Representativeness-Based Sampling Network Design for the Arctic

Forrest M. Hoffman[1], Jitendra Kumar[2], Richard T. Mills[1], and William W. Hargrove[3]

[1]Computational Earth Sciences Group,
Oak Ridge National Laboratory, Oak Ridge, TN
[2]Environmental Sciences Division,
Oak Ridge National Laboratory, Oak Ridge, TN
[3]Eastern Forest Environmental Threat Assessment Center,
Southern Research Station, USDA Forest Service, Asheville, NC

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. Described 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 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 environemntal 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.

Forrest M. Hoffman (forrest@climatemodeling.org), Jitendra Kumar (jkumar@climatemodeling.org), Richard T. Mills (rmills@ornl.gov), William W. Hargrove (hnw@geobabble.org)

1 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 carbon store could be released by widespread thawing of permafrost, resulting in a positive feedback to climate warming [1]. The Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) has documented strong evidence for warming of the Earth's climate over the last century and has attributed the increase in global temperatures primarily to the rising anthropogenic greenhouse gas burden [2]. Climate warming is projected to continue with broad implications for sensitive ecosystems and globally important climate feedbacks [3]. Warming is projected to be especially pronounced at high latitudes and accompanied by significant regional impacts. Evidence of Arctic-wide responses is already being observed [4]. Despite these potential implications, the Arctic has a limited record of low density observations. The The Arctic Cimate Impact Assessment (ACIA) [5] emphasized the need for studies of the complex and interacting processes of the atmosphere, sea ice, ocean, and terrestrial systems to improve the interpretation of past climate and projections of future climate. Committee on Designing an Arctic Observing Network [6] identified critical needs and gaps for observations in the Arctic. It recommended an Arctic Observing Network to satisfy current and future scientific needs and offered recommendations on key physical, biogeochemical, and human dimensions variables to monitor.

Conducting systematic and continuous field observations and long term monitoring are challenging, particularly in the Arctic. 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. Statistical design of the network, particularly the location of sampling sites, is critical under such harsh working conditions to maximize the representativeness of the sampled data, given a fixed number of sampling locations. Required is a methodology that provides a quantitative framework for stratifying sampling domains, informing site selection, and determining the representativeness of measurements. This information is required for up-scaling and extrapolating point measurements to a larger landscape with similar environmental characteristics. This study addresses these needs by developing a quantitative methodology, based on the concept of ecoregions, for objectively delineating sampling domains, identifying optimal sampling locations for these domains, and quantifying representativeness of sites and measurements. This methodology is applied at the landscape scale to inform the design of a sampling network for the U.S. Department of Energy's Next Generation Ecosystem Experiment (NGEE) Arctic project in the State of Alaska. The National Science Foundation's (NSF's) National Ecological Observatory Network (NEON) adopted an objective, data-based methodology to define 20 optimal sampling domains across the conterminous United States [7, 8]. Described here is an extension of that same methodology applied both across space and through time to support identification of measurement sites and provide a framework for scaling measurements and model parameters for the NGEE Arctic project.

2 Delineation of Quantitative Ecoregions

2.1 Ecoregions

Ecoregions have been widely used to stratify geographic domains into nearly homogeneous land areas with respect to their geophysical, biological, and climatic characteristics. Since ecoregions are designed to correspond well with biome distributions and species ranges, they are frequently used as a framework for studying ecosystem structure and function. Qualitative and generalized ecoregion maps of the United States and the world have traditionally been developed by experts for studying ecosystem behavior or to define units for land management [9, 10, 11, 12]. Hargrove and Hoffman [13] used cluster analysis for quantitative delineation of ecoregions using a set of nine environmental characteristics for the conterminous United States at a resolution of 1 km^2, and subsequently demonstrated its application for sampling network design, environmental niche modeling, and comparison of global model predictions [14, 15]. Krohn et al. [16] applied clustering to create hierarchical biophysical regions for Maine at a 21 km^2 resolution. Jensen et al. [17] used agglomerative clustering for hierarchical classification of sub-watersheds in the Columbia River Basin using 19 indirect biophysical variables. In this study, k-means cluster analysis was used to delineate ecoregions having nearly equal within-region heterogeneity.

