Representativeness-based Sampling Network Design for NGEE and Identifying Phenoregions for the Conterminous U.S.

Forrest M. Hoffman[†‡], Jitendra Kumar[†], Damian Maddalena[†], and William W. Hargrove*

[†]Oak Ridge National Laboratory, [‡]University of California-Irvine, and *USDA Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC)

Fourth Workshop on Understanding Climate Change from Data, Boulder, Colorado, USA

July 1, 2014

U.S. DEPARTMENT OF ENERGY

UCI

FOREST SERVICE U.S. DEPARTMENT OF AGRICULTURE

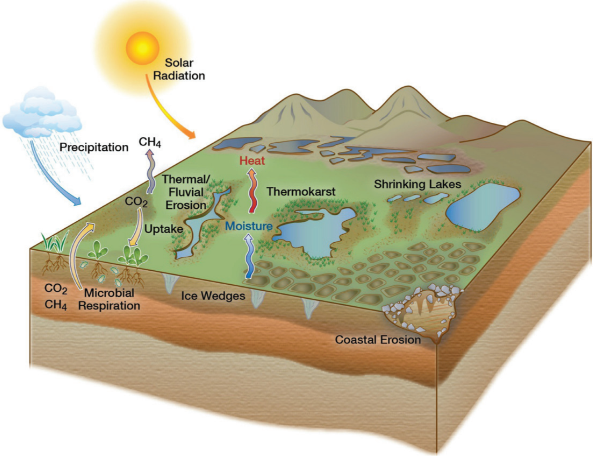
Climate Change Science Institute AT OAK RIDGE NATIONAL LABORATORY

OAK RIDGE NATIONAL LABORATORY

MANAGED BY UT-BATTELLE FOR THE U.S. DEPARTMENT OF ENERGY

Next-Generation Ecosystem Experiments (NGEE Arctic)

http://ngee.ornl.gov/

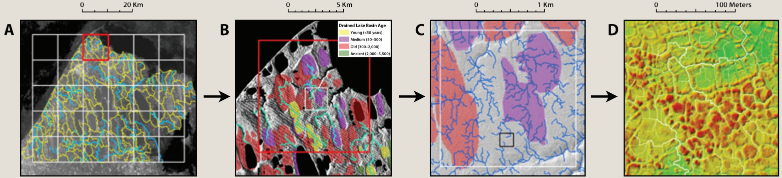


The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Integrating Across Scales

- NGEE Arctic process studies and observations are strongly linked to model development and application for improving process representation, initialization, calibration, and evaluation.
- A hierarchy of models will be deployed at fine, intermediate, and climate scales to connect observations to models and models to each other in a quantitative up-scaling and down-scaling framework.

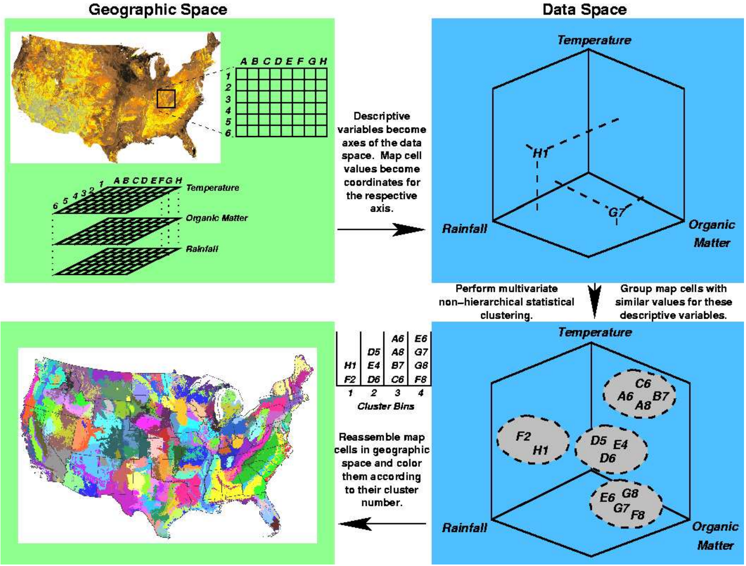
Hydrologic and Geomorphic Features at Multiple Scales. *At the scale of (A) a high-resolution ESM, (B) a single ESM grid cell, (C) a 2 × 2 km domain of high-resolution Light Detection and Ranging (LiDAR) topographic data, and (D) polygonal ground. Yellow outlines in panel A show geomorphologically stable hydrologic basins, connected by stream channels (blue). Colored regions in panels B and C show multiple drained thaw lake basins within a single 10 × 10 km grid cell (B) or a 2 × 2 km domain (C), with progressively more detailed representation of stream channels (blue). Colors in panel D represent higher (red) to lower (green) surface elevations for a fine-scale subregion, with very fine drainage features (white). [Los Alamos National Laboratory, University of Alaska Fairbanks, and University of Texas at El Paso]*



Quantitative Sampling Network Design

- Resource and logistical constraints limit the frequency and extent of observations, necessitating the development of a systematic sampling strategy that objectively represents environmental variability at the desired spatial scale.
- Required is a methodology that provides a quantitative framework for informing site selection and determining the representativeness of measurements.
- Multivariate spatiotemporal clustering (MSTC) was applied at the landscape scale (4 km^2) for the State of Alaska to demonstrate its utility for representativeness and scaling.
- An extension of the method applied by Hargrove and Hoffman for design of National Science Foundation's (NSF's) National Ecological Observatory Network (NEON) domains (Schimel et al., 2007; Keller et al., 2008).

Multivariate Spatiotemporal Clustering (MSTC)

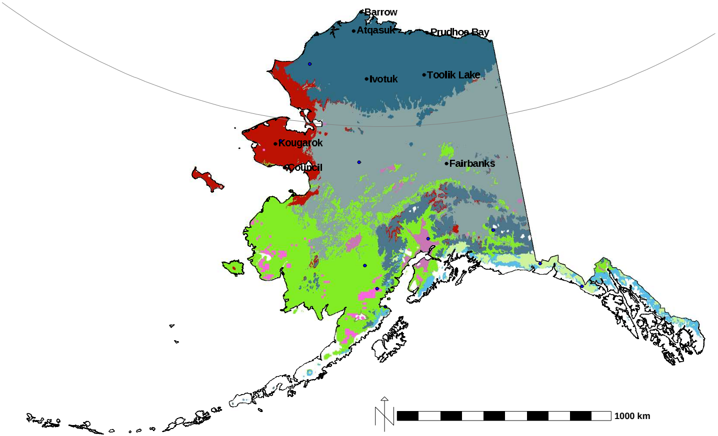


Data Layers

Table: 37 variables averaged for 2000–2009 and 2090–2099

Description	Number/Name	Units	Source
Monthly mean air temperature	12	°C	GCM
Monthly mean precipitation	12	mm	GCM
Day of freeze	mean	day of year	GCM
	standard deviation	days	
Day of thaw	mean	day of year	GCM
	standard deviation	days	
Length of growing season	mean	days	GCM
	standard deviation	days	
Maximum active layer thickness	1	m	GIPL
Warming effect of snow	1	°C	GIPL
Mean annual ground temperature at bottom of active layer	1	°C	GIPL
Mean annual ground surface temperature	1	°C	GIPL
Thermal offset	1	°C	GIPL
Limnicity	1	%	NHD
Elevation	1	m	SRTM

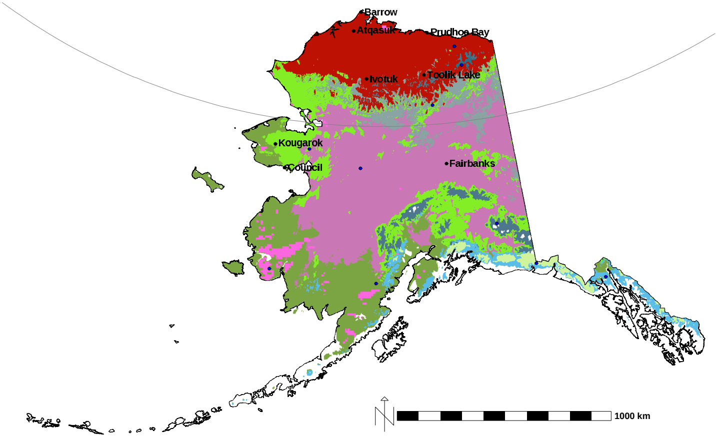
10 Alaska Ecoregions (2000–2009)



(Hoffman et al., 2013)

Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

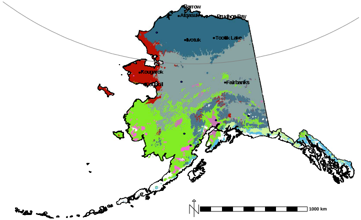
10 Alaska Ecoregions (2090–2099)



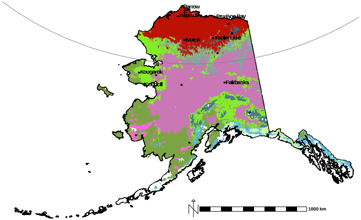
(Hoffman et al., 2013)

Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

10 Alaska Ecoregions, Present and Future



2000–2009



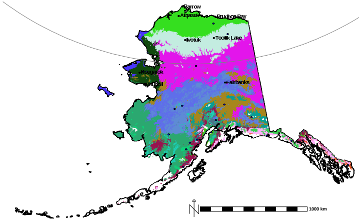
2090–2099

(Hoffman et al., 2013)

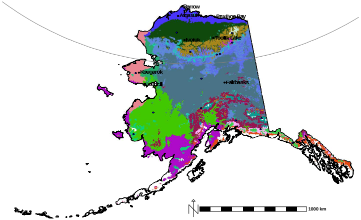
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At this level of division, the conditions in the large boreal forest become compressed onto the Brooks Range and the conditions on the Seward Peninsula “migrate” to the North Slope.

