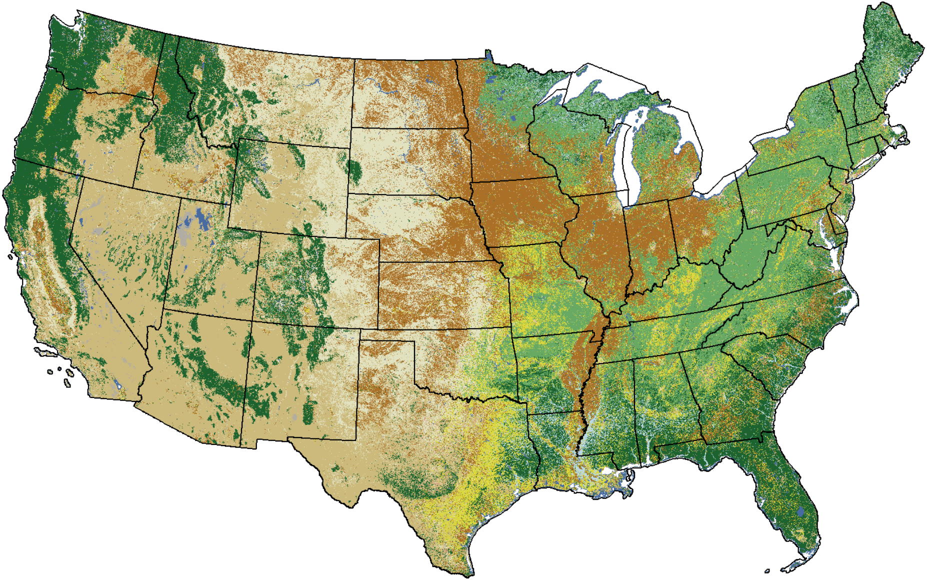
1000 Phenoregions Max Mode (Random Colors)



