High Performance Computational Landscape Ecology

Forrest M. Hoffman[1], William W. Hargrove[2], Jitendra Kumar[1], Zachary L. Langford[3,1], and Damian M. Maddalena[4]

[1]Oak Ridge National Laboratory, [2]USDA Forest Service, [3]University of Tennessee–Knoxville, and [4]University of North Carolina–Wilmington

July 6, 2015

International Association for Landscape Ecology (IALE)
9th World Congress

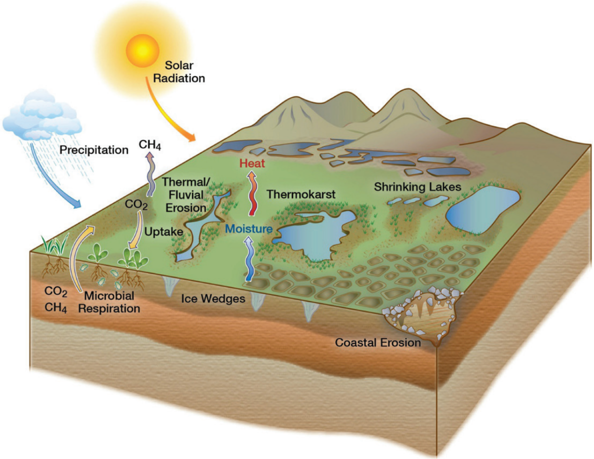
Portland, Oregon, USA

Using Clustering to Define Climate Regimes

Site and Network Representativeness

Next-Generation Ecosystem Experiments (NGEE Arctic)

http://ngee.ornl.gov/



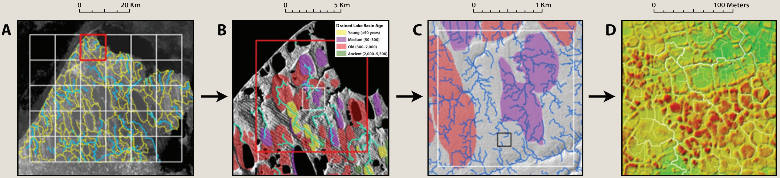
The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

U.S. DEPARTMENT OF ENERGY
OAK RIDGE National Laboratory
Los Alamos NATIONAL LABORATORY
BERKELEY LAB
BROOKHAVEN NATIONAL LABORATORY
UAF UNIVERSITY OF ALASKA FAIRBANKS

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

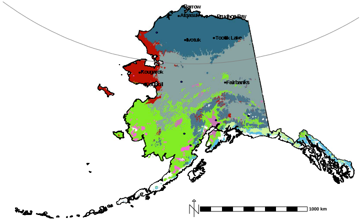
- Resource and logistical constraints limit the frequency and extent of observations, necessitating the development of a systematic sampling strategy that objectively represents environmental variability at the desired spatial scale.
- Required is a methodology that provides a quantitative framework for informing site selection and determining the representativeness of measurements.
- Multivariate spatiotemporal clustering (MSTC) was applied at the landscape scale ($4\ \text{km}^2$) 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.

Data Layers

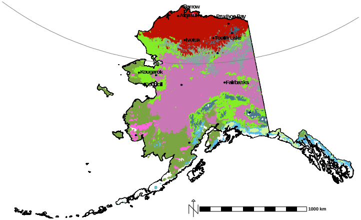
Table: 37 characteristics averaged for the present (2000–2009) and the future (2090–2099).

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, Present and Future



2000–2009



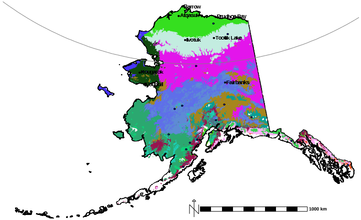
2090–2099

(Hoffman et al., 2013)

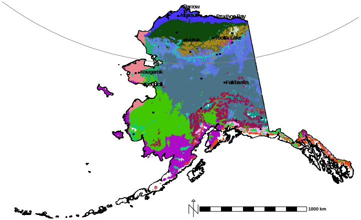
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

