Representativeness-Based Sampling Network Design and Scaling Strategies for Measurements in Arctic and Tropical Ecosystems

Forrest M. Hoffman[1, 2], Jitendra Kumar[2], Zachary Langford[2], Damian Maddalena[2], Nathan Collier[2], Victoria Sloan[2], Richard T. Mills[3], and William W. Hargrove[4]

[1]University of California – Irvine, [2]Oak Ridge National Laboratory (ORNL), [3]Intel Corp., and [4]USDA Forest Service

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. Required is a quantitative methodology for stratifying sampling domains, informing site selection, and determining the representativeness of measurement sites and networks. Multivariate spatiotemporal clustering was applied to down-scaled general circulation model results and data for the State of Alaska at 2 km $\times$ 2 km resolution to define multiple sets of bioclimatic 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 bioclimatic 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.

Quantitative Delineation of Ecoregions

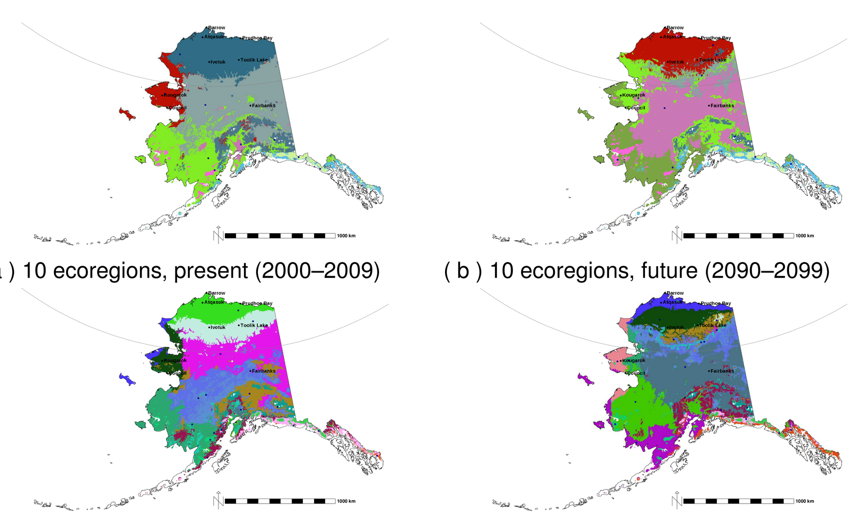
We developed a Multivariate Spatiotemporal Clustering (MSTC) methodology based on *k*-means clustering that uses high performance computing (Hoffman and Hargrove, 1999; Hargrove and Hoffman, 2004) and applied it to create maps of ecoregions for the State of Alaska.

Using 2 km $\times$ 2 km maps of 37 characteristics for the State of Alaska, derived from down-scaled general circulation model results and data (Table 1) for present (2000–2009) and future (2090–2099) decades (Walsh et al., 2008), we created maps of ecoregions at various levels of division, including 10, 20, 50, 100, 500, and 1000 ecoregions.

Table 1: *The 37 characteristics or variables, averaged for 2000–2009 and 2090–2099, used in Multivariate Spatiotemporal Clustering (MSTC) for the State of Alaska.*

Description	Number or 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	SRTM30

Maps for 10 and 20 ecoregions for the present and future are shown in Figure 1. Comparison of present with future maps shows how environmental conditions will shift, according to model projections.



(a) 10 ecoregions, present (2000–2009) (b) 10 ecoregions, future (2090–2099)

(c) 20 ecoregions, present (2000–2009) (d) 20 ecoregions, future (2090–2099)

Figure 1: *The 10 (a and b) and 20 (c and d) most-different quantitatively defined ecoregions for the State of Alaska in the present (a and c) and future (b and d) decades were derived from 37 variables and are shown using random colors. Realized centroids, map locations most closely approximating the mean value within an ecoregion of all the 37 variables, are indicated by the blue dot in each ecoregion.*

The cluster centroids from the MSTC procedure represent the mean values of all characteristics for every ecoregion. Tables 3–6 show the values of all 37 characteristics for the 10 centroids for the State of Alaska in both the present and future decades.