NLCD 2006



Reclassed 1000 Phenoregions Max Mode

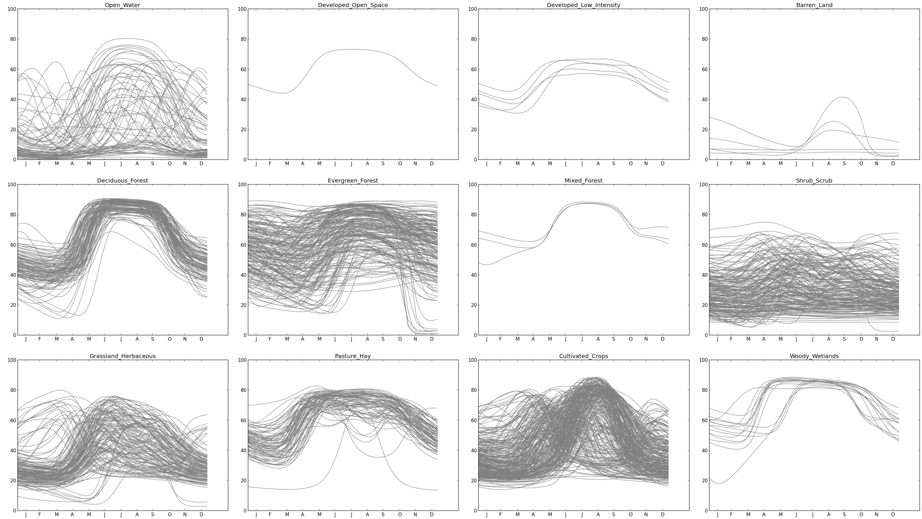


Areas for Stolen Labels on 1000 Phenoregions

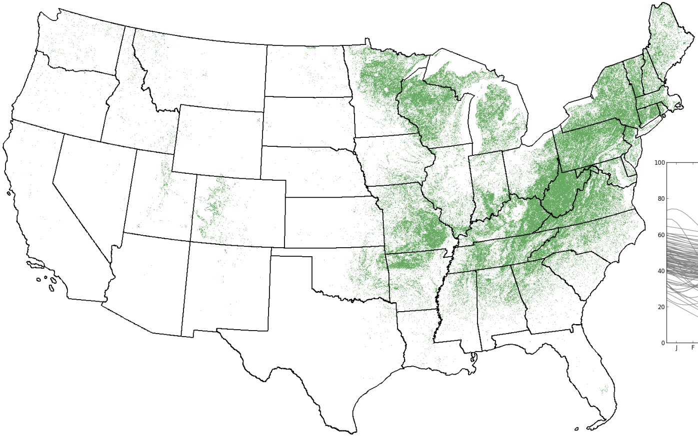
Land Cover Type	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
11-Open Water	37.0	36.6	37.9	38.0	40.4	41.9	38.7	39.5	41.9	41.1
21-Developed Open Space	3.1	3.0	4.2	2.1	2.5	4.0	3.4	4.1	4.0	2.4
22-Developed Low Intensity	8.6	10.9	18.3	10.7	15.5	12.4	12.8	18.2	13.7	13.1
31-Barren Land	9.1	7.6	9.4	9.1	8.8	7.1	7.7	8.3	9.3	9.6
41-Deciduous Forest	254.0	270.0	256.4	266.8	277.3	269.1	266.5	225.3	269.7	265.0
42-Evergreen Forest	321.6	320.2	357.4	322.1	316.3	326.0	315.8	315.3	328.1	316.1
43-Mixed Forest	10.3	11.6	7.3	7.1	7.4	7.2	8.9	6.5	6.2	6.6
52-Shrub/Scrub	523.1	515.7	558.1	501.7	480.7	469.7	537.6	445.7	487.3	464.6
71-Grassland Herbaceous	251.1	267.1	219.8	276.2	251.8	305.6	270.5	332.8	280.9	314.0
81-Pasture Hay	142.6	147.3	141.6	155.9	166.2	138.9	132.1	190.4	127.7	136.1
82-Cultivated Crops	347.2	323.5	304.1	318.1	342.2	332.9	318.8	319.9	347.9	348.6
90-Woody Wetlands	33.4	27.8	26.9	33.3	32.1	26.4	28.6	35.1	24.6	24.1

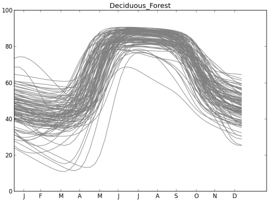
Units of millions of acres

Reclassed Phenoregion Centroid Traces

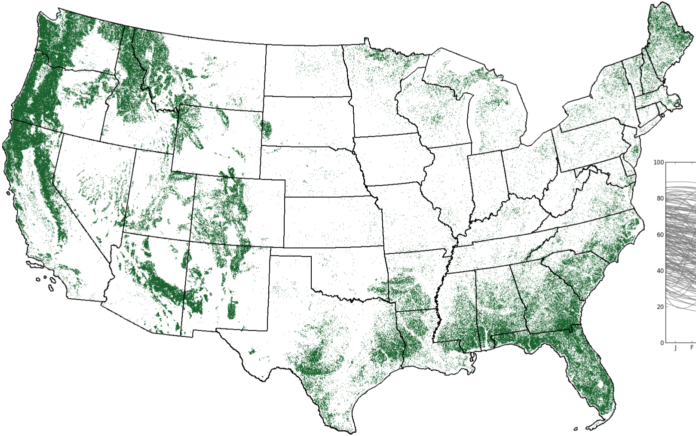


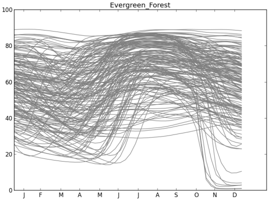
Reclassed Deciduous Phenoregion (Year 2000)