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



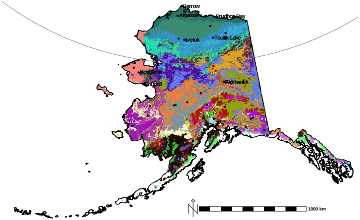
2090–2099

(Hoffman et al., 2013)

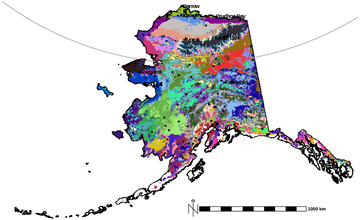
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

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)

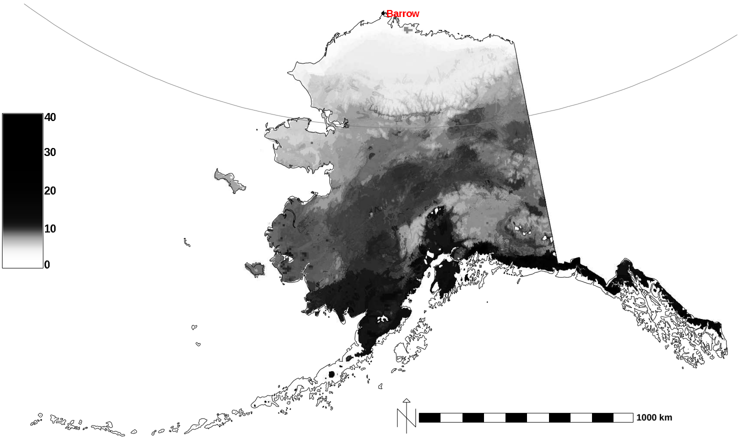
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

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.

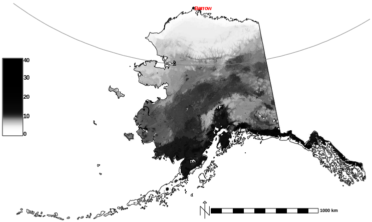
Present Representativeness of Barrow or "Barrow-ness"



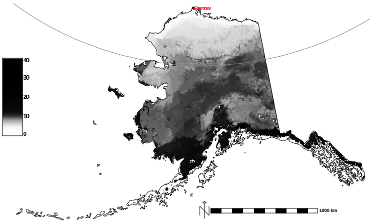
(Hoffman et al., 2013)

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

Present vs. Future Barrow-ness



2000–2009

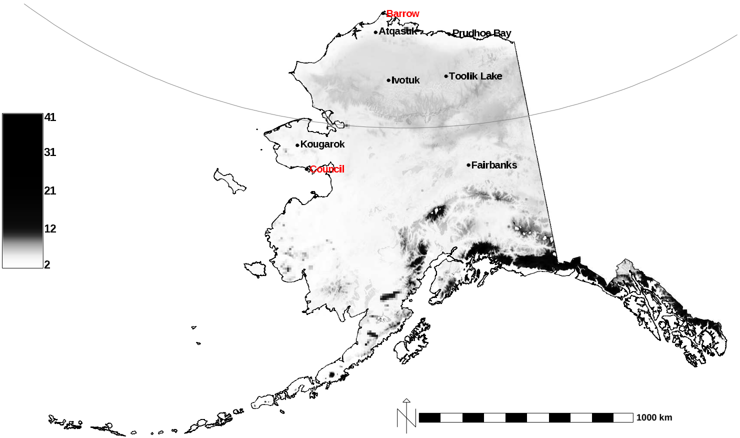


2090–2099

(Hoffman et al., 2013)

As environmental conditions change, due primarily to increasing temperatures, climate gradients shift and the representativeness of Barrow will be reduced in the future.