20 Alaska Ecoregions, Present and Future



2000–2009



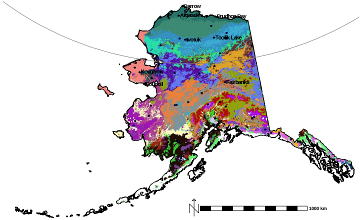
2090–2099

(Hoffman et al., 2013)

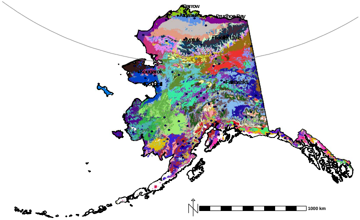
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At this level of division, the two primary regions of the Seward Peninsula and that of the northern boreal forest replace the two regions on the North Slope almost entirely.

50 and 100 Alaska Ecoregions, Present



$k = 50$, 2000–2009



$k = 100$, 2000–2009

(Hoffman et al., 2013)

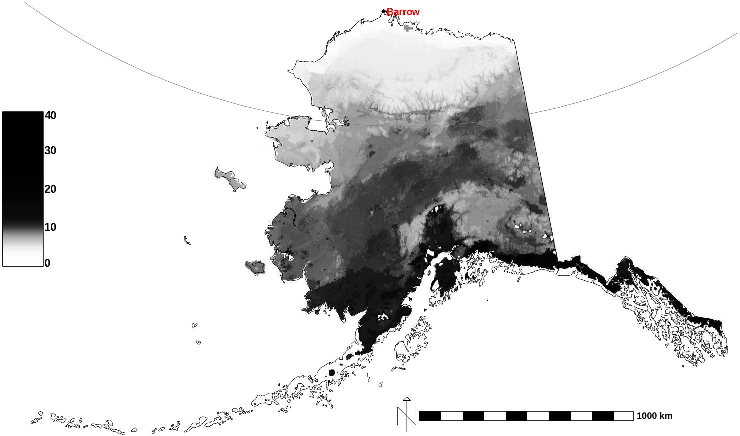
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At high levels of division, some regions vanish between the present and future while other region representing new combinations of environmental conditions come into existence.

NGEE Arctic Site Representativeness

- This representativeness analysis uses the standardized n-dimensional data space formed from all input data layers.
- In this data space, the Euclidean distance between a sampling location (like Barrow) and every other point is calculated.
- These data space distances are then used to generate grayscale maps showing the similarity, or lack thereof, of every location to the sampling location.
- In the subsequent maps, white areas are well represented by the sampling location or network, while dark and black areas as poorly represented by the sampling location or network.
- This analysis assumes that the climate surrogates maintain their predictive power and that no significant biological adaptation occurs in the future.

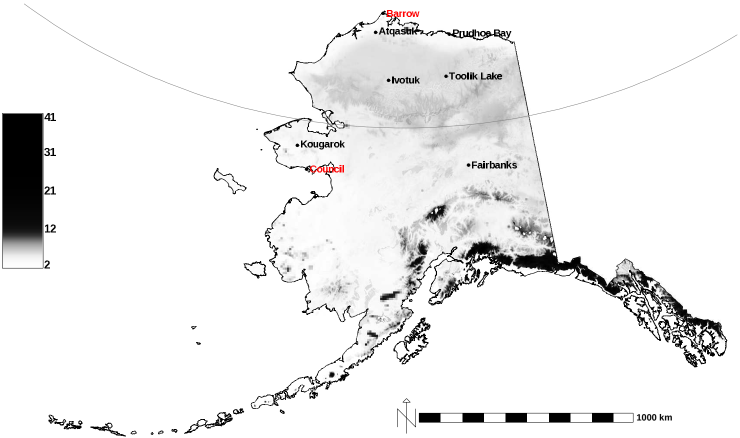
Present Representativeness of Barrow or "Barrow-ness"



(Hoffman et al., 2013)

Light-colored regions are well represented and dark-colored regions are poorly represented by the sampling location listed in **red**.

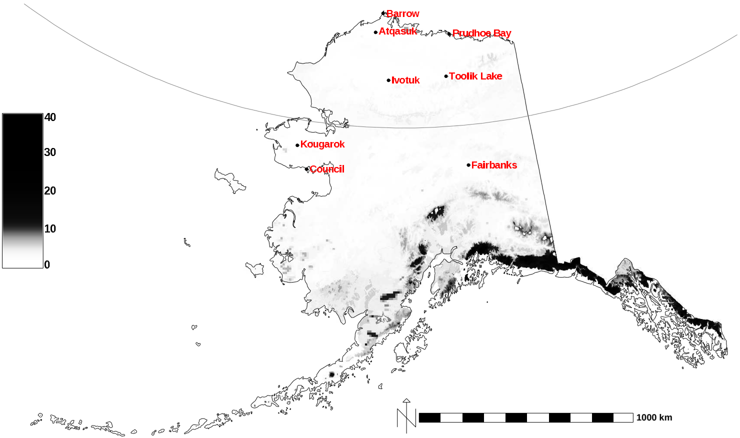
Network Representativeness: Barrow + Council



(Hoffman et al., 2013)

Light-colored regions are well represented and dark-colored regions are poorly represented by the sampling location listed in **red**.

Network Representativeness: All 8 Sites



(Hoffman et al., 2013)

Light-colored regions are well represented and dark-colored regions are poorly represented by the sampling location listed in **red**.

State Space Dissimilarities: 8 Sites, Present (2000–2009)

Table: Site state space distances for the present (2000–2009) with DEM

Sites	Council	Atqasuk	Ivotuk	Toolik Lake	Kougarok	Prudhoe Bay	Fairbanks
Barrow	9.13	4.53	5.90	5.87	7.98	3.57	12.16
Council		8.69	6.37	7.00	2.28	8.15	5.05
Atqasuk			5.18	5.23	7.79	1.74	10.66
Ivotuk				1.81	5.83	4.48	7.90
Toolik Lake					6.47	4.65	8.70
Kougarok						7.25	5.57
Prudhoe Bay							10.38

State Space Dissimilarities: 8 Sites, Future (2090–2099)

Table: Site state space distances for the future (2090–2099) with DEM

Sites	Council	Atqasuk	Ivotuk	Toolik Lake	Kougarok	Prudhoe Bay	Fairbanks
Barrow	8.87	4.89	6.88	6.94	8.04	4.18	11.95
Council		8.82	6.93	7.74	2.43	8.24	5.66
Atqasuk			5.86	5.84	8.15	2.30	10.16
Ivotuk				2.01	7.27	4.75	7.51
Toolik Lake					7.81	5.00	8.33
Kougarok						7.89	6.42
Prudhoe Bay							9.81

State Space Dissimilarities: 8 Sites, Present and Future

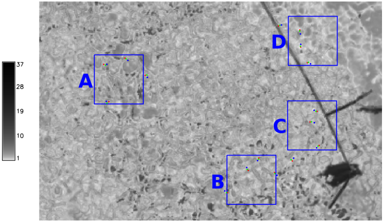
Table: Site state space distances between the present (2000–2009) and the future (2090–2099) with DEM

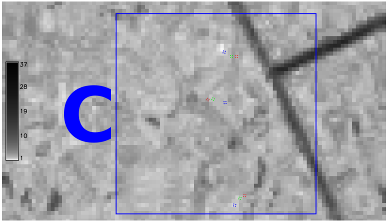
		Future (2090–2099)							
	Sites	Barrow	Council	Atqasuk	Ivotuk	Toolik Lake	Kougarok	Prudhoe Bay	Fairbanks
Present (2000–2009)	Barrow	3.31	9.67	4.63	6.05	5.75	9.02	3.69	11.67
	Council	8.38	1.65	8.10	5.91	6.87	3.10	7.45	5.38
	Atqasuk	6.01	9.33	2.42	5.46	5.26	8.97	2.63	10.13
	Ivotuk	7.06	7.17	5.83	1.53	2.05	7.25	4.87	7.40
	Toolik Lake	7.19	7.67	6.07	2.48	1.25	7.70	5.23	8.16
	Kougarok	7.29	3.05	6.92	5.57	6.31	2.51	6.54	5.75
	Prudhoe Bay	5.29	8.80	3.07	4.75	4.69	8.48	1.94	9.81
	Fairbanks	12.02	5.49	10.36	7.83	8.74	6.24	10.10	1.96

Representativeness: A Quantitative Approach for Scaling

- MSTC provides a quantitative framework for stratifying sampling domains, informing site selection, and determining representativeness of measurements.
- Representativeness analysis provides a systematic approach for up-scaling point measurements to larger domains.
- Methodology is independent of resolution, thus can be applied from site/plot scale to landscape/climate scale.
- It can be extended to include finer spatiotemporal scales, more geophysical characteristics, and remote sensing data.
- Methodology described in Open Access paper:
  Hoffman, F. M., J. Kumar, R. T. Mills, and W. W. Hargrove (2013), "Representativeness-Based Sampling Network Design for the State of Alaska." *Landscape Ecol.*, 28(8):1567–1586. doi:10.1007/s10980-013-9902-0.
- Resulting maps and data available from (the first NGEE Arctic Data DOI): doi:10.5440/1108686.

