

Office of Science



Prospects for Satellite Remote Sensing to Identify Evolving Anthromes and Quantify Carbon Cycle Dynamics

Forrest M. Hoffman¹, Jitendra Kumar¹, Zachary L. Langford¹, V. Shashank Konduri², Russell Limber³, and William W. Hargrove⁴

March 28, 2023

¹Oak Ridge National Laboratory, Oak Ridge, TN, USA ²National Ecological Observatory Network, Boulder, CO, USA ³University of Tennessee, Knoxville, TN, USA ⁴US Department of Agriculture - Forest Service, Asheville, NC, USA Anthromes, CO₂, and Terrestrial Carbon From the deep past to net-zero March 27–30, 2023 Potomac, Maryland, USA

Anthromes – Anthropogenic Biomes – Human Biomes

- Originally conceived by Ellis and Ramankutty (*Front. Ecol. Environ.*, 2008) in a paper titled "Putting People in the Map: Anthropogenic Biomes of the World"
- Designed to recognize the terrestrial biosphere in its contemporary, human-altered and -managed form
- Urban and agriculture are the most obvious, but there are a hierarchy of human settlements (high- to low-density), villages dominated by a variety of agricultural practices, various croplands and rangelands, as well as populated and remote forests and unpopulated forests and barren lands
- Ellis and Ramankutty aimed to develop a framework for inclusion of human-created mosaics with natural systems, and the accompanying ecological interactions, directly into global ecosystem models
- Representation of prognostic human interactions remains woefully inadequate even in today's sophisticated Earth system models

Introduction

- Terrestrial ecosystems (vegetation and soils) represent a large natural sink of carbon
- This land carbon sink may be growing due to CO₂-fertilization and land use/management change
- Satellite remote sensing offers the ability to estimate ecosystem carbon state and dynamics at multiple spatial and temporal scales
- Earth observations are increasing in spatial resolution, temporal frequency, and spectral range
- Exascale computing enables rapid assimilation, simulation, and analysis of Earth system data
- Combined, these technologies can better constrain the carbon cycle and capture human influences



Sentinel-2 monitors vegetation, soil, and water cover



Frontier at Oak Ridge National Laboratory is the #1 fastest supercomputer on the <u>TOP500</u> List and the first supercomputer to break the exaflop barrier (Nov 14, 2022).

Vegetation Live Biomass and Its Changes (2000–2019)



Coefficient of variation (CV) of carbon

0.2

>0.3

0.1

produce large uncertainties in

biomass estimates

Left: (A) Spatial distribution of mean global vegetation carbon density. (B) Spatial distribution of coefficient of variation (CV) in global vegetation carbon density. **Right: (A)** Pixel-level (in 10-km resolution) vegetation carbon trend map. From Xu et al. (Sci. Adv., 2021).

Vegetation Indices

Vegetation indices (e.g., NDVI, EVI, NDRE) provide useful constraints on seasonal phenology and trends in vegetation change due to disturbance and land use and climate change





gov/images/2033/global-enhanced-vegetation-index-measurements

from

Index

anced Vegetation

MODIS https:/

Multivariate Geographic Clustering

- Ecoregions provide a useful framework for mapping biomes and were traditionally created by experts
- Our approach has been to objectively create ecoregions using continuous continental-scale data and clustering
- We developed a highly scalable *k*-means cluster analysis code that uses distributed memory parallelism
- Originally developed on a 486/Pentium cluster, the code now runs on the largest hybrid CPU/GPU architectures on Earth

Hargrove, W. W., F. M. Hoffman, and T. Sterling (2001), The Do-It-Yourself Supercomputer, *Sci. Am.*, 265(2):72–79,

https://www.scientificamerican.com/article/the-do-it-yourself-superc/

MAKING MAPS WITH THE STONE SOUPERCOMPUTER

TO DRAW A MAP of the ecoregions in the continental U.S., the Stone SouperComputer compared 25 environmental characteristics of 7.8 million one-square-kilometer cells. As a simple example, consider the classification of nine cells based on only three characteristics (temperature, rainfall and organic matter in the soil). Illustration A shows how the PC cluster would plot TEMP

76 SCIENTIFIC AMERICAN

tone the cells in a three-dimensional data space and group them into four f7:8 ecoregions. The four-region map divides the U.S. into recognizable idea to the country into 1,000 eco regions provides far more detail (C). Another approach is to represent three composite characteristics with varying TEMPERATURE levels of red, green and blue (D).



Copyright 2001 Scientific American, Inc.