2.2 Multivariate Spatiotemporal Clustering (MSTC)

The k-means algorithm [18] clusters a dataset of n observation vectors ($\vec{X}_1$, $\vec{X}_2$, ..., $\vec{X}_n$) into a user-selected number of groupings or clusters (k). The algorithm begins by calculating the Euclidean distance of each observation to the initial centroid vectors ($\vec{C}_1$, $\vec{C}_2$, ..., $\vec{C}_n$) and classifies or assigns each observation to its nearest centroid. Each centroid vector is recalculated as the vector mean of all observations assigned to it. This classification and re-calculation process is iteratively repeated until fewer than some fixed proportion of observations change their cluster assignment between iterations. In the algorithm used here, convergence is assumed once fewer than 0.05% of the observations change cluster assignments. The results of the k-means algorithm are sensitive to the choice of initial centroids. Various heuristics may be employed for their selection, such as choosing initial centroids to have an even distribution within data space or to be spread along the edges of the distribution of observations. In this study, a multi-stage refinement method based on the work of Bradley and Fayyad [19] is employed.

For geographic or spatial stratification applications, observation vectors consist of map cells, the dimensions of which are the biological or geophysical characteristics or variables under consideration. For spatiotemporal partitioning, observation vectors consist of map cells at different time periods. Hoffman and Hargrove [20] developed a parallel version of the k-means algorithm for use on clusters of inexpensive personal computers [21], and this code was used in a meta-computing environment to cluster data using multiple supercomputers across the Internet [22]. Hoffman et al. [23] later implemented improvements to accelerate convergence, handle empty cluster

cases, and obtain initial centroids through a scalable implementation of the Bradley and Fayyad [19] method. Kumar et al. [24] extended this work to develop a fully distributed, highly scalable k-means parallel clustering tool for analysis of very large data sets, which was employed in the study presented here.

2.3 Input Data Layers

This analysis used a set of 37 environmental characteristics, or variables, shown in Table 1, from down-scaled general circulation model (GCM) results and observational data for the State of Alaska at a nominal resolution of 2 km $\times$ 2 km. These data were used to define a collection of ecoregions at multiple levels of division across two time periods for Alaska. Selection of these 37 variables reflects a compromise between desirability and availability. Model results were averaged for the present (2000–2009) and the future (2090–2099). This analysis combined temperature, precipitation, and related bio-climatic projections from a five-model composite data set of down-scaled GCM results for the A1B emissions scenario [25] described by Walsh et al. [26]; corresponding snow and permafrost projections from the Geophysical Institute Permafrost Lab (GIPL) 1.3 permafrost dynamics model forced with the composite GCM results [27]; limnicity data based on the National Hydrography Dataset (NHD), pre-processed by Arp and Jones [28]; and elevation data from the Shuttle Radar Topography Mission (SRTM). The same limnicity and elevation data were used for both time periods. Because the units of measurement differ between variables, all data were standardized such that each variable had a mean of zero and a standard deviation of one prior to clustering to equalize the contribution from each predictor.

2.4 Alaska Ecoregions

Nowacki and Brock [29] and Gallant et al. [30] produced ecoregion maps for the State of Alaska using two different expert-based methodologies, strongly focused on land form. Later, Nowacki et al. [31] produced a "unified" ecoregion map—combining the two expert-based techniques—by considering limited data and in consultation with experienced ecologists, biologists, geologists, and regional experts. While useful for some purposes, such qualitative maps are based on the subjective expertise of the person or group developing them and suffer from various limitations [32, 33]. The question of whether ecoregions can or should be developed using quantitative statistical methods or should rely upon human expertise has been a matter of debate among geographers [34]. In this study, MSTC was applied to derive ecoregions based on climate and topographic factors for the present and the future at multiple levels of division. The climate and topographic factors discussed in §2.3 describe the environmental conditions of each map cell and are the most important drivers controlling vegetation and primary production. Thus, groupings or clusters of similarly characterized map cells delineated based on these variables define unique ecoregions. As demonstrated by Hargrove and Hoffman [14], both present and projected future climate factors were included in the same analysis so that groups of similar cells were objectively determined across space and through time. MSTC