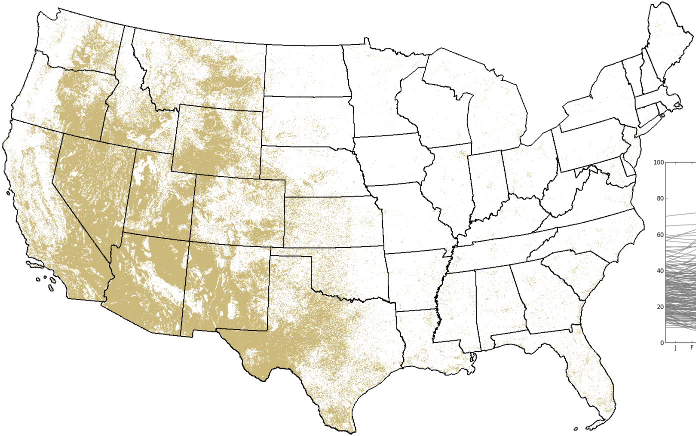


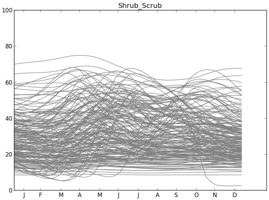
Reclassed Evergreen Phenoregion (Year 2000)





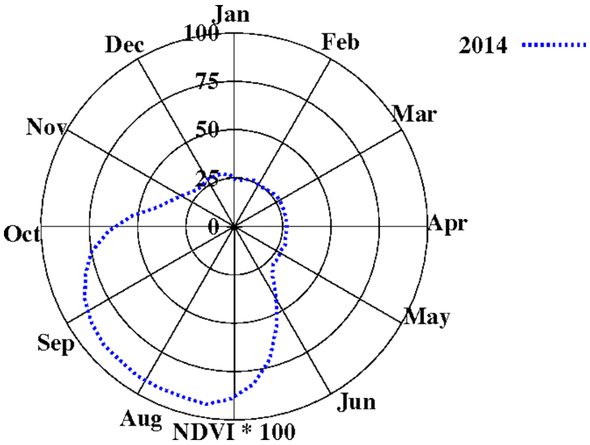
Reclassed Shrub/Scrub Phenoregion (Year 2000)





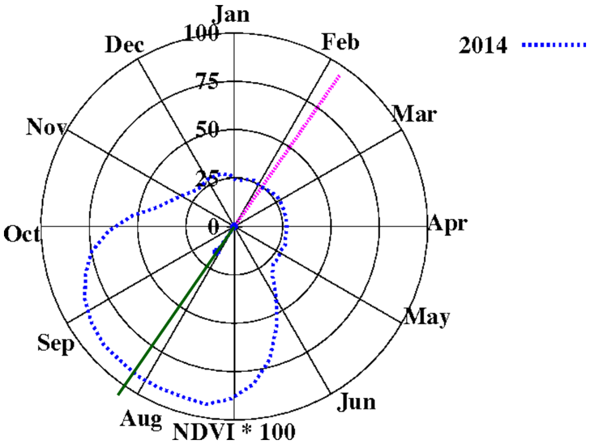
Defining a Phenological Year with Polar Plots

- Plotting the 46 8-day NDVI magnitude vectors clockwise on a polar plot facilitates comparison across years
- Plot is (nearly) circular for evergreen vegetation
- Plot is elliptical for seasonal or deciduous vegetation
- A phenological year can easily be identified differently for every map cell



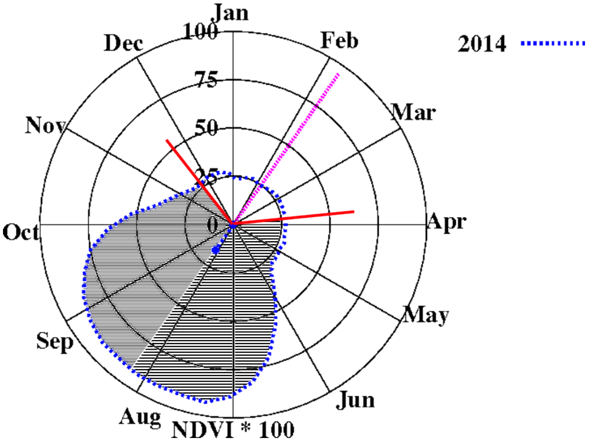
Defining a Phenological Year with Polar Plots

- The distance from the origin to the centroid of the ellipse is a measure of the degree of seasonality
- The green vector represents the halfway point through the phenological year
- The opposing purple vector represents the start/end time of the phenological year
- We define the active growing season using completion thresholds