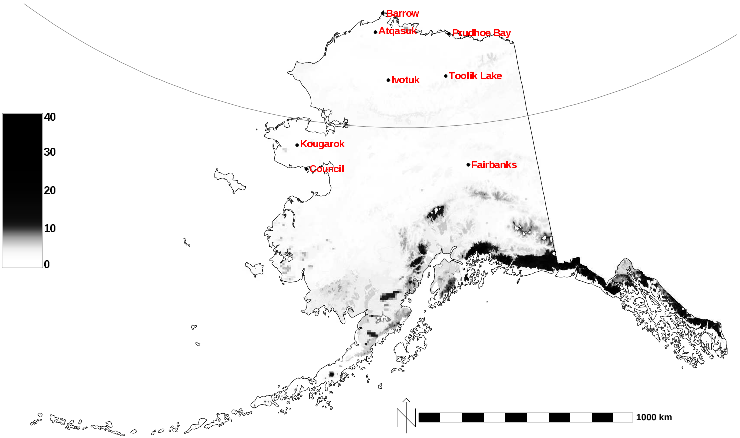
Network Representativeness: Barrow + Council



(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 dissimilarities for the present (2000–2009).

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

(Hoffman et al., 2013)

State Space Dissimilarities: 8 Sites, Present and Future

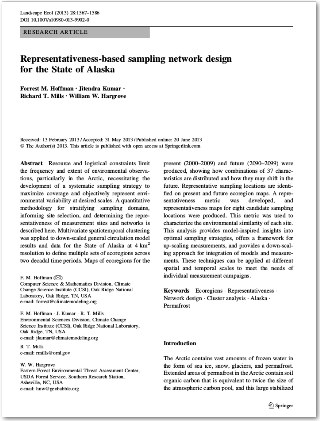
Table: Site state space dissimilarities between the present (2000–2009) and the future (2090–2099).

		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

(Hoffman et al., 2013)

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.



Landscape Ecol (2013) 28:1567–1586
DOI 10.1007/s10980-013-9902-0

RESEARCH ARTICLE

Representativeness-based sampling network design for the State of Alaska

Forrest M. Hoffman · Jitendra Kumar · Richard T. Mills · William W. Hargrove

Received: 13 February 2013 / Accepted: 31 May 2013 / Published online: 20 June 2013

Abstract Resource and logistical constraints limit the frequency and extent of environmental observations, particularly in the Arctic, necessitating the development of a systematic sampling strategy to maximize coverage and objectively represent environmental variability at desired scales. A quantitative methodology for stratifying sampling domains, informing site selection, and determining the representativeness of measurement sites and networks is described here. Multivariate spatiotemporal clustering was applied to down-scaled general circulation model results and data for the State of Alaska at 4 km^2 resolution to define multiple sets of ecoregions across two decadal time periods. Maps of ecoregions for the present (2000–2009) and future (2090–2099) were produced, showing how combinations of 37 characteristics are distributed and how they may shift in the future. Representative sampling locations are identified on present and future ecoregion maps. A representativeness metric was developed, and representativeness maps for eight candidate sampling locations were produced. This metric was used to characterize the environmental similarity of each site. This analysis provides model-inspired insights into optimal sampling strategies, offers a framework for up-scaling measurements, and provides a down-scaling approach for integration of models and measurements. These techniques can be applied at different spatial and temporal scales to meet the needs of individual measurement campaigns.

Keywords Ecoregions · Representativeness · Network design · Cluster analysis · Alaska · Permafrost

F. M. Hoffman (✉)
Computer Science & Mathematics Division, Climate Change Science Institute (CCSI), Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: forrest@climatemodeling.org

F. M. Hoffman · J. Kumar · R. T. Mills
Environmental Sciences Division, Climate Change Science Institute (CCSI), Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: jkumar@climatemodeling.org

R. T. Mills
e-mail: rmills@ornl.gov

W. W. Hargrove
Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, Southern Research Station, Asheville, NC, USA
e-mail: hnw@geobabble.org

