Representativeness-Based Sampling Network Design for the Arctic

Forrest M. Hoffman[1,2], Jitendra Kumar[1,2], William W. Hargrove[3], Martijn Pallandt[4], and Mathias Goeckedei[4]

[1]Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN USA; [1]University of Tennessee, Knoxville, TN USA; [3]US Department of Agriculture Forrest Service, Asheville, NC USA; and [4]Max Planck Institute for Biogeochemistry, Jena, Germany

September 27, 2018

ICEI 2018

10th International Conference on Ecological Informatics (ICEI 2018)

Friedrich Schiller Universität Jena, Germany

Oak Ridge National Laboratory

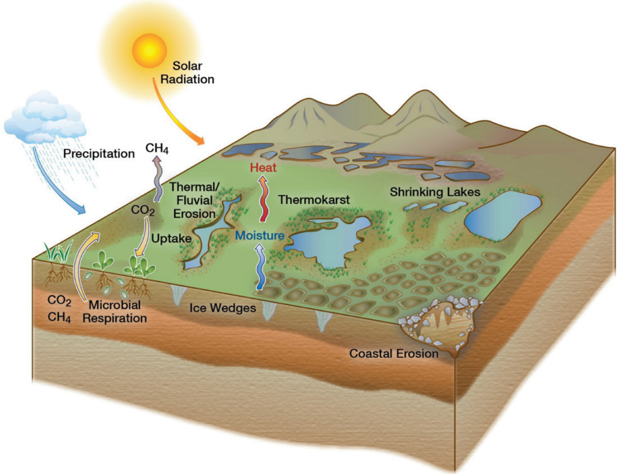
Climate Change Science Institute
Oak Ridge National Laboratory

Forest Service U.S. Department of Agriculture

MPI-BGC Jena

Next-Generation Ecosystem Experiments (NGEE Arctic)

http://ngee.ornl.gov/



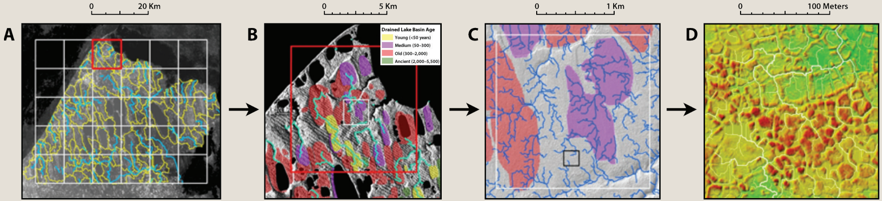
The Next-Generation Ecosystem Experiments (NGEE Arctic) project is supported by the Office of Biological and Environmental Research in the U.S. Department of Energy Office of Science.

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

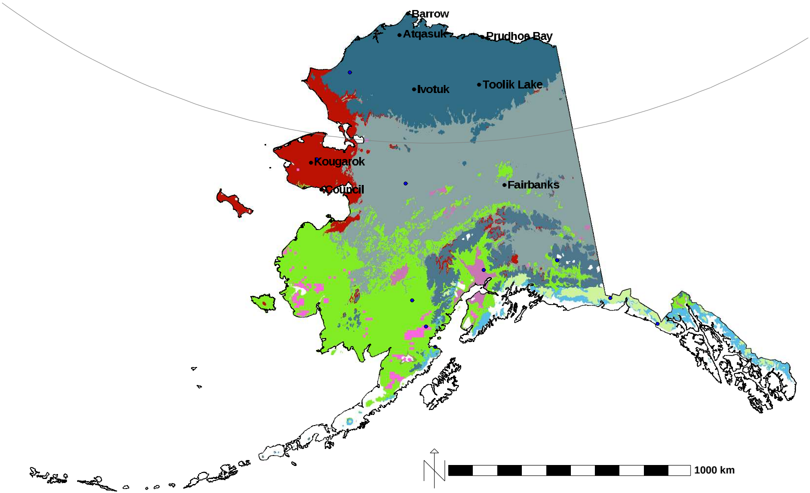
- Resource and logistical constraints limit the frequency and extent of observations, necessitating the development of a systematic sampling strategy that objectively represents environmental variability at the desired spatial scale.
- Required is a methodology that provides a quantitative framework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