Table 3: Precipitation for 10 Alaska Ecoregions

	Monthly Mean Precipitation (mm)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	328.42	284.15	248.03	213.67	213.59	173.93	202.24	283.41	429.71	523.36	387.81	383.70
2	29.06	21.48	22.60	20.85	16.53	35.36	53.89	72.98	55.97	40.90	33.40	33.55
3	23.79	15.13	17.31	17.14	16.84	34.64	48.53	69.06	47.68	36.91	26.46	24.55
4	52.87	45.42	43.99	36.14	41.55	66.09	87.36	116.79	98.97	75.19	56.97	54.83
5	27.86	21.10	20.29	15.67	23.40	55.77	69.13	77.37	56.34	39.13	28.88	26.97
6	46.02	38.39	41.14	34.36	36.75	48.58	61.56	100.36	84.54	62.36	53.71	51.05
7	70.13	58.04	62.02	50.47	52.88	63.39	80.38	128.24	118.58	89.91	82.71	76.47
8	559.21	476.17	428.45	381.38	375.37	287.92	347.00	486.23	755.09	914.55	651.59	693.75
9	115.78	102.92	99.70	77.83	83.27	143.64	182.02	206.01	215.50	180.12	119.10	126.89
10	36.12	31.06	31.52	25.20	27.09	64.58	77.77	98.97	69.45	47.02	42.52	43.39

Table 4: Temperature for 10 Alaska Ecoregions

	Monthly Mean Temperature (°C)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	-5.99	-4.04	-1.44	2.89	6.85	10.35	12.84	12.18	8.02	2.83	-2.42	-4.79
2	-15.50	-18.87	-16.20	-9.48	0.67	8.95	12.71	10.87	5.04	-3.57	-9.19	-13.97
3	-23.36	-25.20	-21.91	-13.14	-1.15	7.97	11.54	8.69	1.00	-10.26	-18.53	-24.92
4	-10.64	-10.70	-7.07	-0.99	6.38	11.53	14.19	12.73	7.49	-0.78	-6.59	-10.36
5	-18.89	-17.05	-11.27	-1.88	7.58	13.47	15.72	12.73	5.76	-4.72	-13.77	-18.82
6	-5.53	-6.60	-3.79	0.60	7.49	12.13	15.02	14.48	10.24	2.59	-2.12	-5.56
7	-2.66	-3.89	-1.33	2.44	8.38	12.64	15.56	15.28	11.24	3.89	0.50	-2.31
8	-11.72	-8.73	-5.78	-0.47	3.01	7.21	10.00	9.06	4.11	-1.25	-7.42	-10.43
9	-14.78	-13.36	-10.05	-3.69	1.69	6.61	9.25	7.79	2.11	-5.33	-11.44	-14.51
10	-12.10	-10.56	-5.20	2.92	11.11	15.91	18.05	15.93	9.81	-0.11	-6.68	-10.07

Table 5: Permafrost properties for 10 Alaska Ecoregions

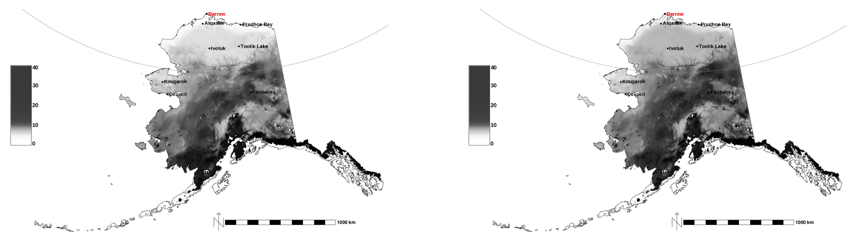
	Freeze Day (d)		Thaw Day (d)		GS Length (d)		Max AL Thick (m)	ΔT_{sn} (°C)	MAGT ALB (°C)	MAGST (°C)	Thermal Offset (°C)
	mean	stdev	mean	stdev	mean	stdev					
1	312.43	8.38	76.71	14.73	235.71	20.48	-0.23	1.07	3.82	4.07	-0.25
2	279.34	5.80	133.42	3.11	145.91	6.51	0.74	2.77	-1.87	-1.32	-0.55
3	262.53	1.62	138.98	2.76	123.55	2.83	0.62	3.63	-5.84	-5.38	-0.45
4	289.40	4.45	107.53	6.30	181.87	9.82	-0.44	1.70	1.28	2.00	-0.72
5	276.72	2.11	110.36	4.29	166.36	5.32	0.63	1.97	-1.48	-0.66	-0.83
6	311.55	9.96	92.86	15.41	218.69	24.00	-0.22	1.02	3.51	4.06	-0.55
7	329.34	17.32	70.29	31.07	259.05	42.78	-0.21	0.52	4.96	5.23	-0.27
8	283.29	4.86	110.22	7.53	173.38	10.28	0.01	1.80	0.36	0.74	-0.38
9	267.14	3.52	126.13	6.38	142.03	7.35	0.53	2.12	-2.01	-1.70	-0.31
10	291.63	5.32	93.33	8.27	198.30	12.38	-0.51	0.99	2.53	3.27	-0.74