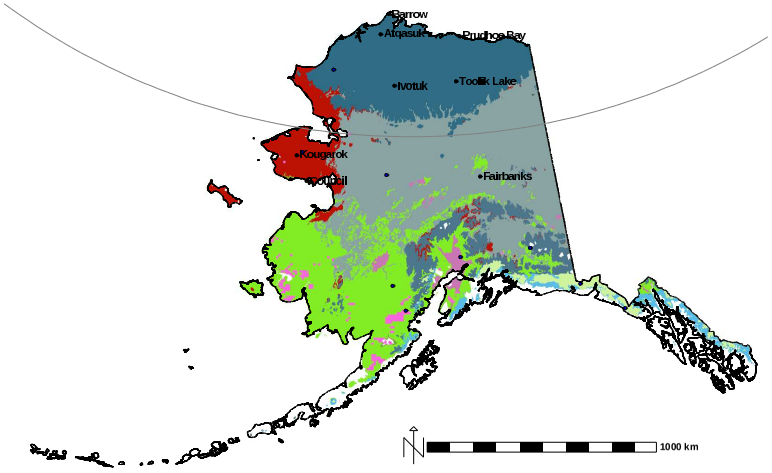
Table 1: The 37 characteristics or variables, averaged for 2000–2009 and 2090–2099, used in Multivariate Spatiotemporal Clustering (MSTC) for the State of Alaska.

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	SRTM

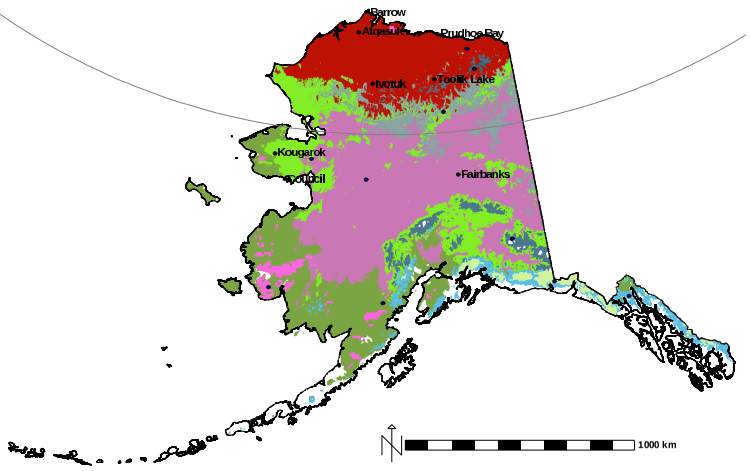
provides a basis for comparison of environmental conditions in the future with those in the present. Ecoregions constructed through this analysis may grow or shrink in spatial area and may shift across the landscape. At high levels of division or under extreme environmental change conditions, some present-day ecoregions may become extinct in the future (*i.e.*, shrink to zero spatial area), while others may exist only in the future (*i.e.*, have no analog in the present). This quantitative delineation of ecoregions across space and through time facilitates assessment of the magnitude of change between present and future environmental conditions and enables the evaluation of the ecological implications of climate change scenarios. From a conservation perspective, this methodology maps changing habitats and species at risk from climate change [35]. From a field sampling perspective, this methodology identifies regions fostering potentially vulnerable ecosystems or supporting large and vulnerable carbon stores that may be sensitive to climate change [36, 37]. Such ecoregions warrant intense observation and benefit from careful, quantifiable, and defensible sampling network design strategies.