Data Layers

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

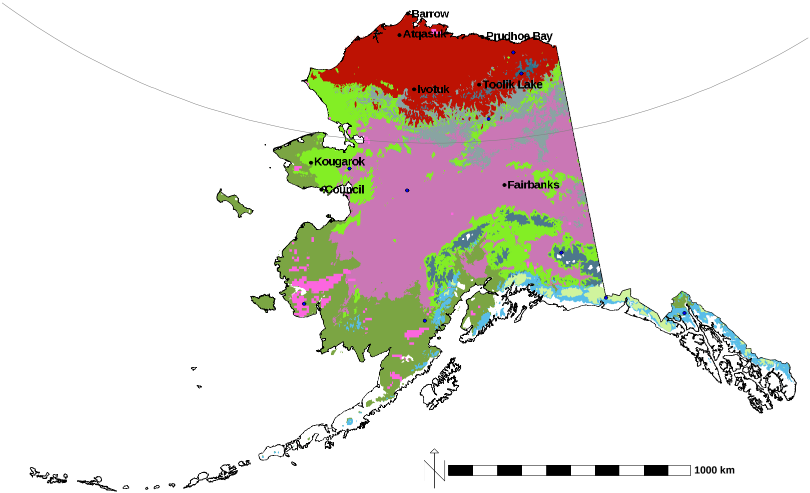
Description	Number/Name	Units	Source
Monthly mean air temperature	12	°C	GCM
Monthly mean precipitation	12	mm	GCM
Day of freeze	mean	day of year	GCM
	standard deviation	days	
Day of thaw	mean	day of year	GCM
	standard deviation	days	
Length of growing season	mean	days	GCM
	standard deviation	days	
Maximum 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 (2000–2009)



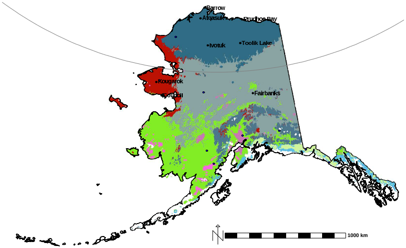
Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

10 Alaska Ecoregions (2090–2099)

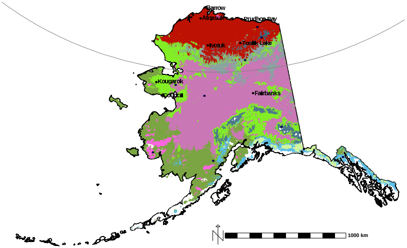


Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

10 Alaska Ecoregions, Present and Future



2000–2009

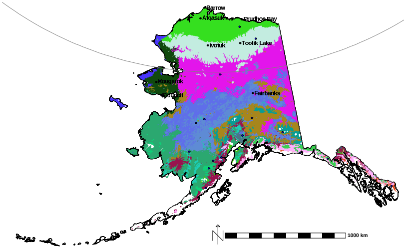


2090–2099

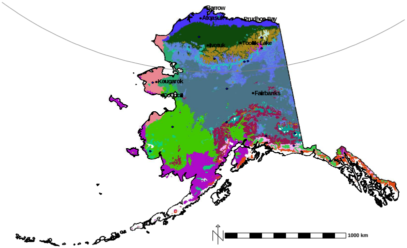
(Hoffman et al., 2013)

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

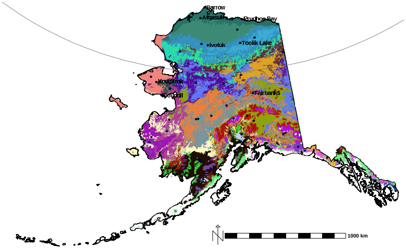


2090–2099

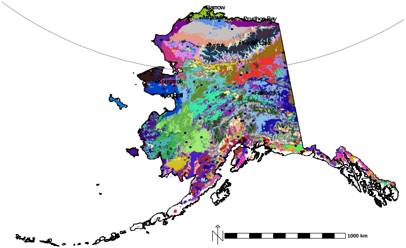
(Hoffman et al., 2013)