Table 6: Limnicity, elevation and areas for 10 Alaska Ecoregions

	Limnicity (%)	Elevation (m)	Present (2000–2009)		Future (2090–2099)	
			Area (km^2)	% Area	Area (km^2)	% Area
1	0.91	911.04	33424	2.45	48356	3.54
2	3.61	395.02	93860	6.87	227188	16.63
3	3.62	543.53	295596	21.63	2316	0.17
4	3.33	440.21	302024	22.10	204408	14.96
5	1.49	412.60	486504	35.61	88952	6.51
6	52.78	37.88	16708	1.22	26308	1.93
7	5.45	189.60	1404	0.10	243244	17.80
8	0.20	1429.68	26352	1.93	22392	1.64
9	0.27	1587.51	92088	6.74	39512	2.89
10	1.47	315.57	18412	1.35	463696	33.94

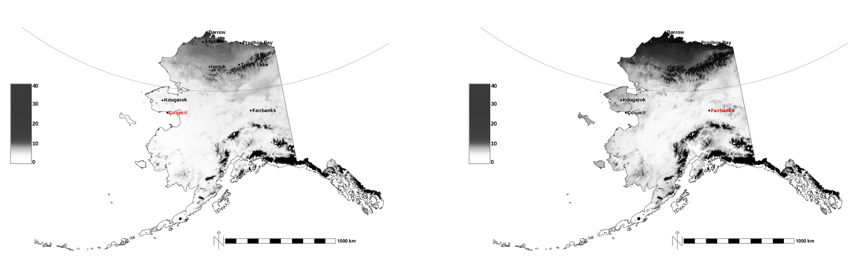
Site and Network Representativeness

To utilize limited point measurements at larger spatial and temporal scales for input to or evaluation of process modeling or for estimating landscape-scale characteristics, the representativeness of those measurements must be quantified in the context of a heterogeneous and evolving landscape. Our dissimilarity metric, calculated as the Euclidean distance between a sampling location and all other points on a map, is useful for informing site selection to maximize network coverage, up-scaling of point measurements, down-scaling of remote sensing data, and extrapolation of measurements to unsampled domains.



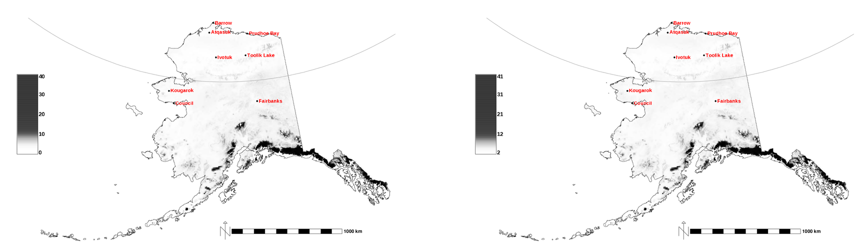
(a) Present period (2000–2009) (b) Future period (2090–2099)

Figure 2: *Point-based representativeness maps of present-day Barrow for the present and future time periods. White to light gray land areas are well-represented by Barrow, while dark gray to black land areas are poorly represented by Barrow.*