Expert-derived ecoregion maps are static and have boundaries based on subjective consideration of geographic properties and expert judgment. In contrast, statistically derived ecoregions can vary with time and are delineated in the data space or state space representing all the characteristics under consideration. Moreover, the state space resolution can be varied by selecting different values of k, the level of division in the clustering algorithm. Figures 1(a) and 1(b) contain maps of the 10 quantitatively defined, most-different Alaskan ecoregions for the present and future, respectively. The cluster centroid of each ecoregion represents the mean value of all the characteristics or state variables for that ecoregion. Tables 2 and 3 show the 10 centroid values of all 37 state variables, as well as the land area and percent land

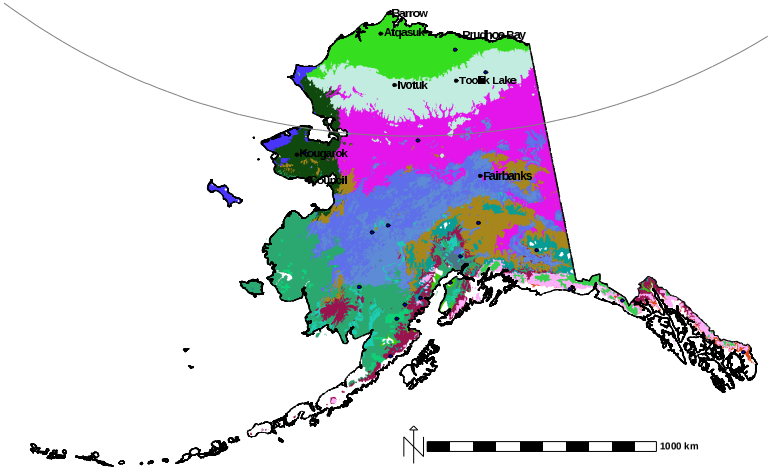
area for both the present and future time periods. Increasing the selected number of clusters in the k-means algorithm allows the definition of a larger number of more specifically defined, less generalized ecoregions. For example, Figures 1(c) and 1(d) contain maps of the 20 quantitatively defined, most-different Alaskan ecoregions for the present and future, respectively. The 20 cluster centroids values are shown in Tables 4 and 5, as well as the land area and percent land area for both the present and future time periods. By continuing to increase the level of division, the state space resolution can be further increased. Maps of Alaska were produced for $k = 5$, 10, 20, 50, 100, 200, 500, and 1000 ecoregions [38]. To demonstrate the additional state space resolution provided by higher levels of division, maps of 50 and 100 ecoregions for the present and future are shown in Figure 2. Since cluster centroids are calculated in the 37-dimensional state space, they may not actually exist in geographic space. However, the map cell closest to the calculated centroid in state space is easily identified. This cell is called the *realized centroid* for the ecoregion, and it best represents the combination of environmental conditions for the entire ecoregion. The location of these representative realized centroids is indicated by the blue dot in each ecoregion in Figures 1 and 2.



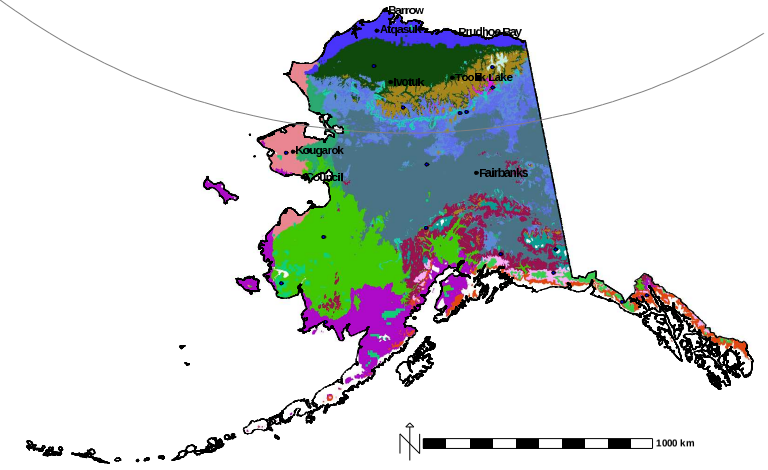
(a) 10 ecoregions, present (2000–2009)



(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.