Barrow Environmental Observatory (BEO)

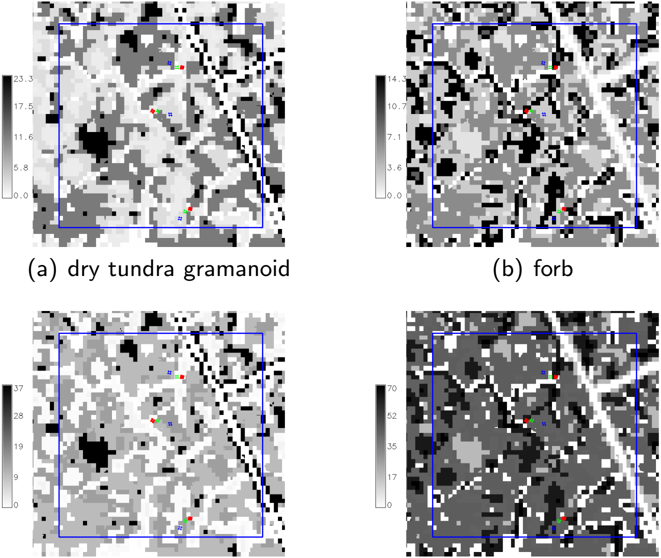




(Kumar et al., in prep)

Representativeness map for vegetation sampling points for A, B, C, and D sampling area (left) and zoomed in on the C samping area (right) developed from WorldView2 satellite images for the year 2010 and LiDAR data.

Vegetation sampling locations represent polygon troughs (red), edges (green), and centers (blue).



(a) dry tundra gramanoid

(b) forb

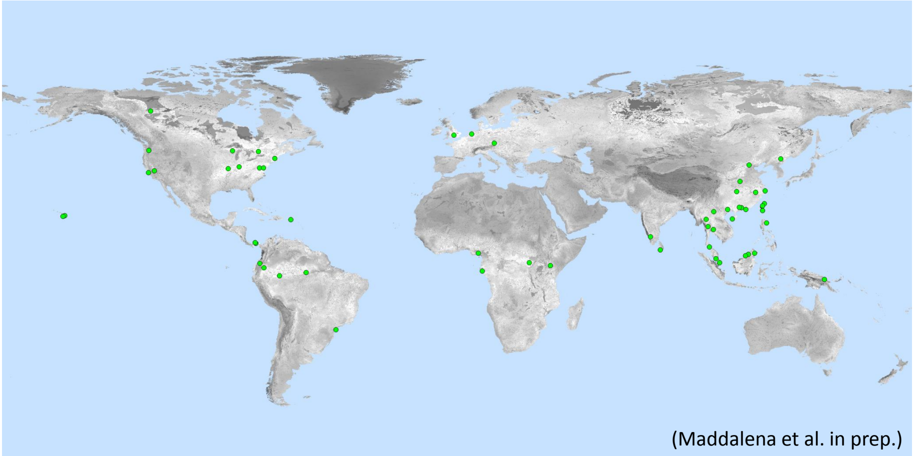
(c) lichen

(d) moss
(Kumar et al., in prep)

Example plant functional type (PFT) distributions scaled up from vegetation sampling locations.

An Approach for Selecting Distributed Sampling Sites for NGEE Tropics

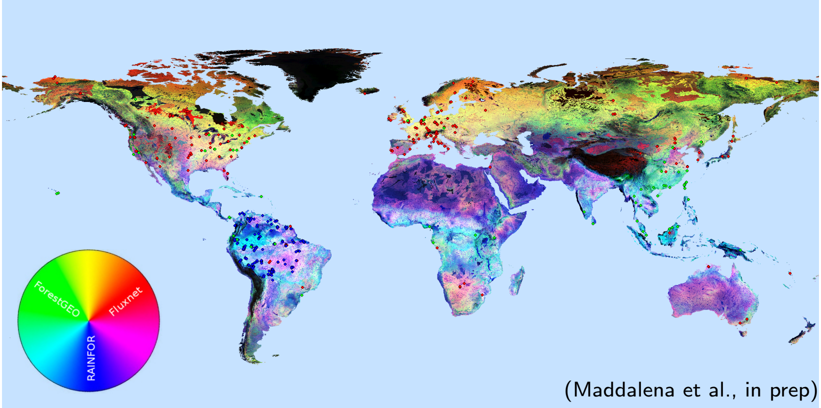
ForestGEO Network Global Representativeness



Light-colored regions are well represented and dark-colored regions are poorly represented by the ForestGEO sampling network.

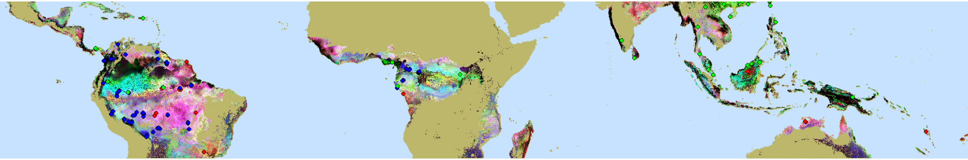
Input layers include 17 global bioclimatic, topographic, and edaphic variables (e.g., biotemperature, seasonal precipitation, slope/aspect, soil C and N).

Triple-Network Global Representativeness

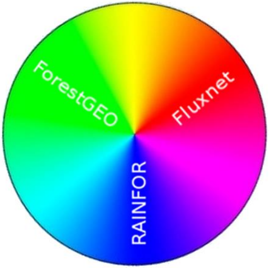


Map indicates which sampling network offers the most representative coverage at any location. Every location is made up of a combination of three primary colors from Fluxnet (red), ForestGEO (green), and RAINFOR (blue).

3-Network Tropical Forest Representativeness



- Individual networks can be combined to determine how well they represent pan-tropical forests
- Here, every location is a combination of three colors, one for each network
- Analysis enables targeting representative sites (or underrepresented sites) as required by science objectives
- Analysis also offers a data-based spatial and temporal scaling framework



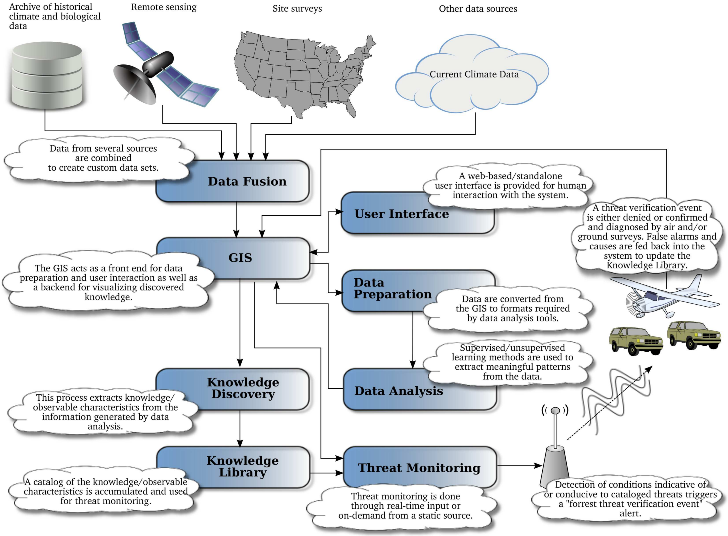
Light colors = well represented
Dark colors = poorly represented

The USDA Forest Service, NASA Stennis Space Center, DOE Oak Ridge National Laboratory, and DOI Eros Data Center have created a system to monitor threats to U.S. forests and wildlands:

- Tier 1: Strategic — The *ForWarn* system 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 was in place and managed by the USDA Forest Service, but Tier 1 was needed to optimally direct its labor-intensive efforts and discover new threats sooner.

Design Plan for the *ForWarn* Early Warning 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.