(a) Council (b) Fairbanks

Figure 3: *Point-based representativeness for other potential present-day NGEE Arctic sites for the present time period. White to light gray land areas are well-represented by the site, while dark gray to black land areas are poorly represented by the site.*



(a) Present period (2000–2009) (b) Future period (2090–2099)

Figure 4: *Point-based representativeness maps for a network of eight sites for the present and future time periods. White to light gray land areas are well-represented by the network of sites, while dark gray to black land areas are poorly represented by the network of sites.*

This same unit-less measure of dissimilarity can be computed between any two locations of interest to produce tables quantitatively characterizing dissimilarity of candidate sampling locations. Below we show that site dissimilarities, computed for eight candidate sampling locations in Alaska, may apply to the present (Table 7), the future (Table 8), or across time (Table 9).

Table 7: *Site state space distances 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

Table 8: *Site state space distances for the future (2090–2099).*

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

Table 9: *Site state space distances 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

Representativeness Within the Barrow Environmental Observatory (BEO)

The same representativeness methodology may be applied using any surrogate variables that hold predictive power for the quantities of interest being measured or scaled. Here we use multi-spectral remote sensing imagery from the WorldView2 satellite to determining the representativeness of vegetation sampling points and to extrapolate those limited samples across space to derive maps of plant functional types (PFTs) for input to models.

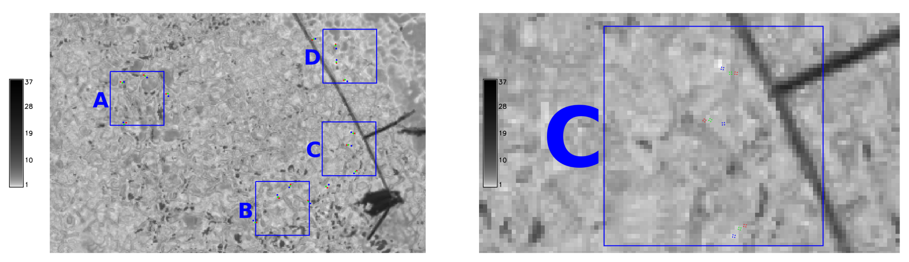
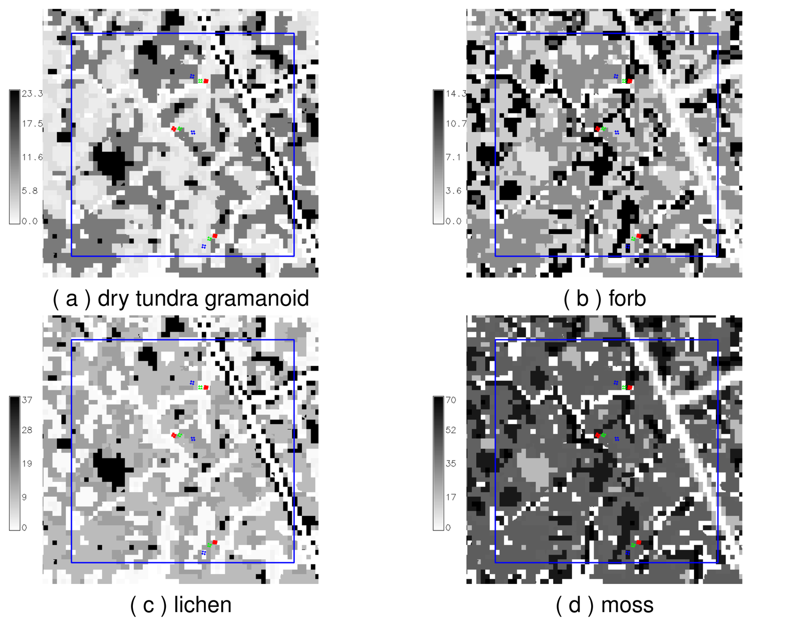