AUGUST 2001



New Analysis Reveals Representativeness of the AmeriFlux Network

PAGES 529, 535

The AmeriFlux network of eddy flux covariance towers was established to quantify variation in carbon dioxide and water vapor exchange between terrestrial ecosystems and the atmos-

BY WILLIAM W. HARGROVE, FORREST M. HOFFMAN, AND BEVERLY E. LAW phere, and to understand the underlying mechanisms responsible for observed fluxes and carbon pook. The network is primarily funded by the U.S. Department of Energy, NASA, the National Oceanic and Atmospheric Administration, and the National Science Foundation. Similar regional networks elsewhere in the world—for example, CarboEurope, AsiaPlux, OzPlux, and Pluxnet Canada—participate in

carbon observation network within the North American Carbon Program (NACP). The NACP seeks to provide long-term, mechanistically detailed sphially resolved carbon fluxes across both of these roles, the AmeriFlux network should be ecologically representative of the environments contained within the geographic isboundaries of the program. A new ecoregionno, scale analysis of the existing AmeriFlux network reveals that, while central continental and flux towes are needed to expresent activonmental flux towes are needed to expresent activonmental

VOLUME 84 NUMBER 48 2 DECEMBER 2003

synthesis activities across larger geographic

areas [Baldocchi et al., 2001; Law et al., 2002] The existing AmeriFlux network will also

form a backbone of "Tier 4" intensive measurement sites as one component of a fourtiered

PAGES 529-544



Fig. 1. The representativeness of an existing spatial array of sample locations or study sites—for example, the AmeriFlux network of carbon dioxide eddy flux covariance towers—can be mapped relative to a set of quantitative ecoregions, suggesting locations for additional samples or sites. Distance in data space to the closest ecoregion containing a site quantifies how well an existing network represents each ecoregion in the map. Environments in darker ecoregions are poorly represented by this network.

Network Representativeness

- The *n*-dimensional space formed by the data layers offers a natural framework for estimating representativeness of individual sampling sites
 - The Euclidean distance between individual sites in data space is a metric of similarity or dissimilarity
- Representativeness across multiple sampling sites can be combined to produce a map of network representativeness

Hargrove, W. W., and F. M. Hoffman (2003), New Analysis Reveals Representativeness of the AmeriFlux Network, *Eos Trans. AGU*, 84(48):529, 535, doi:<u>10.1029/2003EO480001</u>.



Fig. 1 Map of the CTFS-ForestGEO network illustrating its representation of biodimatic, edaphic, and topographic conditions globally. Site numbers correspond to ID# in Table 2. Shading indicates how well the network of sites represents the suite of environmental factors included in the analysis; light-colored areas are well-represented by the network, while dark colored areas are poorly represented. Stippling covers nonforest areas. The analysis is described in Appendix S1.

Table 1 Attributes of a CTFS-ForestGEO census

Attribute	Utility		
Very large plot size	Resolve community and population dynamics of highly diverse forests with many rare species with sufficient sample sizes (Losos & Leigh, 2004; Condit et al., 2006); quantify spatial patterns at multiple scales (Condit et al., 2000; Wieggand et al., 2007ab; Detto & Muller-Landau, 2013; Lutz et al., 2013); characterize gap dynamics (Feeley et al., 2007b); calibrate and validate remote sensing and models, particularly those with large spatial grain (Mascaro et al., 2011; Réiou-Mechani et al., 2014).		
Includes every freestanding woody stem ≥1 cm DBH	Characterize the abundance and diversity of understory as well as canopy trees; quantify the demography of juveniles (Condit, 2000; Muller-Landau <i>et al.</i> , 2006a,b).		
All individuals identified to species	Characterize patterns of diversity, species-area, and abundance distributions (Hubbell, 1979, 2001; He & Legendre, 2002; Condit et al., 2005; John et al., 2007; Shen et al., 2009; He & Hubbell, 2011; Wang et al., 2011; Cheng et al., 2012); test theories of competition and coexistence (Brown et al., 2013); describe poorly known plant species (Gereau & Kenfack, 2000; Davies, 2001; Davies et al., 2001; Sonké et al., 2002; Kenfack et al., 2004; 2006)		
Diameter measured on all stems	Characterize size-abundance distributions (Muller-Landau et al., 2006b; Lai et al., 2013; Lutz et al., 2013); combine with allometries to estimate whole-ecosystem properties such as biomass (Chave et al., 2008; Valencia et al., 2009; Lin et al., 2012; Ngo et al., 2013; Muller-Landau et al., 2014)		
Mapping of all stems and fine-scale topography	Characterize the spatial pattern of populations (Condit, 2000); conduct spatially explicit analyses of neighborhood influences (Condit et al., 1992; Hubbell et al., 2001; Uriante et al., 2004, 2005; Küger et al., 2011, 2012; Lutz et al., 2014); Characterize microhabitat specificity and controls on demography, biomass, etc. (Harms et al., 2001; Valencia et al., 200 Chuyong et al., 2011); align on the ground and remote sensing measurements (Asner et al., 20, Mascaro et al., 2011);		
Census typically repeated every 5 years	Characterize demographic rates and changes therein (Russo et al., 2005; Muller- Landau et al., 2006a,b; Feeley et al., 2007a; Lai et al., 2013; Stephenson et al., 2014); characterize changes in community composition (Losos & Leigh, 2004; Chave et al., 2008; Feeley et al., 2011; Swenson et al., 2012; Chisholm et al., 2014); characterize changes in biomass or productivity (Chave et al., 2008; Banin et al., 2014; Muller-Landau et al., 2014)		