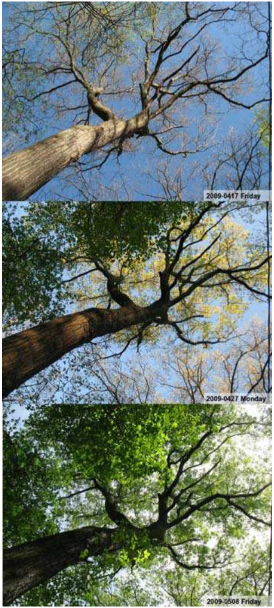
$$\text{NDVI} = \frac{(\sigma_{\text{nir}} - \sigma_{\text{red}})}{(\sigma_{\text{nir}} + \sigma_{\text{red}})} \tag{1}$$

- 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

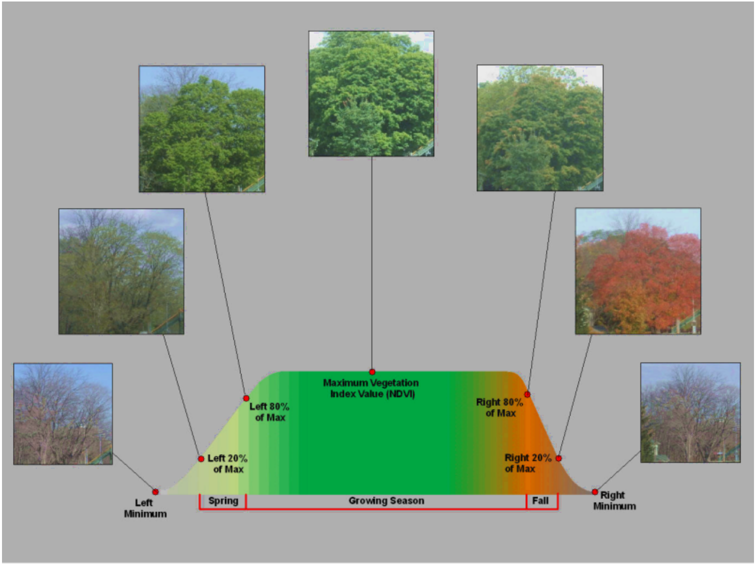
- The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard the Terra (EOS AM, N$\rightarrow$S) and Aqua (EOS PM, S$\rightarrow$N) satellites.
- Both view the entire surface of Earth every 1 to 2 days, acquiring data in 36 spectral bands.
- The MOD 13 product provides Gridded Vegetation Indices (NDVI and EVI) to characterize vegetated surfaces.
- Available are 6 products 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 8-day Terra+Aqua 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.
- *ForWarn* 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.

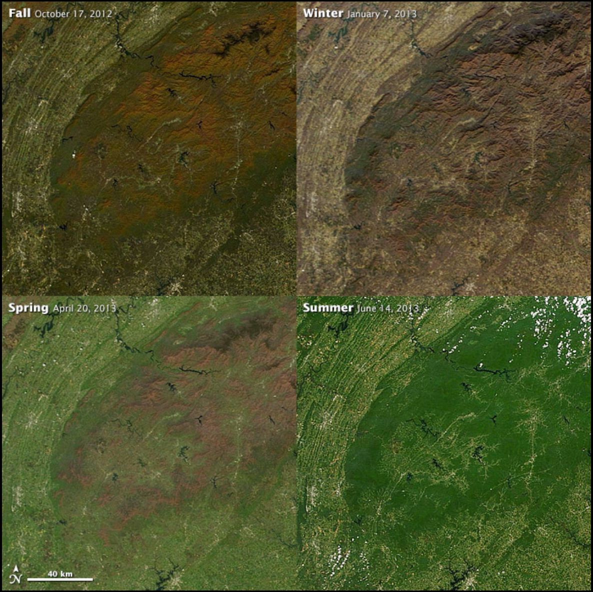


Up-looking photos of a scarlet oak showing the timing of leaf emergence in the spring (Hargrove et al., 2009).

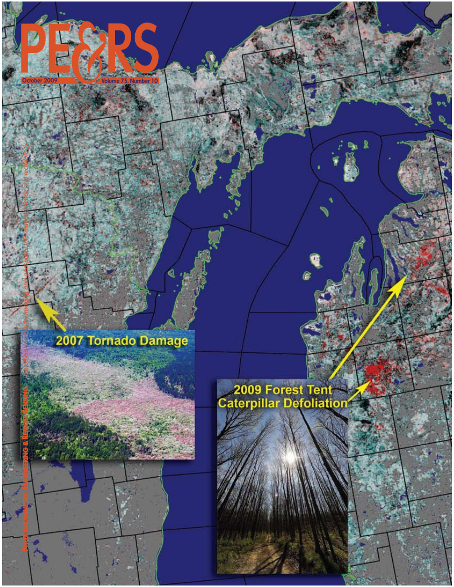
Annual Greenness Profile Through Time



MODIS Snapshots by Season – Walker Branch

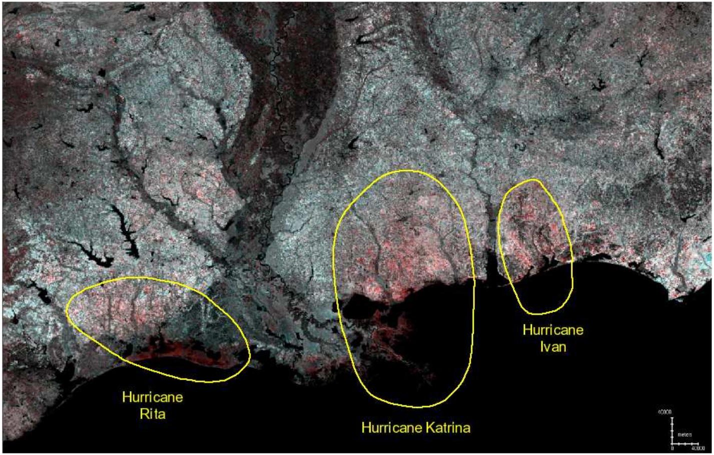


- To detect vegetation disturbances, the current NDVI measurement is compared with the normal, expected baseline for the same location.
- Substantial decreases from the baseline represent potential disturbances.
- Any increases over the baseline may represent vegetation recovery.
- Maximum, mean, or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



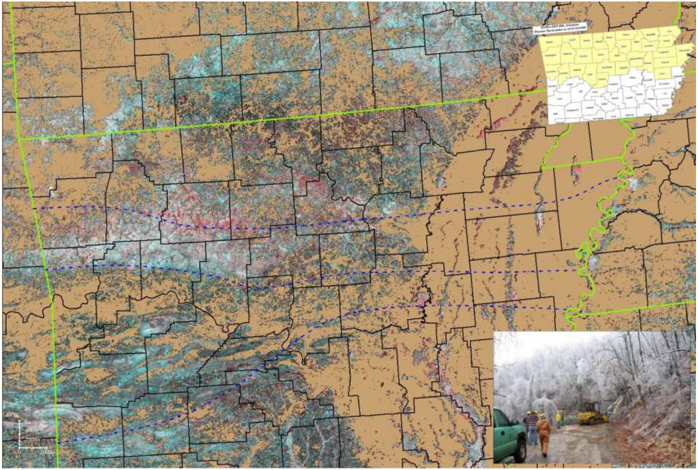
June 10–23, 2009, NDVI is loaded into blue and green; maximum NDVI from 2001–2006 is loaded into red (Hargrove et al., 2009).

Three Hurricanes

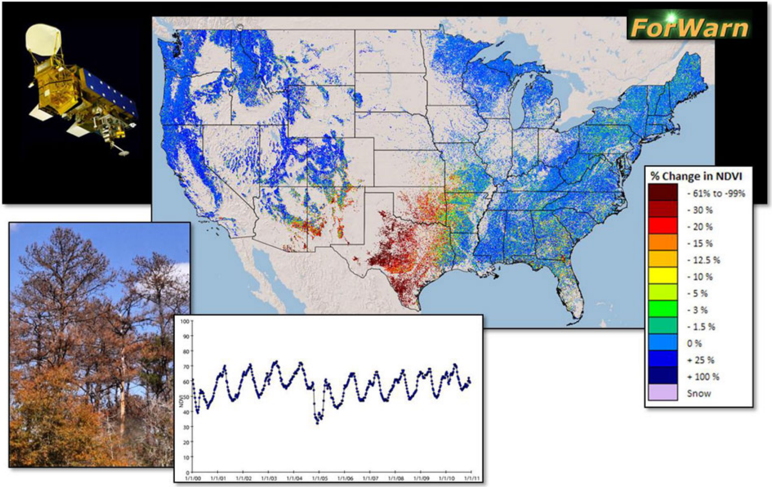


Computed by assigning 2006 20% left value to green & blue, and 20% left from 2004 to red (Hargrove et al., 2009). Red depicts areas of reduced greenness, primarily east of storm tracks and in marshes.

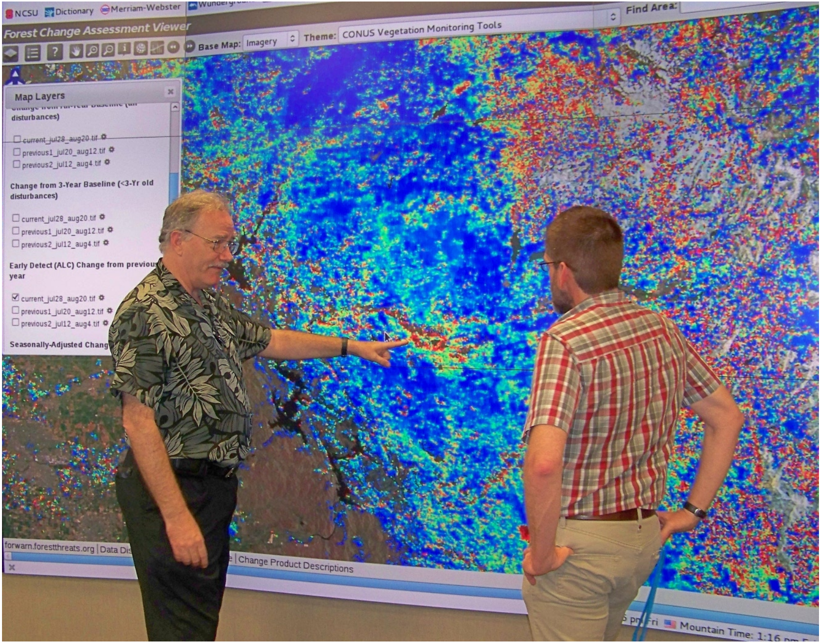
Arkansas Ozarks Ice Storm, Jan. 26–29, 2009



Computed by assigning 2009 max NDVI for June 10–July 15 into blue & green, and 2001–2006 max NDVI for June 10–July 27 into red. Storm resulted in 35,000 without power and 18 fatalities.