Table 2: 10 Alaska Ecoregions with DEM, precipitation, and temperature

		Monthly Mean Precipitation (mm)												Monthly Mean Temperature (°C)											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	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	-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		29.06	21.48	22.60	20.85	16.53	35.36	53.89	72.98	55.97	40.90	33.40	33.55	-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		23.79	15.13	17.31	17.14	16.84	34.64	48.53	69.06	47.68	36.91	26.46	24.55	-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		52.87	45.42	43.99	36.14	41.55	66.09	87.36	116.79	98.97	75.19	56.97	54.83	-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		27.86	21.10	20.29	15.67	23.40	55.77	69.13	77.37	56.34	39.13	28.88	26.97	-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		46.02	38.39	41.14	34.36	36.75	48.58	61.56	100.36	84.54	62.36	53.71	51.05	-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		70.13	58.04	62.02	50.47	52.88	63.39	80.38	128.24	118.58	89.91	82.71	76.47	-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		559.21	476.17	428.45	381.38	375.37	287.92	347.00	486.23	755.09	914.55	651.59	693.75	-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		115.78	102.92	99.70	77.83	83.27	143.64	182.02	206.01	215.50	180.12	119.10	126.89	-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		36.12	31.06	31.52	25.20	27.09	64.58	77.77	98.97	69.45	47.02	42.52	43.39	-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 3: 10 Alaska Ecoregions with DEM, other environmental factors, and area

		Freeze Day (d)		Thaw Day (d)		GS Length (d)		Max AL	ΔT_{sn}	MAGT	MAGST	Thermal	Limnicity	Elevation	Present (2000–2009)		Future (2090–2099)	
		mean	stdev	mean	stdev	mean	stdev	Thick (m)	(°C)	ALB (°C)	(°C)	Offset (°C)	(%)	(m)	Area (km^2)	% Area	Area (km^2)	% Area
1		312.43	8.38	76.71	14.73	235.71	20.48	-0.23	1.07	3.82	4.07	-0.25	0.91	911.04	33424	2.45	48356	3.54
2		279.34	5.80	133.42	3.11	145.91	6.51	0.74	2.77	-1.87	-1.32	-0.55	3.61	395.02	93860	6.87	227188	16.63
3		262.53	1.62	138.98	2.76	123.55	2.83	0.62	3.63	-5.84	-5.38	-0.45	3.62	543.53	295596	21.63	2316	0.17
4		289.40	4.45	107.53	6.30	181.87	9.82	-0.44	1.70	1.28	2.00	-0.72	3.33	440.21	302024	22.10	204408	14.96
5		276.72	2.11	110.36	4.29	166.36	5.32	0.63	1.97	-1.48	-0.66	-0.83	1.49	412.60	486504	35.61	88952	6.51
6		311.55	9.96	92.86	15.41	218.69	24.00	-0.22	1.02	3.51	4.06	-0.55	52.78	37.88	16708	1.22	26308	1.93
7		329.34	17.32	70.29	31.07	259.05	42.78	-0.21	0.52	4.96	5.23	-0.27	5.45	189.60	1404	0.10	243244	17.80
8		283.29	4.86	110.22	7.53	173.38	10.28	0.01	1.80	0.36	0.74	-0.38	0.20	1429.68	26352	1.93	22392	1.64
9		267.14	3.52	126.13	6.38	142.03	7.35	0.53	2.12	-2.01	-1.70	-0.31	0.27	1587.51	92088	6.74	39512	2.89
10		291.63	5.32	93.33	8.27	198.30	12.38	-0.51	0.99	2.53	3.27	-0.74	1.47	315.57	18412	1.35	463696	33.94