Figure 5: *Representativeness map for vegetation sampling points for the full A, B, C, and D sampling area (left) and zoomed in on the C sampling area (right) developed from multiple WorldView2 satellite images for the year 2010 and LiDAR data. Replicated vegetation sampling locations are shown as colored icons representing polygon troughs (red), edges (green), and centers (blue). White to light gray areas are well-represented by the collection of vegetation sampling locations, while dark gray to black areas are poorly represented by the sampling locations.*



(a) dry tundra gramanoid (b) forb

(c) lichen (d) moss

Figure 6: *Using WorldView2 satellite imagery and LiDAR data, plant functional type (PFT) distributions were scaled up to each sampling area based on their proportion at vegetation sampling locations. Four such PFT distribution maps are shown for area C.*

Global Forest Monitoring and Tropical Forest Representativeness

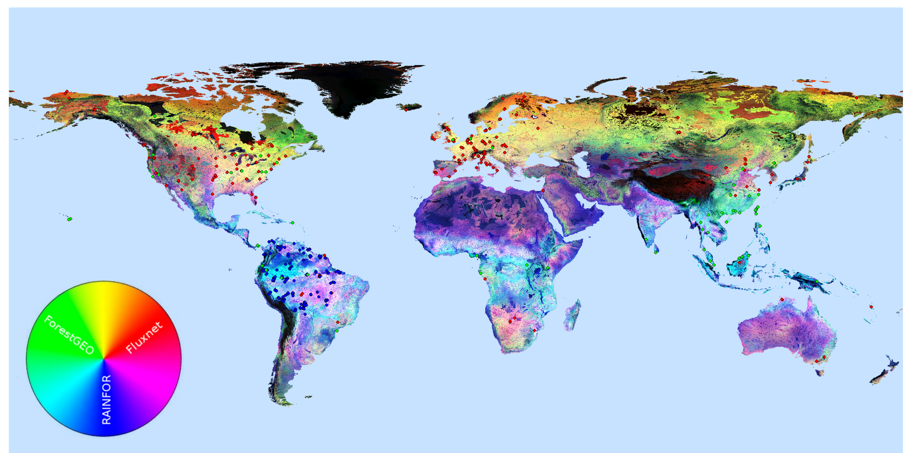


Figure 7: *A three-network representativeness map may be constructed by assigning each network to a primary color. Here, we show a color map indicating which networks offer the most representative coverage at any location. Every location on the map is then made up of combinations of these three primary colors, depending on how well each network represents each location.*

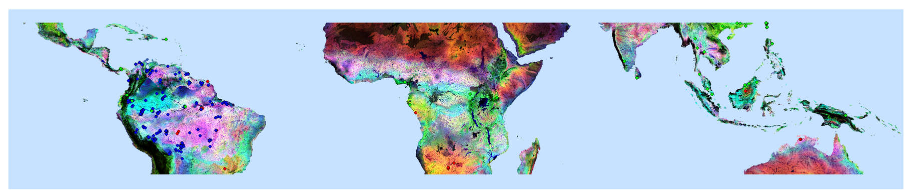


Figure 8: *Similar network representativeness maps may be constructed for any region of interest, like tropical forests only.*

References

Forrest M. Hoffman and William W. Hargrove. Multivariate geographic clustering using a Beowulf-style parallel computer. In Hamid R. Arabnia, editor, *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA '99)*, volume III, pages 1292–1298. CSREA Press, June 1999. ISBN 1-892512-11-4.

William W. Hargrove and Forrest M. Hoffman. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environ. Manage.*, 34(Supplement 1): S39–S60, April 2004. doi: 10.1007/s00267-003-1084-0.

John E. Walsh, William L. Chapman, Vladimir Romanovsky, Jens H. Christensen, and Martin Stendel. Global climate model performance over Alaska and Greenland. *J. Clim.*, 21 (23):6156–6174, December 2008. doi: 10.1175/2008JCLI2163.1.

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

Acknowledgments

This research was sponsored by the Terrestrial Ecosystem Science (TES) and the Regional and Global Climate Modeling (RGCM) programs within the U.S. Department of Energy Biological and Environmental Research (BER) program and the U.S. Department of Agriculture Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC). This research used resources of the Oak Ridge Leadership Computing Facility 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.

Want a copy of this poster to read later? Just scan this QR code with your smartphone!