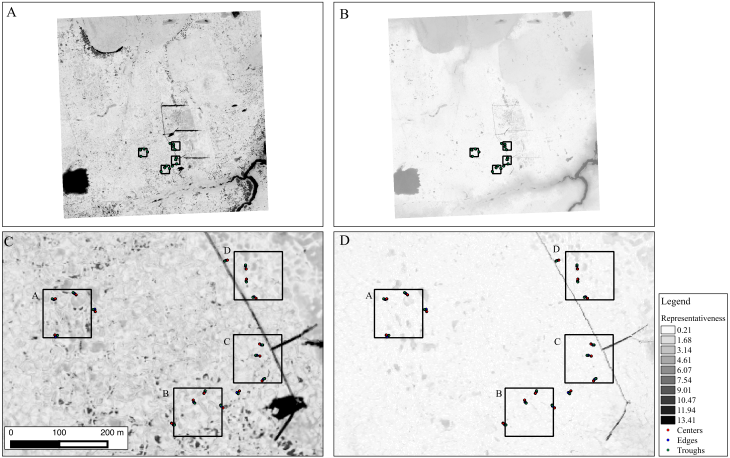
Introduction

The Arctic contains vast amounts of frozen water in the form of sea ice, snow, glaciers, and permafrost. Extended areas of permafrost in the Arctic contain soil organic carbon that is equivalent to twice the size of the atmospheric carbon pool, and this large stabilized

- 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.

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.

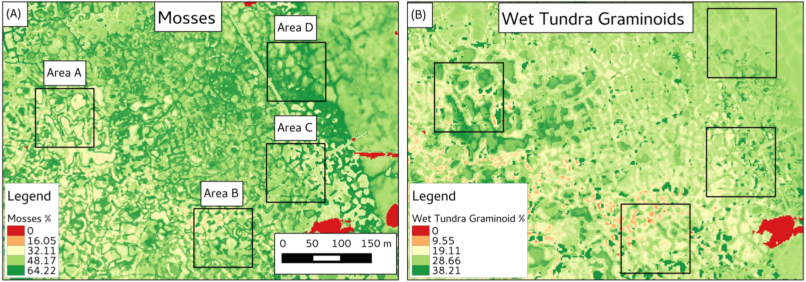
Barrow Environmental Observatory (BEO)



(Langford et al., in prep)

Representativeness map for vegetation sampling points in A, B, C, and D sampling area including phenology (left) and for a single date (right), based on WorldView2 satellite images for the year 2010 and LiDAR data.

Barrow Environmental Observatory (BEO)

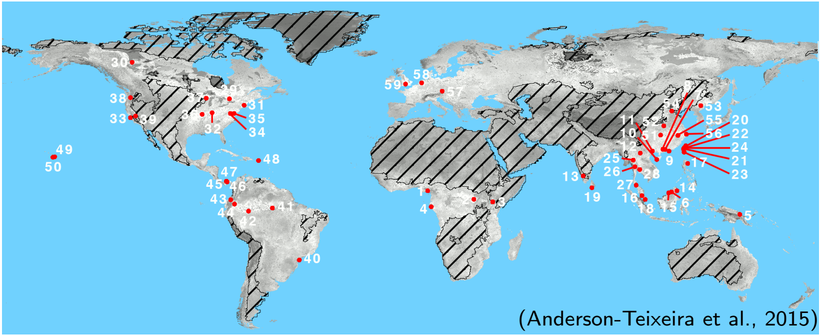


(Langford et al., in prep)

Mosses and wet tundra graminoids PFT % for Areas A, B, C, D.

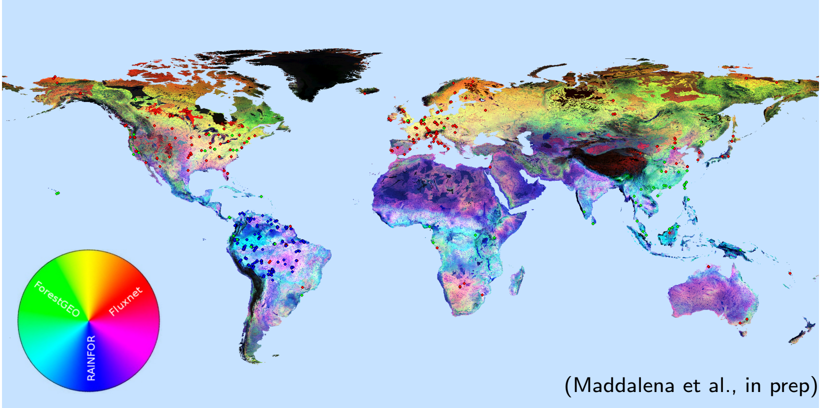
Example plant functional type (PFT) distributions scaled up from vegetation sampling locations.