ForWarn is a forest change recognition and tracking system that uses high-frequency, moderate resolution satellite data to provide near real-time forest change maps for the continental United States that are updated every eight days. Maps and data products are available in the **Forest Change Assessment Viewer** at `http://forwarn.forestthreats.org/fcav/`



ForWarn researchers get EVEREST-sized look at woodland disturbances

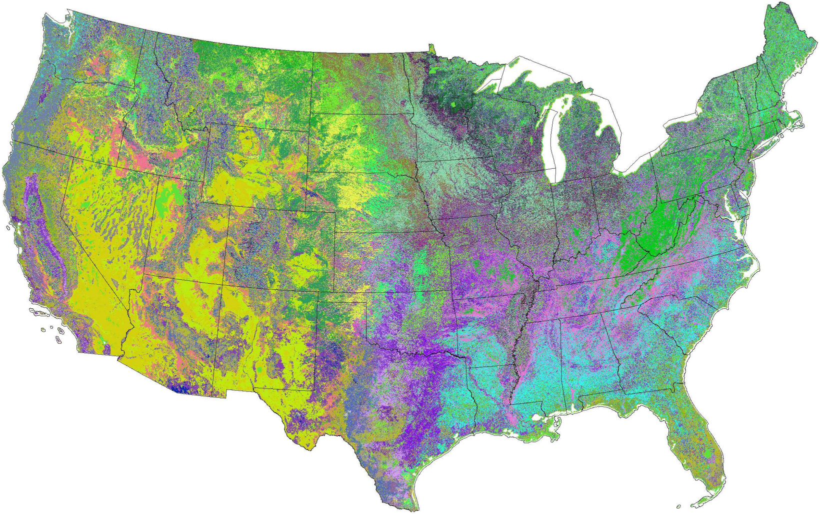
ForWarn Awards

- 2012 Director's Science Delivery Award (September 2012)
  Dr. Robert Doudrick, Station Director of the USDA Forest Service Southern Research Station
- 2013 Interagency Partnership Award (December 2012)
  National Federal Laboratory Consortium (FLC) for Technology Transfer (plus congratulatory letters from Secretary of Energy Ernie Moniz and Secretary of Agriculture Thomas Vilsack)
- 2012 Most Distinguished Scientific or Technical Contribution Award (December 2012)
  ORNL Computer Science & Mathematics Division (CSMD)
- 2012 Partnership Award (March 2013)
  Southeast Regional Federal Laboratory Consortium (FLC) for Technology Transfer
- NASA Group Achievement Award (August 2013)
  Charles Bolden, NASA Administrator
- 2013 Southern Research Station Director's Award for Partnerships (December 2013)
  Dr. Robert Doudrick, Station Director of the USDA Forest Service Southern Research Station
- *Pending:* 2013 Chief's Honor Award (March 17, 2014 in Washington, DC)
  Thomas L. Tidwell, Chief, USDA Forest Service

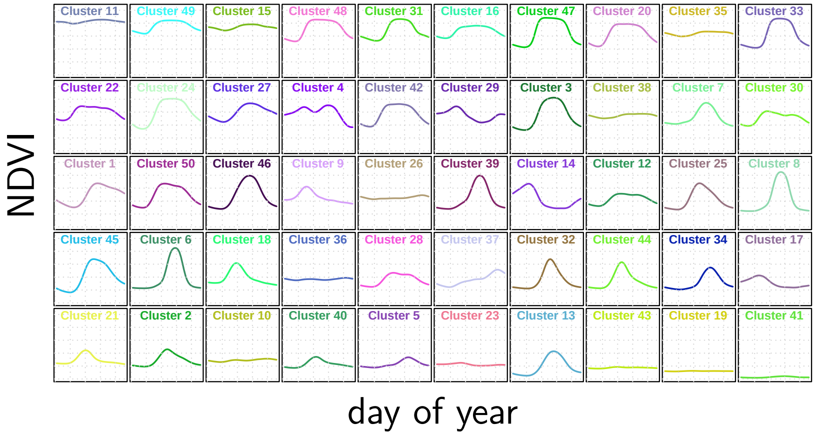
Clustering MODIS NDVI into Phenoregions

- 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.
- $>$87B NDVI values, consisting of $\sim$146.4M cells for the CONUS at 250 m resolution with 46 maps per year for 13 years (2000–2012), analyzed using k-means clustering.
- The annual traces of NDVI for every year and map cell are combined into one 327 GB single-precision binary data set of 46-dimensional observation vectors.
- Clustering yields 13 phenoregion maps in which each cell is classified into one of k phenoclasses that represent prototype annual NDVI traces.

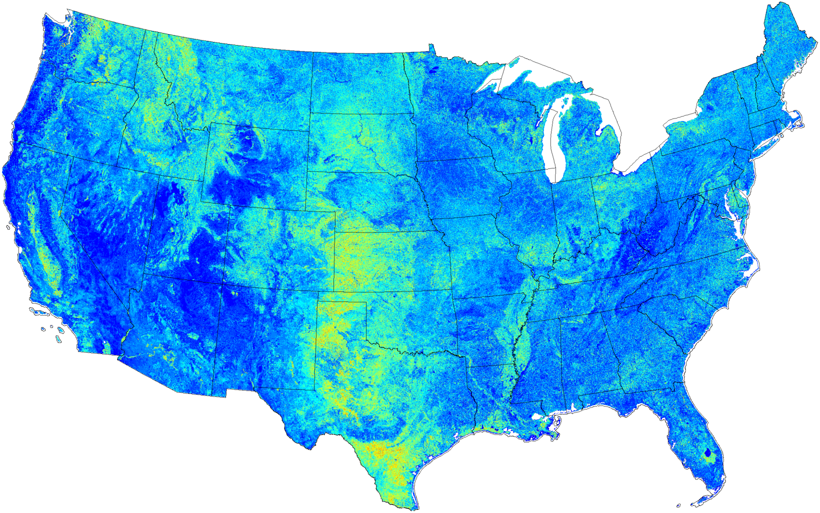
50 Phenoregions for year 2012 (Random Colors)



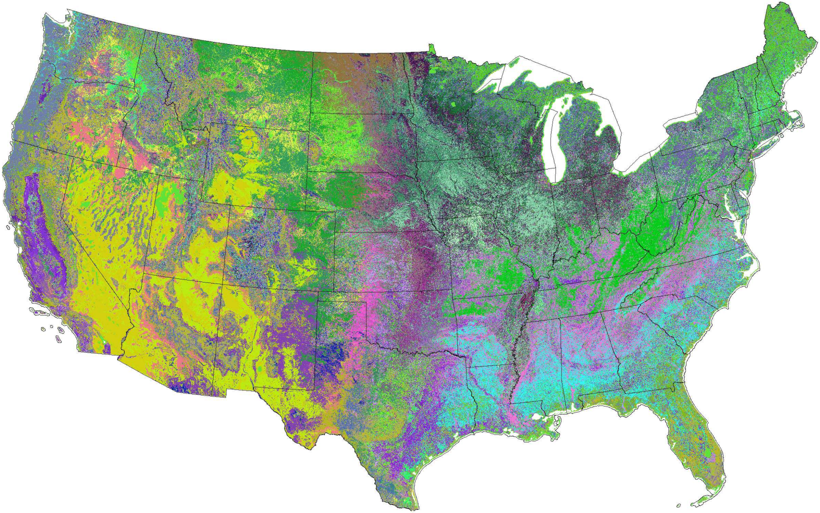
50 Phenoregion Prototypes (Random Colors)