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

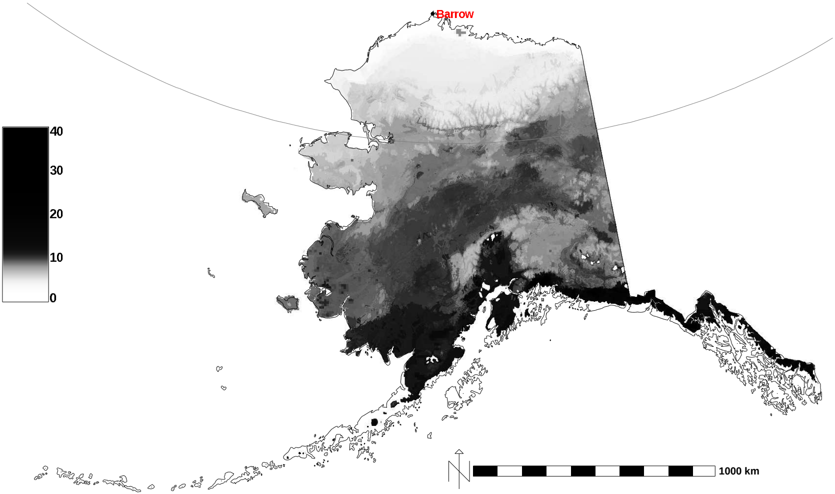
(Hoffman et al., 2013)

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.

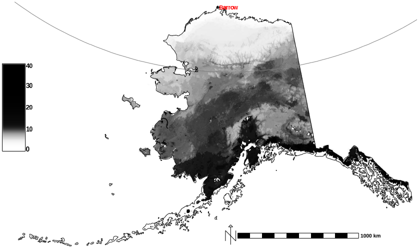
Present Representativeness of Barrow or "Barrow-ness"



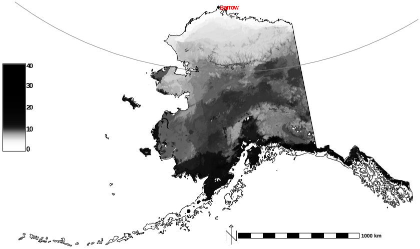
(Hoffman et al., 2013)

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

Present vs. Future Barrow-ness



2000–2009

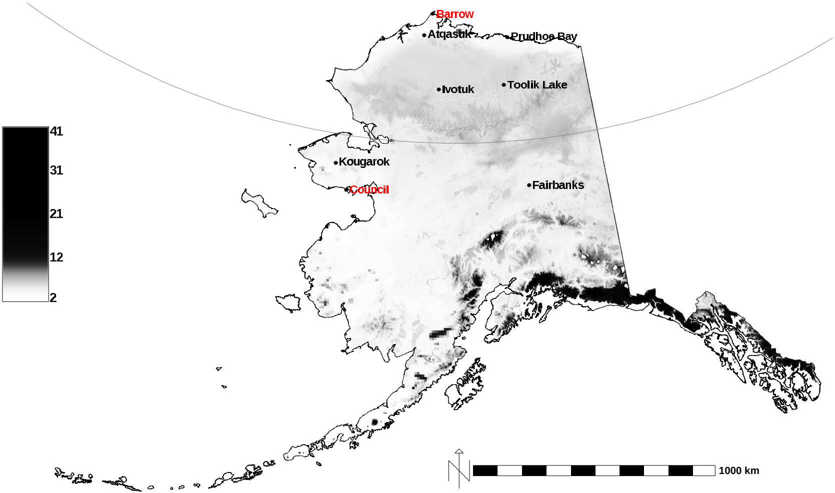


2090–2099

(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**.

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

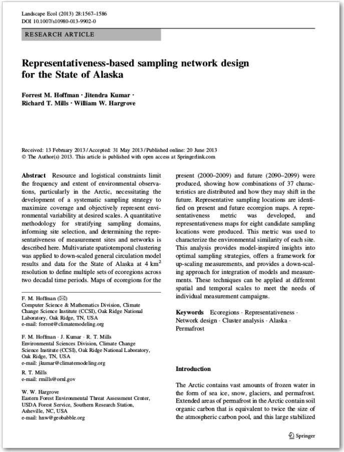
Table: Site state space dissimilarities between the present (2000–2009) and the future (2090–2099).