ForestGEO Network Global Representativeness



Map illustrating ForestGEO network representation of 17 bioclimatic, edaphic, and topographic conditions globally. Light-colored regions are well represented and dark-colored regions are poorly represented by the ForestGEO sampling network. Stippling covers non-forest areas.

Triple-Network Global Representativeness

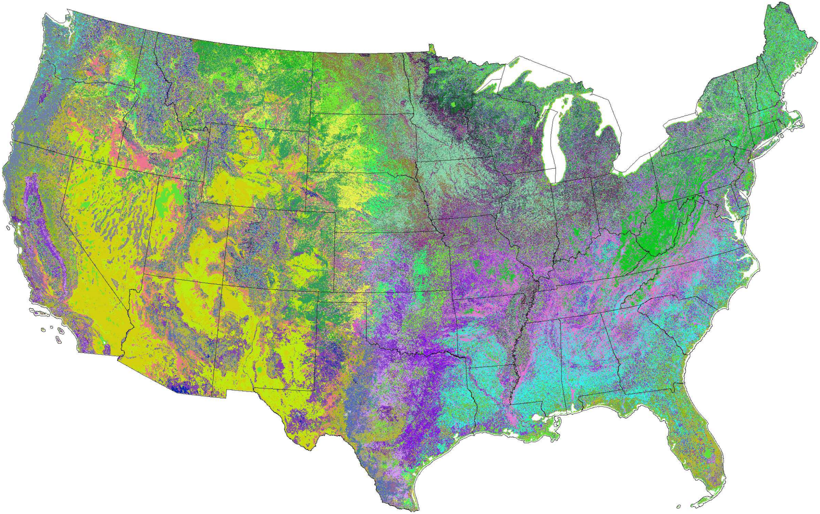


Map indicates which sampling network offers the most representative coverage at any location. Every location is made up of a combination of three primary colors from Fluxnet (red), ForestGEO (green), and RAINFOR (blue).

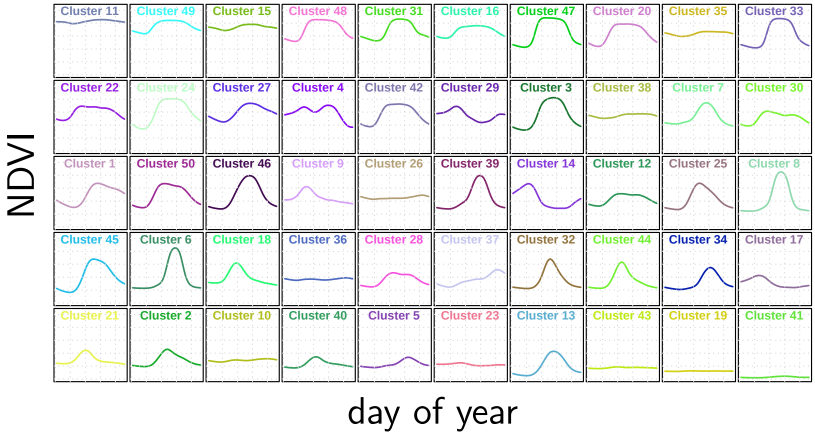
Clustering MODIS NDVI into Phenoregions

- 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.