Table 4: 20 Alaska Ecoregions with DEM, precipitation, and temperature

		Monthly Mean Precipitation (mm)												Monthly Mean Temperature (°C)											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1		42.10	33.50	33.72	23.23	18.06	33.21	60.39	89.64	69.79	38.81	39.92	41.97	-3.95	-8.88	-8.00	-4.26	3.34	9.72	13.84	13.40	9.32	2.26	-0.49	-2.98
2		22.83	15.95	17.46	16.53	10.77	23.18	39.50	55.51	38.58	32.05	28.60	28.47	-15.12	-19.83	-17.89	-11.29	-2.19	5.88	10.74	9.96	5.48	-0.94	-6.51	-12.11
3		49.62	42.37	44.68	37.71	39.97	50.78	62.74	102.58	89.56	68.10	57.94	54.75	-4.67	-5.68	-3.07	1.30	7.87	12.28	15.17	14.72	10.56	3.03	-1.51	-4.79
4		45.86	38.02	38.09	31.45	36.53	47.69	67.19	105.07	88.09	65.10	47.69	43.89	-9.12	-9.90	-6.82	-1.74	5.69	10.84	13.70	12.67	8.08	0.42	-5.18	-9.53
5		28.07	18.09	20.78	20.22	21.89	45.43	63.41	85.20	62.91	43.22	31.33	30.05	-22.22	-23.60	-19.83	-11.40	0.65	8.90	11.67	8.51	0.59	-11.03	-18.53	-24.22
6		364.13	312.26	271.95	238.48	228.77	179.90	205.24	285.08	443.86	578.52	434.85	424.60	-3.11	-1.26	0.87	4.72	8.60	11.88	14.23	13.76	9.90	5.11	0.36	-1.57
7		117.60	105.59	98.30	80.81	88.54	118.92	144.50	180.84	192.33	173.79	133.26	129.07	-9.45	-8.50	-5.20	0.64	6.02	10.43	12.98	11.88	6.72	-0.65	-5.97	-8.67
8		53.79	42.97	48.02	36.58	38.60	58.39	79.35	125.29	98.08	64.47	61.35	58.74	-6.33	-6.43	-2.44	2.95	9.79	14.08	16.57	15.73	11.01	2.44	-1.95	-5.79
9		25.38	17.39	18.26	14.29	20.55	47.78	60.69	69.13	52.44	35.97	25.87	23.16	-21.31	-19.23	-13.53	-3.24	7.14	13.50	15.74	12.40	4.95	-6.07	-15.92	-21.49
10		156.53	135.01	135.19	105.88	106.76	174.21	225.45	260.02	292.60	249.95	159.55	170.86	-15.81	-14.01	-11.07	-4.85	-0.36	4.50	7.28	6.00	0.25	-6.43	-12.27	-15.20
11		47.67	42.28	37.39	29.64	38.78	87.58	110.83	115.09	92.95	69.57	49.69	51.51	-15.52	-15.16	-11.11	-4.39	3.51	9.30	11.80	9.74	3.58	-5.43	-12.08	-15.52
12		641.20	539.78	490.23	435.77	422.73	319.13	387.06	541.90	846.19	1029.32	725.43	787.13	-12.52	-9.51	-6.58	-1.22	2.27	6.58	9.46	8.52	3.44	-1.91	-8.19	-11.16
13		18.51	11.49	12.95	13.33	10.45	21.10	29.64	48.73	28.09	28.97	20.40	17.84	-24.79	-27.31	-24.70	-15.54	-3.63	6.65	11.28	8.83	1.47	-9.33	-18.53	-25.82
14		80.31	68.74	72.43	61.82	63.48	68.58	82.12	131.32	134.42	109.45	97.72	90.52	-1.17	-2.03	0.08	3.68	8.86	12.53	15.46	15.43	11.66	4.73	1.64	-0.81
15		46.21	38.81	39.10	33.88	38.21	64.95	79.37	107.54	89.99	66.13	49.73	47.73	-11.59	-11.38	-7.12	-0.62	7.39	12.70	15.19	13.40	7.75	-1.08	-7.13	-10.94
16		392.07	340.33	299.20	260.04	266.12	214.63	254.60	360.12	549.93	650.39	467.93	481.65	-9.78	-7.13	-4.10	0.90	4.56	8.38	11.02	10.12	5.55	0.16	-5.83	-8.71
17		28.86	20.98	22.60	21.05	15.70	31.47	49.93	71.02	56.82	40.68	32.43	32.56	-15.98	-19.02	-16.23	-9.19	1.48	10.23	13.86	11.64	5.33	-3.95	-9.81	-14.73
18		31.69	26.04	25.60	21.05	26.69	61.40	76.54	92.91	68.11	45.84	34.54	34.52	-15.30	-14.18	-8.83	-0.63	8.02	13.58	15.86	13.26	6.89	-3.05	-10.52	-14.71
19		28.63	23.03	22.14	17.04	23.79	56.15	69.04	83.43	57.99	40.08	30.42	28.79	-16.72	-15.19	-9.21	-0.39	8.53	14.05	16.28	13.58	7.07	-3.00	-11.36	-16.35
20		34.17	29.98	28.80	23.42	26.50	68.75	81.15	95.54	64.90	45.61	40.73	41.68	-12.87	-11.12	-5.63	2.76	11.17	16.08	18.19	15.86	9.49	-0.58	-7.50	-10.44