		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

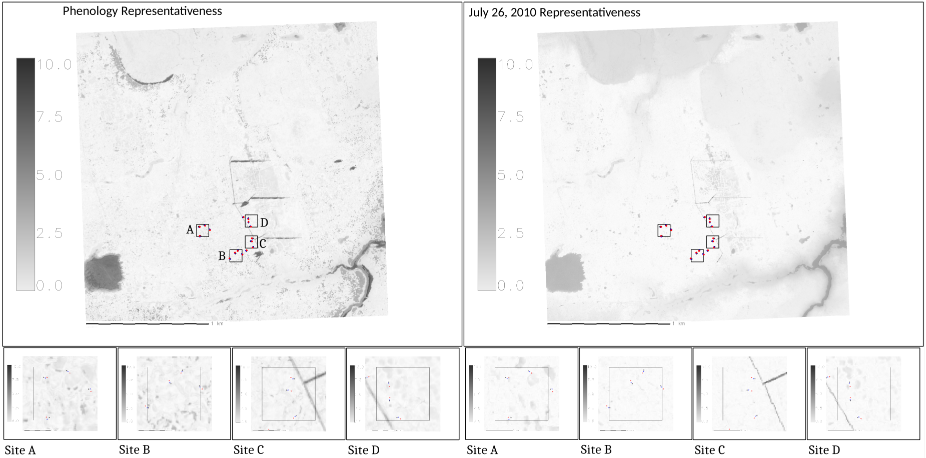
Introduction

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

Springer

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.

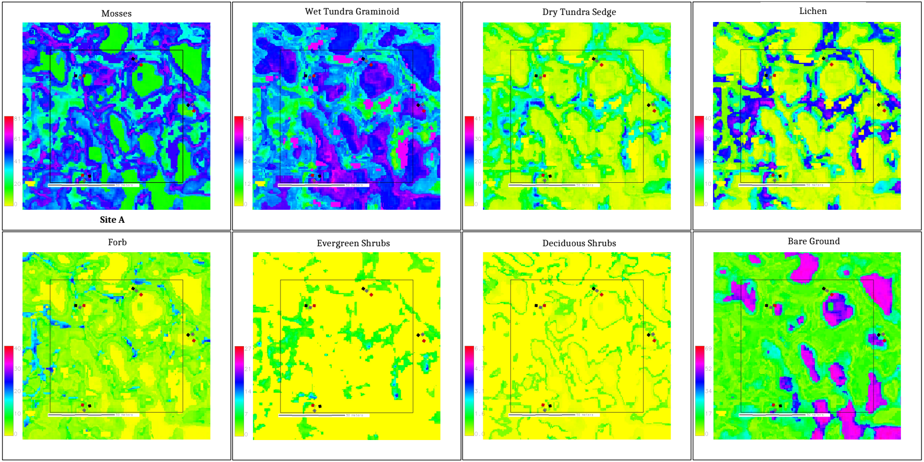
Barrow Environmental Observatory (BEO)



(Langford et al., 2016)

Representativeness map for vegetation sampling points in A, B, C, and D sampling area with phenology (left) and without (right), based on WorldView2 satellite images for the year 2010 and LiDAR data.

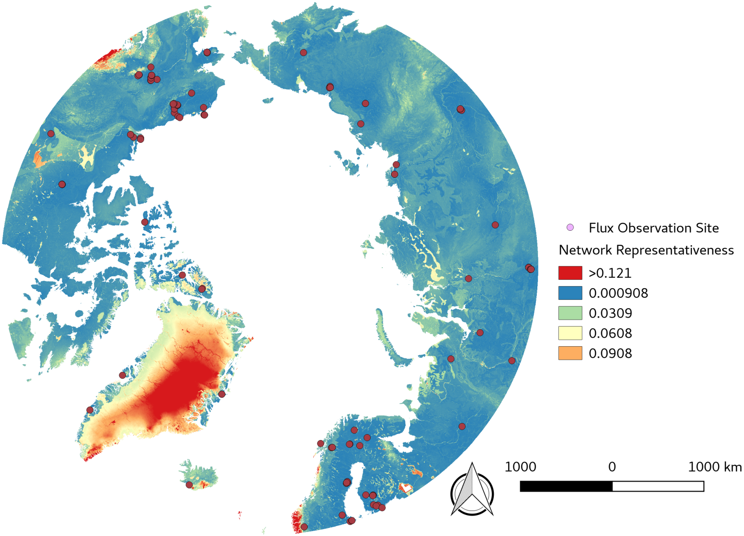
Barrow Environmental Observatory (BEO)



(Langford et al., 2016)

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

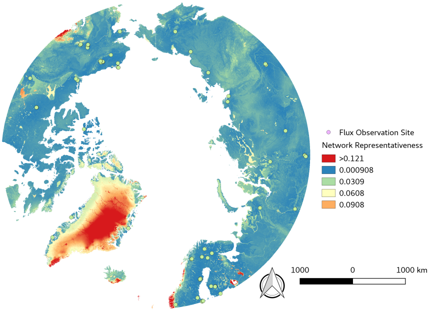
Pan-Arctic Representativeness of Measurement Sites