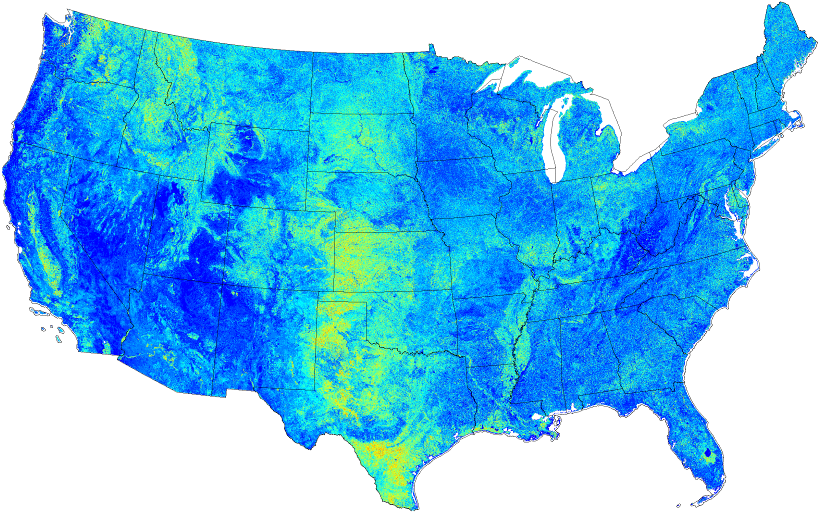
50 Phenoregions for year 2012 (Random Colors)



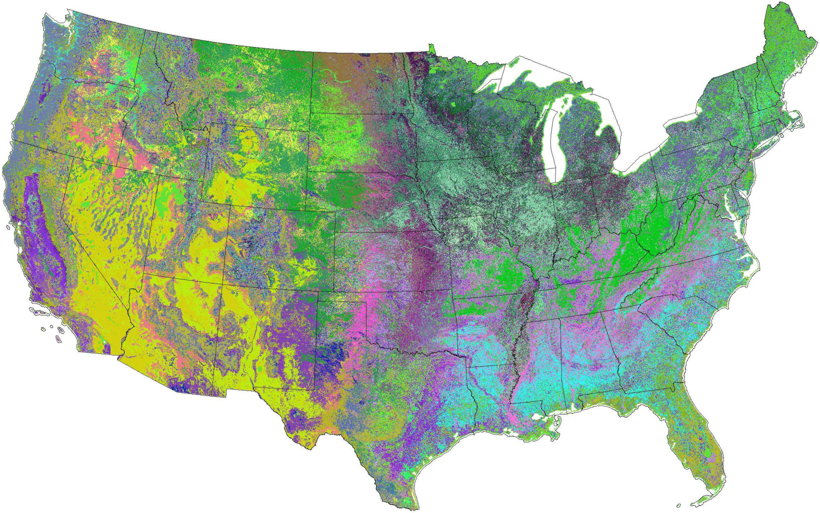
50 Phenoregion Prototypes (Random Colors)