Optimizing Sampling Networks

- The CTFS-ForestGEO global forest monitoring network is aimed at characterizing forest responses to global change
 - The figure at left shows the global representativeness of the CTFS-ForestGEO sites in 2014
- Non-forested areas are masked with hatching, and as expected, they are consistently darker than the forested regions, which are represented to varying degrees by the monitoring sites

Anderson-Teixeira, K. J., et al. (2015), CTFS-ForestGEO: A Worldwide Network Monitoring Forests in an Era of Global Change, *Glob. Change Biol.*, 21(2):528–549, doi:<u>10.1111/gcb.12712</u>.

Representativeness for Alaska

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
Day of freeze	standard deviation	days	
Day of thaw	mean	day of year	GCM
Day of thaw	standard deviation	days	
length of growing season	mean	days	GCM
Length of growing season	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

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:<u>10.1007/s10980-013-9902-0</u>. 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 © The Author(s) 2013. This article is published with open access at Springerlink.com

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 km² resolution to define multiple sets of ecoregions across two decadal time periods. Maps of ecoregions for the

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: <u>kumar@climatemodeling.org</u>

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

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.

 $\begin{array}{lll} \textbf{Keywords} & Ecoregions \cdot Representativeness \cdot \\ Network \ design \cdot Cluster \ analysis \cdot Alaska \cdot \\ Permafrost \end{array}$

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

10 Alaska Ecoregions, Present and Future

(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

(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.

Sampling Network Design



NSF's NEON Sampling Domains

Gridded data from satellite and airborne remote sensing, models, and synthesis products can be combined to design optimal sampling networks and understand representativeness as it evolves through time



50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI Every 8 days (46 images/year) Clustered from year 2000 to present





50 Phenoregion Prototypes (Random Colors)

(Hargrove et al., in prep.)

EarthInsights

day of year



GSMNP: Spatial distribution of the 30 vegetation clusters across the national park

Extracted canopy height and structure from airborne LiDAR



(Kumar et al., in prep.)

10

10 km

GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

forests with slightly lower mean height with dense understory vegetation

low height grasslands and heath balds that are small in area but distinct landscape type



EarthInsights

Global Fire Regimes



Regions that exhibit similar fire seasonality globally

From MODIS "Hotspots" at 1 km resolution from 2002–2018

EarthInsights

(Norman et al., submitted)

Vegetation Distribution at Barrow Environmental Observatory



Leveraging Advances in Machine Learning for Earth Sciences

Existing machine learning techniques can improve understanding of biospheric processes and representation in Earth system models



Spanning Spatial & Temporal Scales for Ecosystem Modeling



Arctic Vegetation Mapping from Multi-Sensor Fusion

Used Hyperion Multispectral and IfSAR-derived Digital Elevation Model, applied cluster analysis, and trained a convolutional neural network (CNN) with Alaska Existing Vegetation Ecoregions (AKEVT)



Langford, Z. L., et al. (2019), Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks, *Remote Sens.*, 11(1):69, doi:10.3390/rs11010069.

Satellite Data Analytics Enables Within-Season Crop Identification



Figure: a) Comparison of cluster-then-label crop map with USDA Crop Data Layer (CDL) shows similar patterns at continental scale. b) Good spatial agreement is found at three selected regions, but cluster-then-label crop maps lack sharpness at field boundaries due to coarser resolution of MODIS data.

Earliest date for crop type classification



Konduri, V. S., J. Kumar, W. W. Hargrove, F. M. Hoffman, and A. R. Ganguly (2020), Mapping Crops Within the Growing Season Across the United States, *Remote Sens. Environ.*, 251, 112048, doi:<u>10.1016/j.rse.2020.112048</u>.