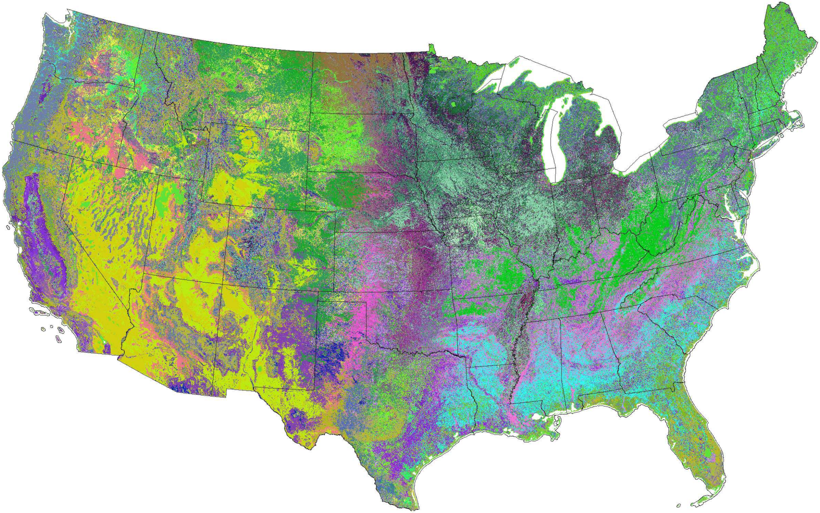
50 Phenoregions Persistence



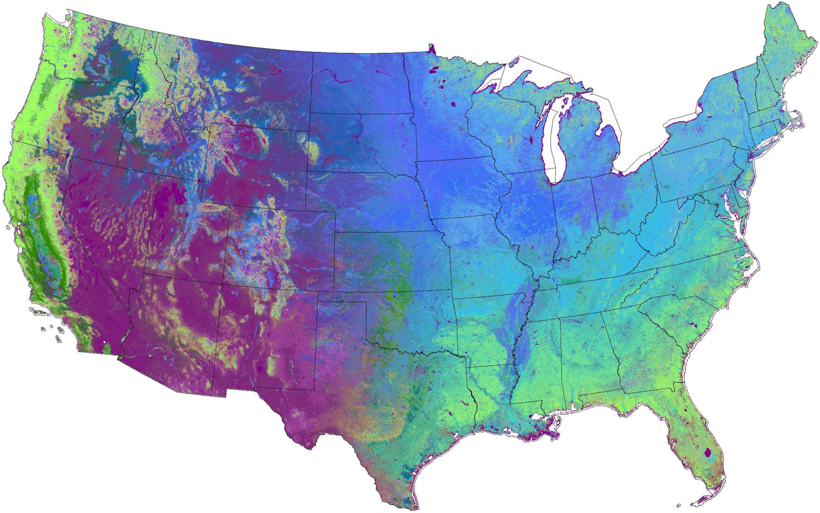
50 Phenoregions Mode (Random Colors)



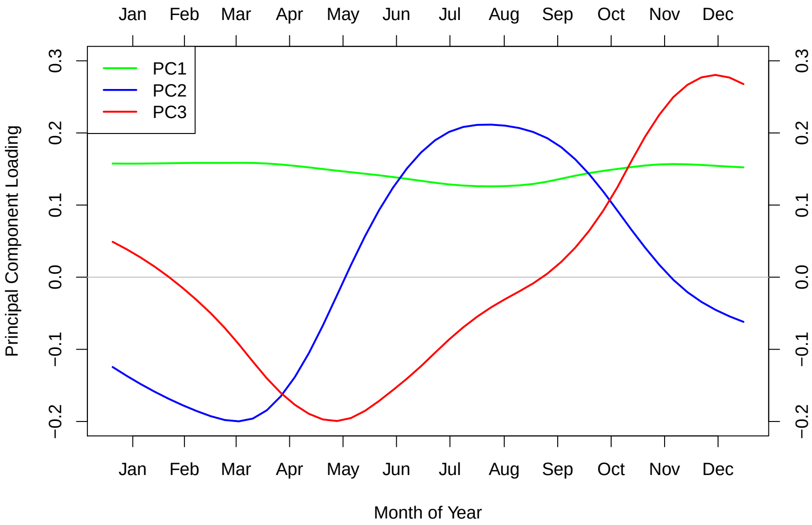
50 Phenoregions Max Mode (Random Colors)



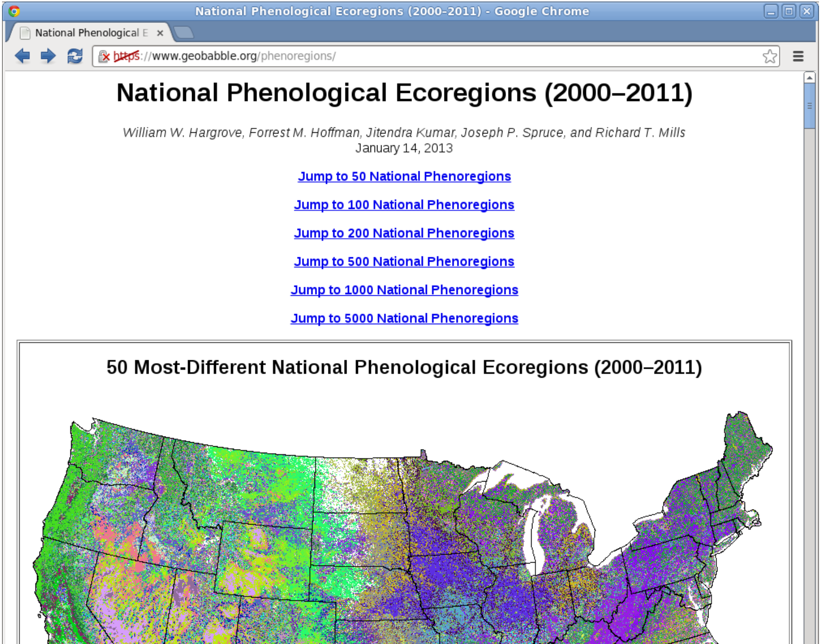
50 Phenoregions Max Mode (Similarity Colors)



50 Phenoregions Max Mode (Similarity Colors Legend)

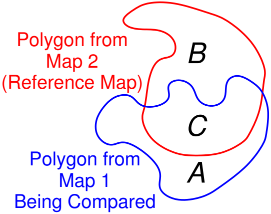


Phenoregions Clearinghouse



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.

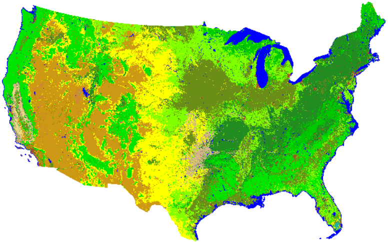
Two 2-Way Comparisons with Land Cover Maps

Cluster	IGBP Land Cover	Olson's Global Ecoregions
1	Evergreen Needleleaf Forest	cool conifer forest
2	Grasslands	cool grasses and shrubs
3	Cropland/Natural Vegetation Mosaic	cool forest and field
4	Croplands	cool forest and field
5	Grasslands	cool grasses and shrubs
6	Croplands	corn and beans cropland
7	Cropland/Natural Vegetation Mosaic	cool forest and field
8	Croplands	corn and beans cropland
9	Grasslands	hot and mild grasses and shrubs
10	Grasslands	cool grasses and shrubs
11	Evergreen Needleleaf Forest	cool conifer forest
12	Grasslands	hot and mild grasses and shrubs
13	Water	inland water
14	Savannas	savanna (woods)
15	Evergreen Needleleaf Forest	cool conifer forest
16	Evergreen Needleleaf Forest	conifer forest
17	Open Shrublands	semi desert sage
18	Grasslands	cool grasses and shrubs
19	Open Shrublands	semi desert shrubs
20	Deciduous Broadleaf Forest	deciduous broadleaf forest
21	Grasslands	cool grasses and shrubs
22	Croplands	broadleaf crops
23	Open Shrublands	semi desert sage
24	Deciduous Broadleaf Forest	cool broadleaf forest
25	Cropland/Natural Vegetation Mosaic	crops, grass, shrubs

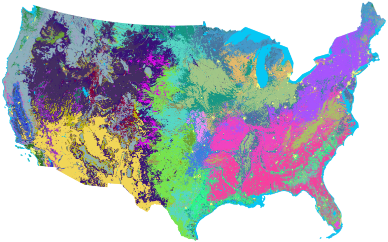
Two 2-Way Comparisons with Land Cover Maps

Cluster	IGBP Land Cover	Olson's Global Ecoregions
26	Evergreen Needleleaf Forest	cool conifer forest
27	Evergreen Needleleaf Forest	cool conifer forest
28	Grasslands	hot and mild grasses and shrubs
29	Woody Savannas	woody savanna
30	Grasslands	hot and mild grasses and shrubs
31	Deciduous Broadleaf Forest	cool broadleaf forest
32	Croplands	cool crops and towns
33	Deciduous Broadleaf Forest	cool broadleaf forest
34	Grasslands	hot and mild grasses and shrubs
35	Evergreen Needleleaf Forest	cool conifer forest
36	Grasslands	dry woody scrub
37	Grasslands	hot and mild grasses and shrubs
38	Evergreen Needleleaf Forest	cool conifer forest
39	Croplands	corn and beans cropland
40	Open Shrublands	semi desert sage
41	Water	inland water
42	Deciduous Broadleaf Forest	deciduous broadleaf forest
43	Open Shrublands	semi desert shrubs
44	Grasslands	cool grasses and shrubs
45	Evergreen Needleleaf Forest	cool conifer forest
46	Croplands	corn and beans cropland
47	Deciduous Broadleaf Forest	cool broadleaf forest
48	Evergreen Needleleaf Forest	conifer forest
49	Evergreen Needleleaf Forest	conifer forest
50	Croplands	woody savanna

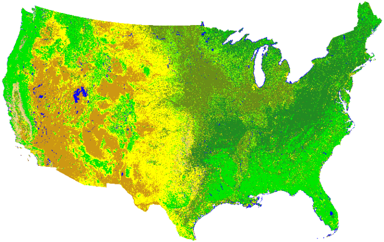
Phenoregions Reclassed Using Land Cover Types