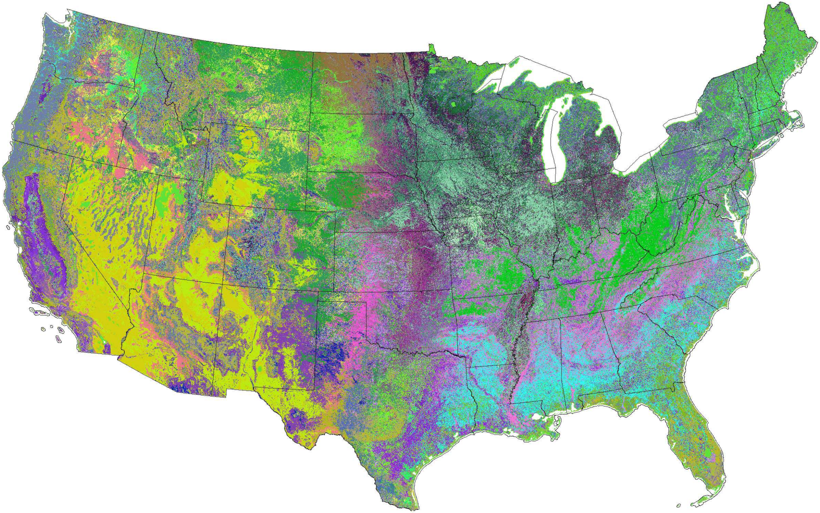
50 Phenoregions Persistence



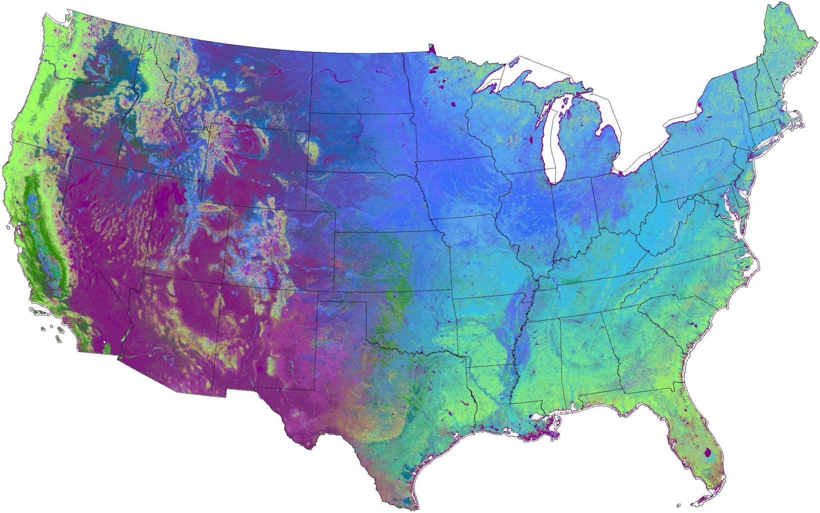
50 Phenoregions Mode (Random Colors)



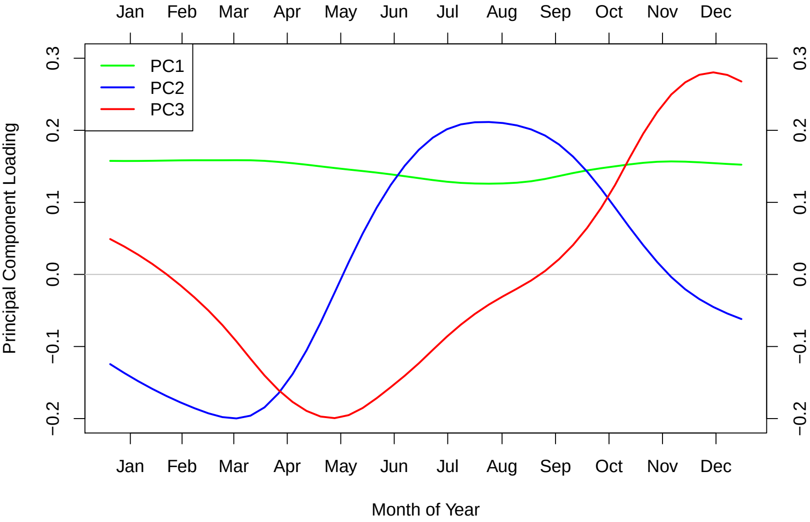
50 Phenoregions Max Mode (Random Colors)



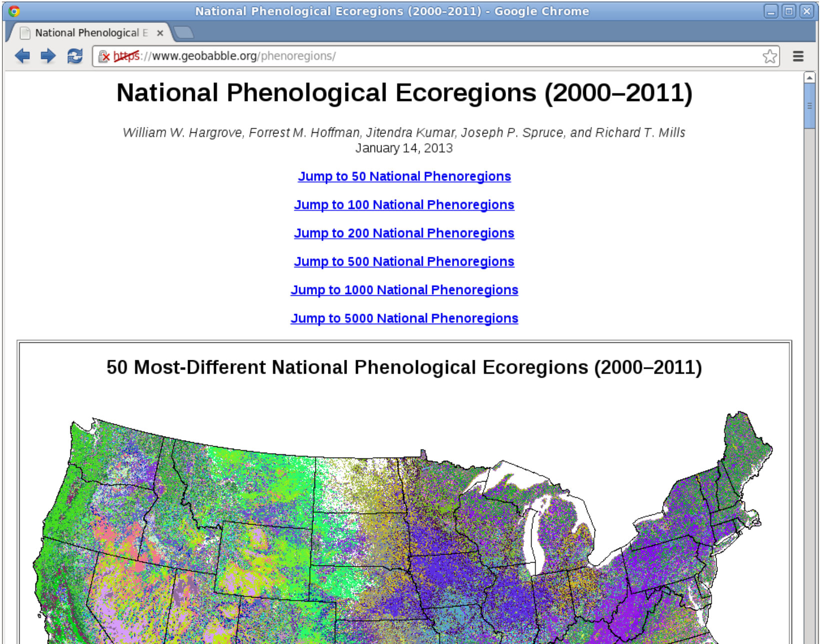
50 Phenoregions Max Mode (Similarity Colors)