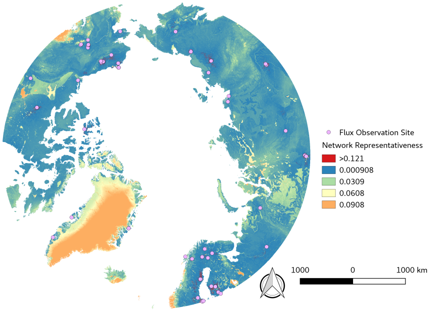
Network of 131 flux observation sites across the Arctic

Work with Martijn Pallandt and Mathias Goeckedei at MPI-BGC Jena

Pan-Arctic Representativeness of Measurement Sites

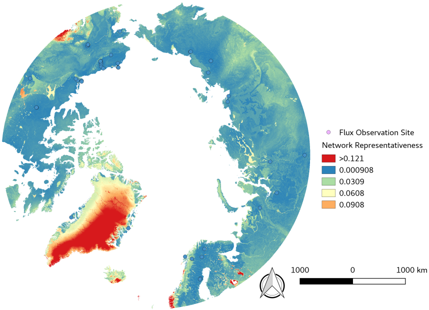


93 currently active sites

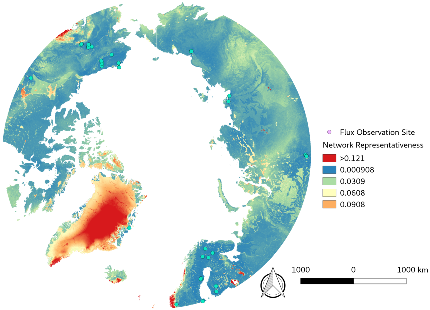


92 sites with >=5 years of data

Pan-Arctic Representativeness of Measurement Sites

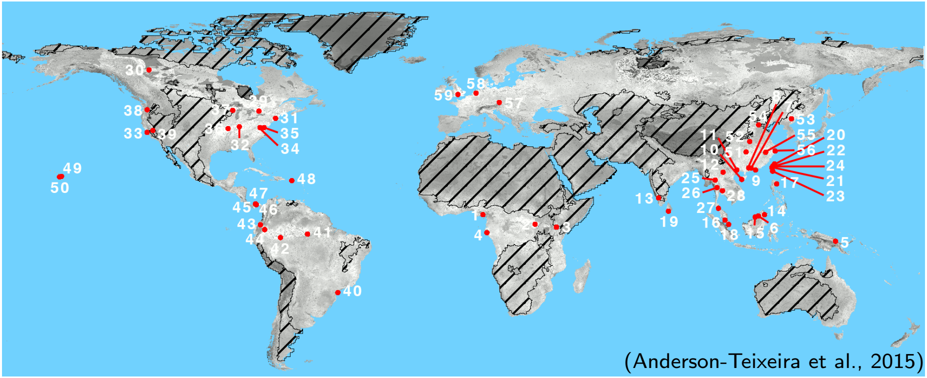


52 sites measuring CH_4



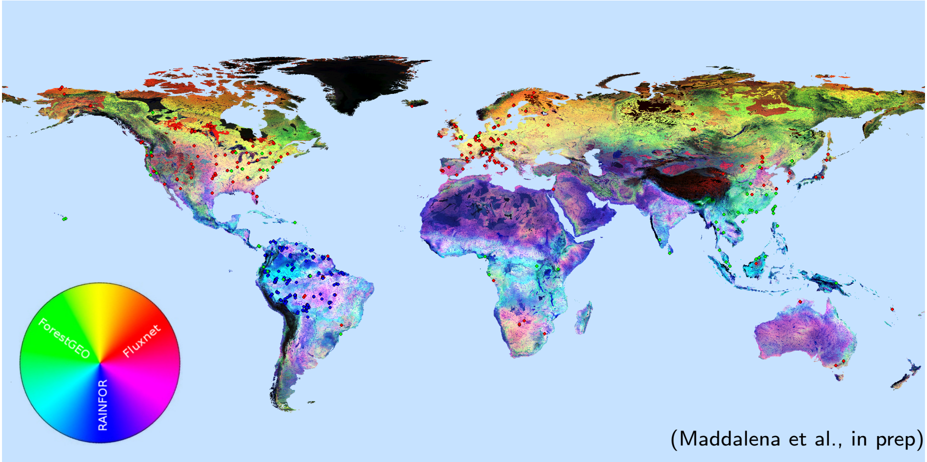
53 sites with winter operations

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



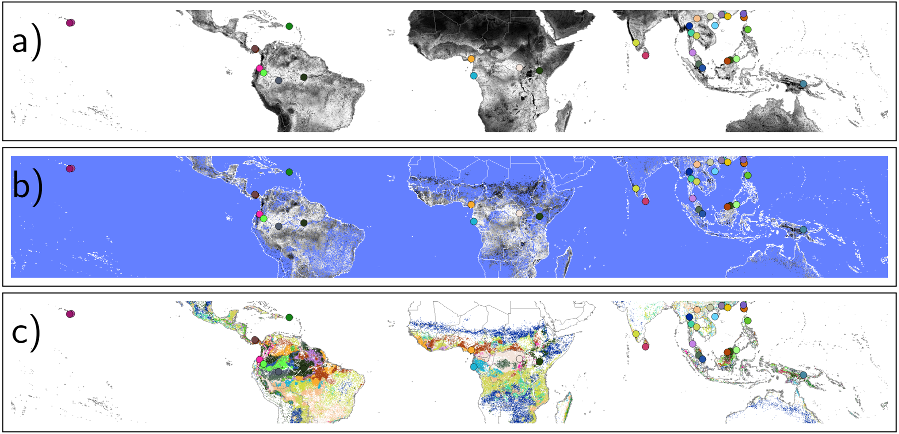
Map indicates which sampling network offers the most representative coverage at any location. Every location is made up of a combination of three primary colors from Fluxnet (red), ForestGEO (green), and RAINFOR (blue).