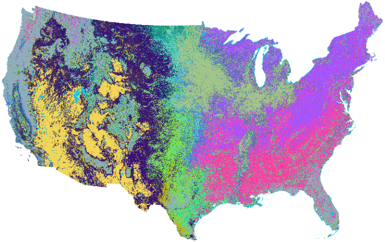
(a) IGBP Land Cover



(c) Olson's Global Ecoregions

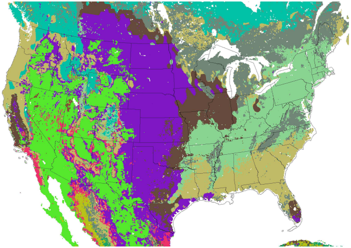


(b) 50 Phenoregions Reclassed

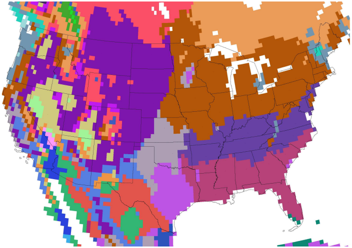


(d) 50 Phenoregions Reclassed

Expert-Derived Land Cover/Vegetation Type Maps



Foley Land Cover



Holdridge Life Zones

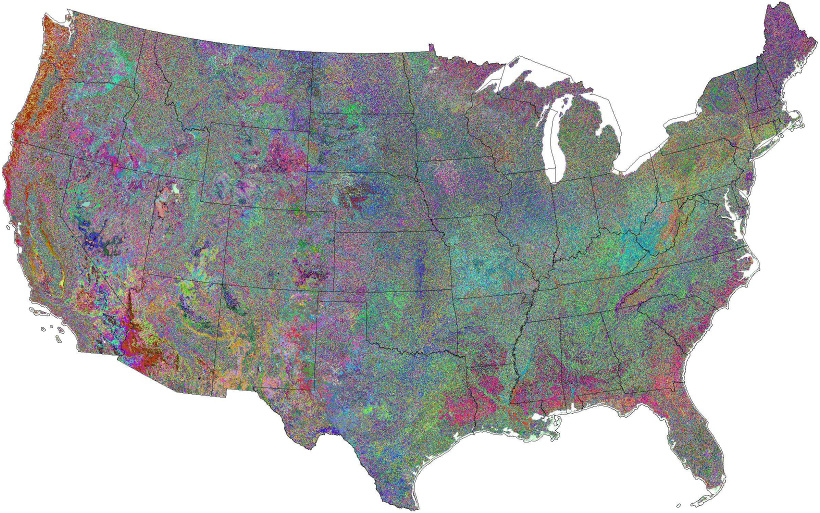
	Expert Map	# Cats
1.	DeFries UMd Vegetation	12
2.	Foley Land Cover	14
3.	Fedorova, Volkova, and Varlyguin World Vegetation Cover	31
4.	GAP National Land Cover	578
5.	Holdridge Life Zones	25
6.	Küchler Types	117
7.	BATS Land Cover	17
8.	IGBP Land Cover	16
9.	Olson Global Ecoregions	49
10.	Seasonal Land 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 **Eastern Deciduous Forest Biome**.
- **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 ecoregionalizations drawn by human experts.
- We consider an entire library of ecoregion and land cover maps, and choose the label with the highest GOF score for every phenoregion polygon.

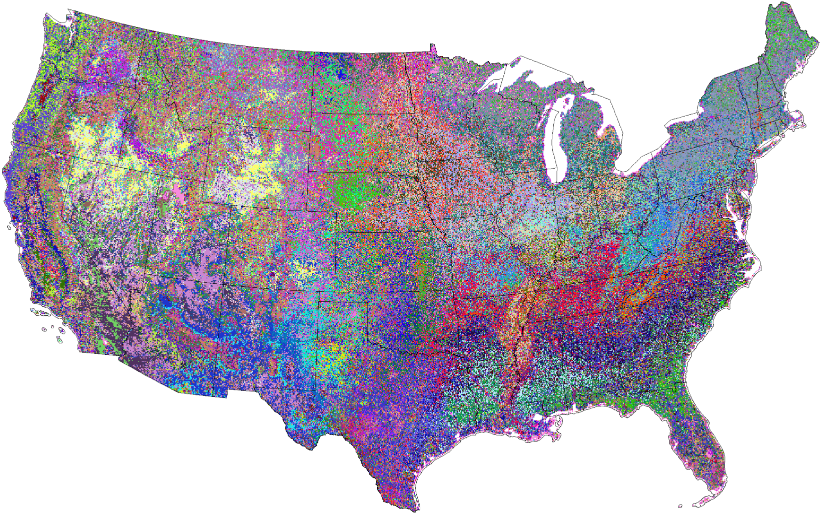
Patchwork Crazy Quilt of Multiple Land Cover Types

1000 Phenoregions Max Under (Random Colors)

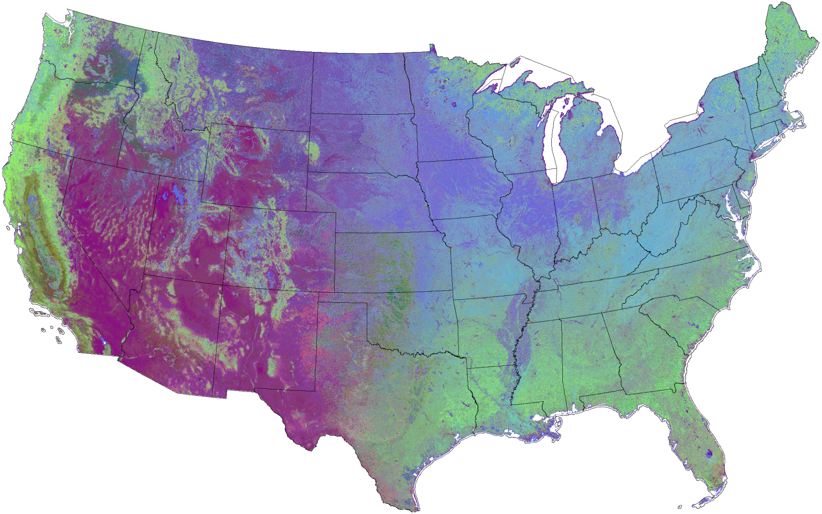


Category	Land Cover Label	Land Cover Map
1	Acadian Low-Elevation Spruce-Fir-Hardwood Forest	landfire vegetation type
2	Agriculture-Pasture and Hay	landfire vegetation type
3	Alpine meadows & barren	ktlamb
4	Barren	landcover.slcr
5	Barren or Sparsely Vegetated	landcover.usgs
6	Bluestem/Grama	ktlamb
7	Bluestem Hills, MLRA 76	mlra
8	Boreal Evergreen Forest/Woodland	foleylandcover
9	Boreal	fvvcode
10	Boreal moist forest	holdridgezonesnormal
11	Broadleaf Deciduous Forest	landcover.usgs
12	Brown Glaciated Plain, MLRA 52	mlra
13	California Central Valley and Southern Coastal Grassland	GAP 240m laea
14	California Central Valley Mixed Oak Savanna	GAP 240m laea
15	California oakwoods	ktlamb
16	California steppe	ktlamb
.	.	.
.	.	.
.	.	.
222	Warm temperate moist forest	holdridgezonesnormal
223	Warm Temperate Moist Forest	leemansholdridgezones
224	[water]	ktlamb
225	Water	landcover.slcr
226	Western Great Plains Mesquite Woodland and Shrubland	GAP 240m laea
227	Western Great Plains Shortgrass Prairie	landfire vegetation type
228	Western ponderosa	ktlamb
229	Western Rio Grande Plain, MLRA 83B	mlra
230	Western spruce/Fir	ktlamb
231	Wheatgrass/Bluegrass	ktlamb
232	Wheatgrass/Needlegrass	ktlamb
233	Willamette and Puget Sound Valleys, MLRA 2	mlra
234	Woodland/Cropland Mosaic	landcover.usgs
235	Woody wetlands	NLCD2006 240m laea

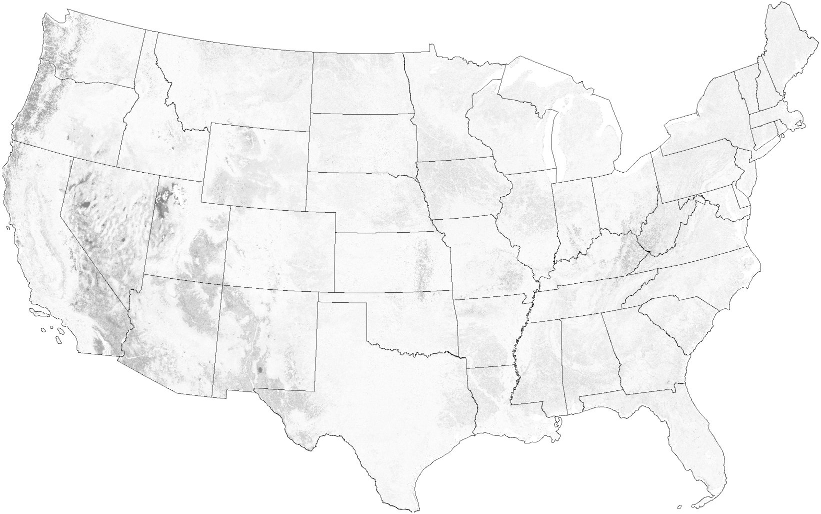
1000 Phenoregions Reclassed into 235 Land Cover Types