50 Phenoregions Max Mode (Similarity Colors Legend)



Phenoregions Clearinghouse



Detecting and Tracking Shifts in Phenoregions

See Jitu's talk this afternoon on a new application of Phenoregions for detecting land cover change:

Detecting and Tracking Shifts in National Vegetation Composition Across the MODIS Era
Jitendra Kumar, Oak Ridge National Laboratory
Monday, July 06, 2015 — 5:20 pm

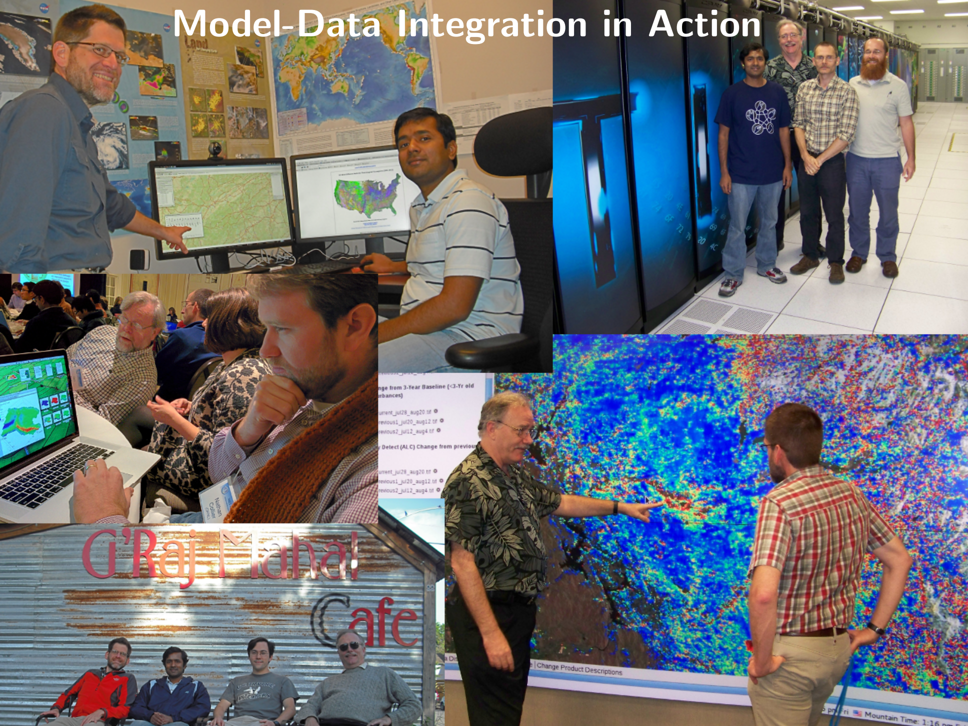
Computational Approaches for Landscape Ecology

- Moore's Law is no longer sustainable since:
  more transistors = more power = more heat loss.
- Speed is now a function of algorithm scalability.
- Future computational approaches must rely on:
  - distributed memory and shared memory parallelism (*threading*)
  - vectorization and cache reuse
  - algorithm acceleration techniques

Insights into computational directions at Tuesday's poster session:

Scalable algorithms for analysis of large geospatiotemporal data sets and applications to landscape ecology
Richard Mills, Intel Corporation
Tuesday, July 07, 2015 — 7:00 pm

Model-Data Integration in Action



References

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, W. W. Hargrove, D. J. Erickson, and R. J. Oglesby. Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models. *Earth Interact.*, 9(10):1–27, Aug. 2005. doi: 10.1175/EI110.1.

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. 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.

Acknowledgements

U.S. DEPARTMENT OF ENERGY

Office of Science

This research was sponsored 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 and the U. S. Department of Agriculture Forest Service. 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.