Defining a Phenological Year with Polar Plots

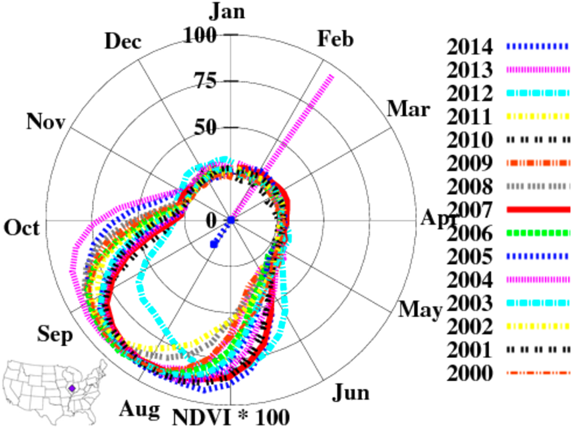
- We selected **completion thresholds of 15% and 80%** of the area accumulated under the curve to define the active growing season
- Starting at the 50% mid-growing season point and sweeping backward by 35% area gives the timing of the 15% threshold
- Starting at the 50% mid-growing season point and sweeping forward by 30% area gives the timing of the 80% threshold



Defining a Phenological Year with Polar Plots

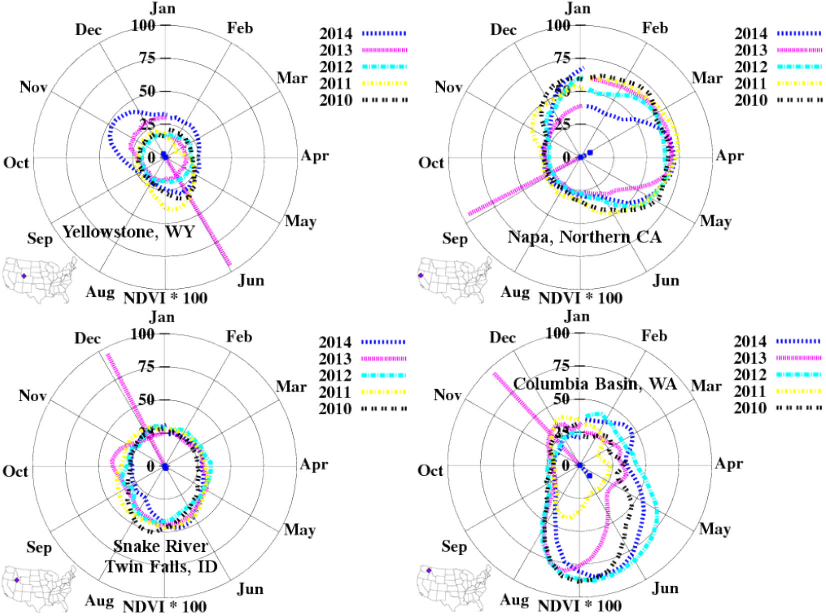
15 years of agriculture in Iowa, showing annual cycles and their variations.

- Here, the vector from the origin to the centroid points to early August
- Thus, the phenological year begins in early February
- Every year has nearly the same area under the curve



In 2012, the U.S. experienced a severe drought, but also an early spring that resulted in compensatory effects on carbon uptake (Wolf et al., 2016). The 2013 growing season extended later than most, and 2002 has one of the smallest areas, indicating reduced carbon uptake.