1000 Phenoregions Reclassed into 235 Land Cover Types



1000 Phenoregions Reclassed Goodness of Fit

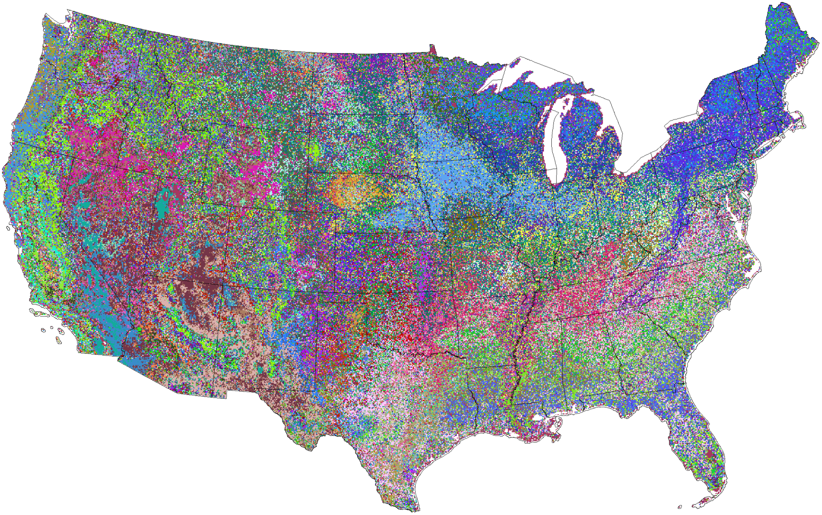


Composition of the 235 Land Cover Types Map

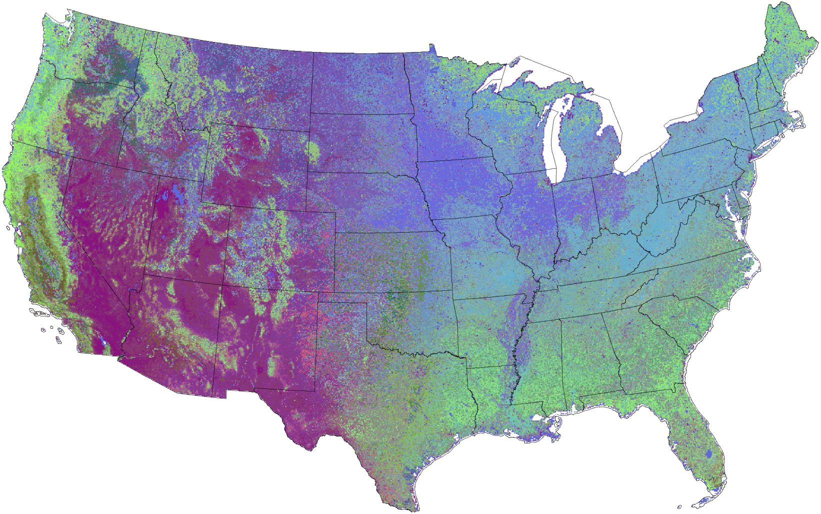
	Map	Cats	WCats	WClusts	%Area
10.	Seasonal Land Cover Regions	194	43	160	19.45
9.	Olson Global Ecoregions	49	12	96	12.36
3.	Fedorova, Volkova, and Varlyguin World Vegetation Cover	31	4	93	10.69
17.	Landfire Vegetation Types	443	27	85	9.09
6.	Küchler Types	117	34	81	7.87
14.	Major Land Resource Areas	197	42	107	7.18
12.	Leemans-Holdridge Life Zones	26	8	54	5.27
11.	USGS Land Cover	24	7	21	4.85
4.	GAP National Land Cover	578	19	124	4.48
5.	Holdridge Life Zones	25	9	38	4.15
2.	Foley Land Cover	14	7	48	3.86
15.	National Land Cover Database 2006	16	8	47	3.24
13.	Matthews Vegetation Types	19	5	18	2.49
16.	Wilson, Henderson, & Sellers Primary Vegetation Types	23	2	9	1.46
7.	BATS Land Cover	17	4	10	1.23
8.	IGBP Land Cover	16	3	4	0.80
1.	DeFries UMd Vegetation	12	2	5	0.25
	TOTAL		**235**	**1000**	**100%**

#	Category	Land Cover Label	Land Cover Map	Percent Area
1	176	Subboreal	fvvcode	5.28%
2	179	Subtropical	fvvcode	4.25%
3	73	Evergreen Coniferous Forest	landcover.usgs	3.87%
4	67	Open Shrubland	foleylandcover	3.74%
5	35	corn and beans cropland	landcover.oge	3.48%
6	29	cool conifer forest	landcover.oge	2.93%
7	32	Cool temperate moist forest	holdridgezonesnormal	2.55%
8	64	Desert Shrubland/Grassland (Creosote, Saltbush, Mesquite, Sand Sage)	landcover.slcr	2.27%
9	55	Deciduous Forest (Oak, Hickory, Sweet Gum, Southern Pines) with Cropland and Pasture	landcover.slcr	2.25%
10	28	cool broadleaf forest	landcover.oge	2.23%
11	66	Sparsely Vegetated Desert Shrublands	landcover.slcr	2.14%
12	188	Warm temperate moist forest	holdridgezonesnormal	2.06%
13	180	Subtropical moist forest	holdridgezonesnormal	2.05%
14	160	semi desert sage	landcover.oge	1.87%
.	.	.	.	
.	.	.	.	
.	.	.	.	
187	120	Northern hardwoods/Spruce	ktlamb	0.01%
188	102	Laurentian-Acadian Alkaline Conifer-Hardwood Swamp	landfire vegetation type	0.01%
189	51	NASS-Vineyard	landfire vegetation type	0.01%
190	2	Alpine meadows & barren	ktlamb	0.01%
191	143	Pseudotsuga menziesii Forest Alliance	landfire vegetation type	0.01%
192	134	Olympic and Cascade Mountains, MLRA 3	mlra	0.01%
193	79	Evergreen Needleleaf Forest (Lodgepole Pine and Douglas Fir)	landcover.slcr	0.01%
194	125	North Pacific Maritime Mesic Subalpine Parkland	GAP 240m laea	0.00%
195	80	Evergreen Needleleaf Forest (Lodgepole Pine, Englemann Spruce, Ponderosa Pine)	landcover.slcr	0.00%
196	157	Saltbrush/Greasewood	ktlamb	0.00%
197	106	Mediterranean California Red Fir Forest	GAP 240m laea	0.00%

1000 Phenoregions Reclassed into 197 Land Cover Types



1000 Phenoregions Reclassed into 197 Land Cover Types



Uses for Label Stealing

- Borrowing ecoregion, land cover, or vegetation type labels for unsupervised classifications.
- Automated attribution of disturbance agents through comparison of a *ForWarn* disturbance map with ADS aerial sketchmaps, wildfire perimeters, tornado track maps, and fuel treatment maps through time.
- Determination of the most important driving variable for phenoregions maps through comparison with separate maps of slope, aspect, solar input, elevation, soil types, etc.
- Automated recognition of species composition of forest vegetation through comparison of a phenoregions map with individual tree species range maps.

AGU Fall Meeting Session Advertisement

Big Data in the Geosciences: New Analytics Methods and Parallel Algorithms

Co-conveners: Jitendra Kumar (ORNL), Robert L. Jacob (ANL), Forrest M. Hoffman (ORNL), and Miguel D. Mahecha (MPI-Jena)

Earth and space science data are increasingly large and complex–often representing high spatial/temporal/spectral resolution and dimensions from remote sensing or model results–making such data difficult to analyze, visualize, interpret, and understand by traditional methods. This session focuses on application and development of new geoscientific data analytics approaches (statistical, data mining, assimilation, machine learning, etc.) and parallel algorithms and software employing high performance computing resources for scalable analysis and novel applications of traditional methods on large geoscience data sets. Analysis methods that operate in-situ with parallel simulations to reduce output data volumes are also of interest. Abstracts focused on analysis, synthesis and knowledge extraction from large and complex Earth science data from all disciplines are invited.

Abstract submissions are due 6 August 2014, 23:59 EDT/03:59 +1 GMT

Acknowledgments

The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the U.S. Department of Energy (DOE) Office of Science. This research was sponsored by the Office of Biological and Environmental Research in the U.S. Department of Energy Office of Science and the U.S. Department of Agriculture (USDA) Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC). This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) 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

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.

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

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.

F. M. Hoffman, J. Kumar, R. T. Mills, and W. W. Hargrove. Representativeness-based sampling network design for the State of Alaska. *Landscape Ecol.*, 28(8): 1567–1586, Oct. 2013. doi:10.1007/s10980-013-9902-0.

M. Keller, D. Schimel, W. Hargrove, and F. Hoffman. A continental strategy for the National Ecological Observatory Network. *Front. Ecol. Environ.*, 6(5):282–284, June 2008. doi:10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2. Special Issue on Continental-Scale Ecology.

D. Schimel, W. Hargrove, F. Hoffman, and J. McMahon. NEON: A hierarchically designed national ecological network. *Front. Ecol. Environ.*, 5(2):59, Mar. 2007. doi:10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2.

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.