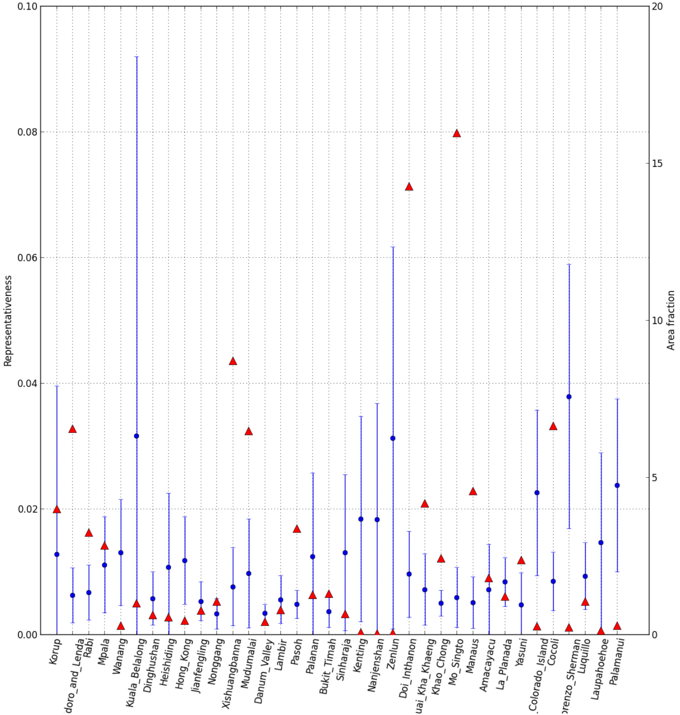
Global Forest Site Constituency

- For a fixed network of sampling sites, *constituency analysis* yields the spatial area represented best by any given site based on Euclidean distance in data space.
- For a given constituency, we can calculate a mean and standard deviation site representativeness.
- Thus,
  - a site with a large **constituency** provides broad spatial coverage;
  - a site with high **mean representativeness** (low dissimilarity) is a strong archetype of its constituency; and
  - a site with a large **standard deviation representativeness** provides broad data space coverage and is, therefore, the best (possibly poor) representative of a diverse constituency.
- These three metrics are (mostly) independent measures of network optimality.