Defining a Phenological Year with Polar Plots



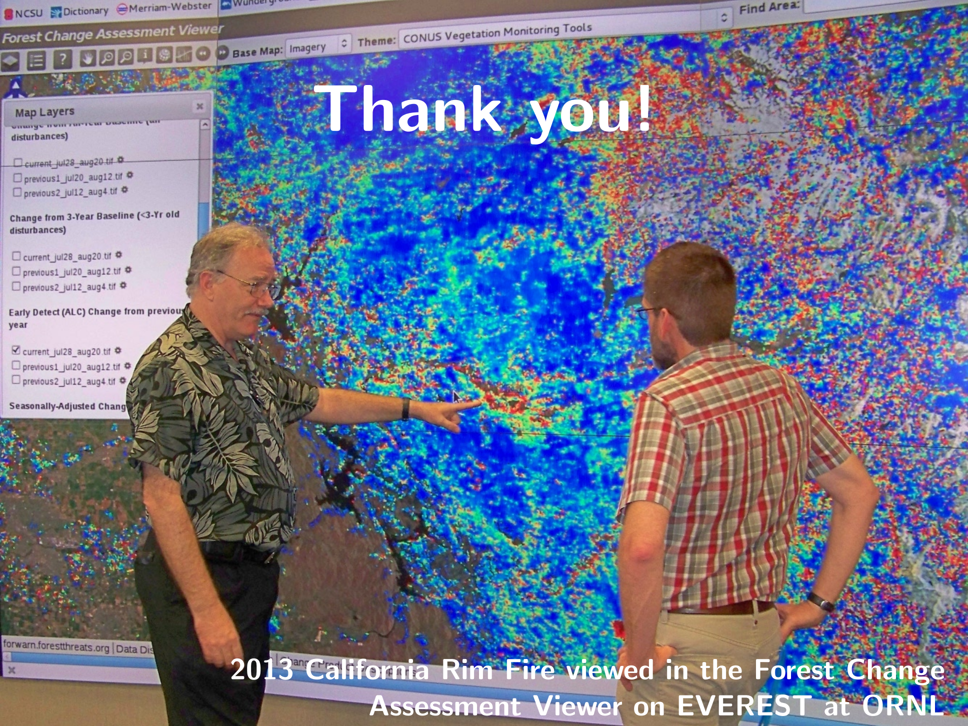
Summary and Conclusions

- **Phenoregions** quantitatively derived from NDVI observations provide a useful framework for understanding vegetation distributions and dynamics. [Multivariate Spatio-Temporal Clustering (MSTC)]
- **Label stealing** enables automated "supervision" of unsupervised data mining for naming phenoregions. [Mapcurves]
- **Polar plots** offer an intuitive visualization technique, provide an easy method for defining a site-specific phenological year, and facilitate comparisons of vegetation dynamics.
  - Effects of snow can be reduced by integrating NDVI only from 15% to 80% of area under the curve
  - Location/biome-specific phenological (a)synchrony (of vegetation and birds and insects) *and their relationships* are obvious
  - Changes in ecological (a)synchrony (due to climate or disturbance) are expressed as shifts in threshold timing and shape of the annual trace
  - Comparisons across latitudes/hemispheres is enabled by overlaying and rotating annual NDVI traces to align start/middle/end of local seasons

Acknowledgments

This research was sponsored by the U. S. Department of Agriculture Forest Service Eastern Forest Environmental Threat Assessment Center (EFETAC) and by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the U. S. Department of Energy Office of Science. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.

Thank you!



2013 California Rim Fire viewed in the Forest Change Assessment Viewer on EVEREST at ORNL

References

W. W. Hargrove, F. M. Hoffman, and P. F. Hessburg. Mapcurves: A quantitative method for comparing categorical maps. *J. Geograph. Syst.*, 8(2):187–208, July 2006. doi: 10.1007/s10109-006-0025-x.

F. M. Hoffman. Analysis of reflected spectral signatures and detection of geophysical disturbance using hyperspectral imagery. Master's thesis, University of Tennessee, Department of Physics and Astronomy, Knoxville, Tennessee, USA, Nov. 2004.

M. A. White, F. Hoffman, W. W. Hargrove, and R. R. Nemani. A global framework for monitoring phenological responses to climate change. *Geophys. Res. Lett.*, 32(4): L04705, Feb. 2005. doi: 10.1029/2004GL021961.

S. Wolf, T. F. Keenan, J. B. Fisher, D. D. Baldocchi, A. R. Desai, A. D. Richardson, R. L. Scott, B. E. Law, M. E. Litvak, N. A. Brunsell, W. Peters, and I. T. van der Laan-Luijkx. Warm spring reduced carbon cycle impact of the 2012 US summer drought. *Proc. Nat. Acad. Sci.*, 113(21):5880–5885, May 2016. doi: 10.1073/pnas.1519620113.