Watershed-Scale Plant Communities Determined from AVIRIS-NG with DNN



At the watershed scale, vegetation community distribution follows topographic and water controls. At a fine scale, nutrients limit the distribution of vegetation types.

EarthInsights

(Konduri et al., in prep.)

Tropical Phenology Study Area: Costa Rica and Panama

20) Shrubs
30) Herbaceous vegetation
40) Cultivated and managed vegetation/agriculture (cropland)
50) Urban/built up
60) Bare/sparse vegetation
70) Snow and Lee
80) Permanent water bodies
90) Herbaceous wetland
100) Moss and Lichen
111) Closed forest, evergreen needle leaf
112) Closed forest, deciduous needle leaf
113) Closed forest, deciduous broad leaf
115) Closed forest, mixed
116) Closed forest, wergreen needle leaf
120) Open forest, evergreen broad leaf
121) Open forest, evergreen needle leaf
123) Open forest, evergreen broad leaf
123) Open forest, deciduous needle leaf
124) Open forest, deciduous needle leaf
125) Open forest, deciduous broad leaf
126) Open forest, mixed
126) Open forest, unknown
200) Open sea

Buchhorn, M. ; Smets, B. ; Bertels, L. ; De Roo, B. ; Lesiv, M. ; Tsendbazar, N. - E. ; Herold, M. ; Fritz, S. Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe 2020.

Land Cover / Vegetation Types in Costa Rica and Panama

Class	Costa Rica [% area]	Panama [% area]
Shrubs	0.75	0.62
Herbaceous vegetation	1.30	2.39
Cropland	29.76	19.73
Urban	1.15	0.67
Bare	0.00	0.00
Water	2.59	1.10
Wetland	1.53	1.43
Evergreen Broadleaf, Closed	44.64	53.46
Deciduous Broadleaf, Closed	0.03	0.00
Mixed, Closed Forest	0.03	0.00
Closed Forest	1.52	2.47
Evergreen Broadleaf, Open	4.81	4.77
Deciduous Broadleaf, Open	0.02	0.00
Mixed, Open Forest	0.00	0.00
Open Forest	11.88	13.25

Satellite Remote Sensing Time Series

- Sentinel-2 time series for the period 2017–2022
- Normalized Difference Red Edge Index (NDRE):
 NDRE = (NIR RE) / (NIR + RE) | NDRE = (B8 B5) / (B8 + B5)
- Spatial resolution: 20m
- Temporal resolution: 15 days
- Data processing:
 - NDRE data processing, cloud/shadow removal performed within Google Earth Engine platform
 - The remaining analyses were conducted on our own clusters
 - A Big Data problem: ~324 million 20-m pixels × 150 time intervals = ~362 GB (double precision)
- NDRE data are noisy in space and time, so require corrections
- Noise filtering and regression-based gap filling were applied

Time Series of Normalized Difference Red Edge (NDRE) Index



NDRE Phenology by Land Cover Type for Costa Rica and Panama



Clustering Annual NDRE Phenology

- We applied clustering to identify the 10 (k = 10) most-different phenologies in the region
- Clusters 5, 6, and 8 constituted the largest areal extent
- Clusters 1, 9, and 10 exhibit no significant phenological signal

(Limber, Kumar, et al., in prep)



Cluster 5 Dominated by Cropland and Open Forest



Cluster 6 Dominated by Evergreen Broadleaf Closed Forest



Cluster 8: Mix of Cropland, Evergreen Broadleaf Closed & Open



Dynamic Spatial Pattern of NDRE Phenoregions



identify significant trends that may be driven by anthropogenic activities

Low Height Vegetation Exhibits the Strongest Seasonality



Vegetation Structure is Dynamic (NASA GEDI L2B)



Seasonal structural changes are exhibited across the full vertical profile of vegetation in Costa Rica and Panama. (Sampling biases in GEDI likely introduce minor artifacts) **Next steps:** Analyze meteorological drivers from reanalysis to correlate with changes in vegetation phenology and structure

Conclusions

- Satellite remote sensing complements in situ measurements and airborne remote sensing in providing constraints on the global carbon cycle
- Advanced statistical and machine learning methods offer powerful approaches for designing sampling strategies, combining multi-platform remote sensing data, extracting useful natural and anthropogenic signals from the data, and improving understanding of human–ecosystem interactions
- We may soon have enough sensors and platforms; spatial, temporal, and spectral resolution; and storage and computational capacity to begin realizing the human-integrated vision of Ellis and Ramankutty
- People are already in the map!
- Given the global extent and increasing effects of our changing climate system,
 All biomes are anthromes!