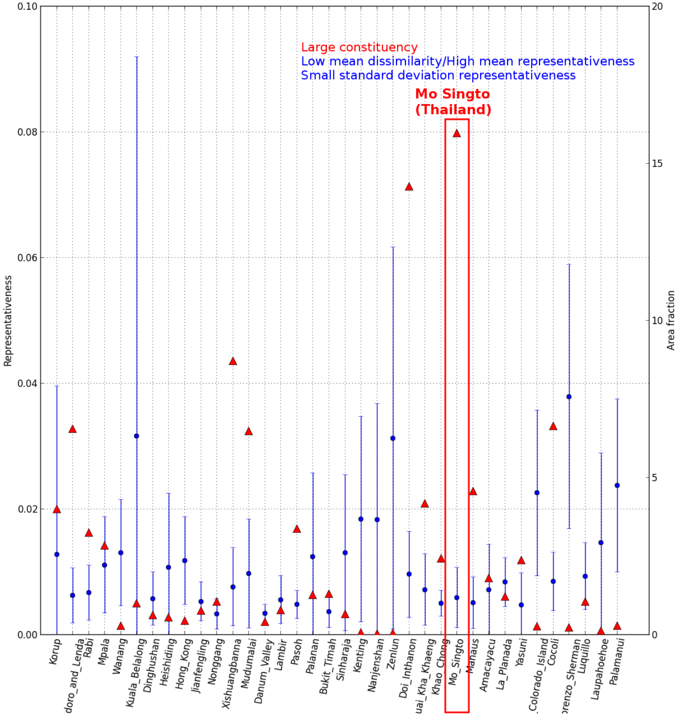
Tropical Forest Site Constituency

For insights into tropical forest network design, we perfomed representativeness and constituency analysis using the 36 CTFS-ForestGEO tropical sites to compute network a) representativeness, b) representativeness for tropical forests, and c) constituency for tropical forests.

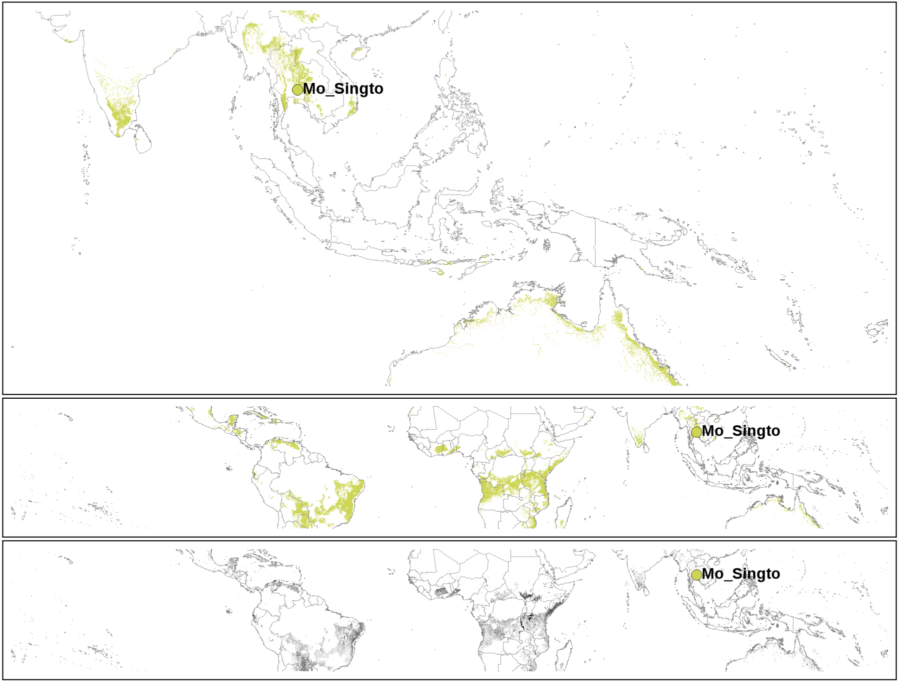


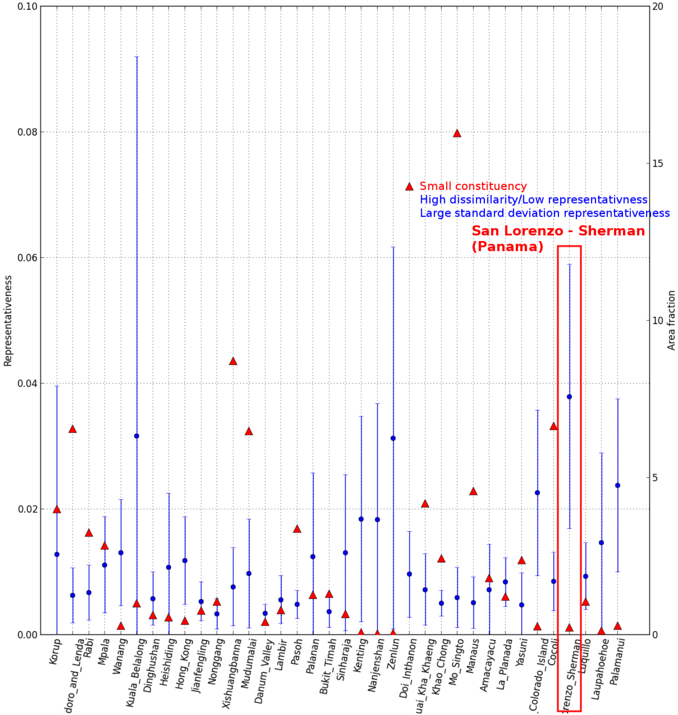


Representativeness
Area fraction
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto
Manaus
Amacayacu
La_Planada
Yasuni
Cocoli
Luquillo
Laupahoehoe
Palamanui

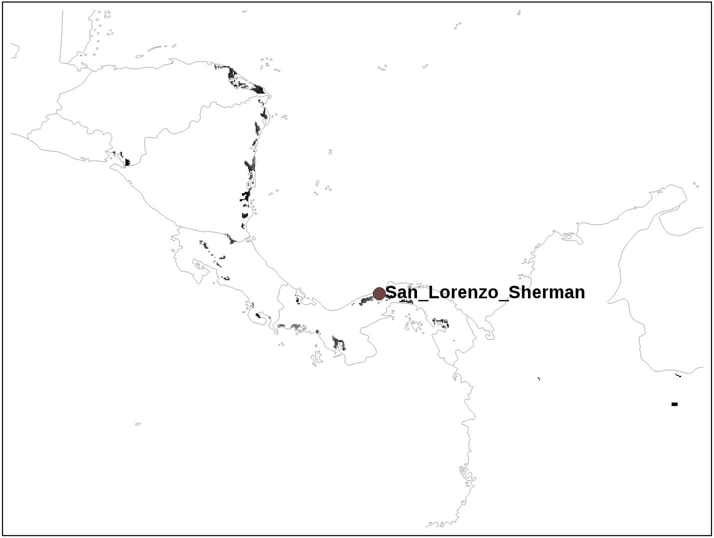


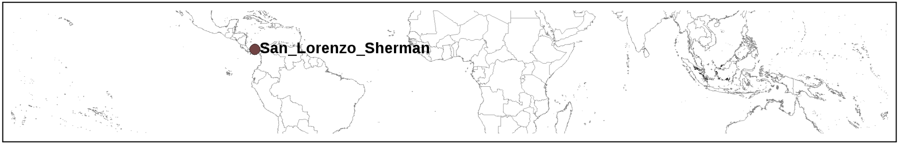
Large constituency
Low mean dissimilarity/High mean representativeness
Small standard deviation representativeness
Mo Singto
(Thailand)
Representativeness
Area fraction
0.10
0.08
0.06
0.04
0.02
0.00
20
15
10
5
0
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Mo_Singto
Amacayacu
La_Planada
Yasuni
Cocoli
Luquillo
Laupahoehoe
Palamanui


Mo_Singto
Mo_Singto
Mo_Singto



Small constituency
High dissimilarity/Low representativness
Large standard deviation representativeness
San Lorenzo - Sherman (Panama)
Representativeness
Area fraction
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto
Manaus
Amacayacu
La_Planada
Yasuni
Colorado_Island
Cocoli
Luquillo
Laupahoehoe
Palamanui


San_Lorenzo_Sherman


San_Lorenzo_Sherman

Conclusions

- **Multivariate Spatiotemporal Clustering (MSTC)** provides a quantitative framework for stratifying sampling domains, informing site selection, and determining representativeness of measurements.
- **Representativeness Analysis** and **Constituency Analysis** provide a systematic approach for optimizing site selection and up-scaling point measurements to larger domains.
- Methodology is *independent of resolution and surrogate data*, thus can be applied from site/plot scale to landscape/global scale with site measurements, remote sensing, and models.

For more information or to collaborate, please contact:

Forrest M. Hoffman
forrest@climatemodeling.org

Acknowledgements

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 National Center for Computational Sciences (NCCS) 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.

References

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.

Z. Langford, J. Kumar, F. M. Hoffman, R. J. Norby, S. D. Wullschleger, V. L. Sloan, and C. M. Iversen. Mapping Arctic plant functional type distributions in the Barrow Environmental Observatory using WorldView-2 and LiDAR datasets. *Remote Sens.*, 8 (9):733, Sept. 2016. doi: 10.3390/rs8090733.

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.