Table 5: 20 Alaska Ecoregions with DEM, other environmental factors, and area

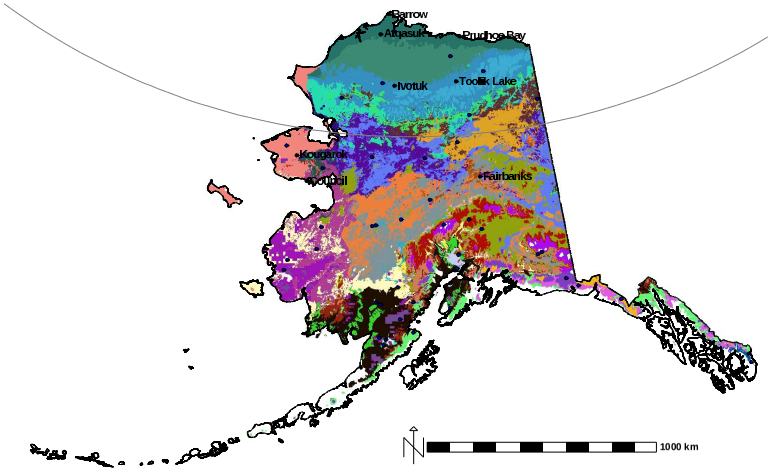
		Freeze Day (d)		Thaw Day (d)		GS Length (d)		Max AL	ΔT_{sn}	MAGT	MAGST	Thermal	Limnicity	Elevation	Present (2000–2009)		Future (2090–2099)	
		mean	stdev	mean	stdev	mean	stdev	Thick (m)	(°C)	ALB (°C)	(°C)	Offset (°C)	(%)	(m)	Area (km^2)	% Area	Area (km^2)	% Area
1		322.94	27.01	114.85	17.96	208.09	31.64	-0.37	1.37	2.65	3.10	-0.45	3.53	122.03	*Does not exist*		46492	3.40
2		288.09	9.93	143.19	2.66	144.90	10.39	0.56	2.84	-2.31	-1.69	-0.61	11.37	61.50	18200	1.33	60368	4.42
3		315.99	10.37	86.80	17.67	229.19	26.16	-0.20	0.95	4.08	4.54	-0.46	63.30	36.94	10100	0.74	16564	1.21
4		293.58	4.10	110.32	5.72	183.26	9.21	-0.29	1.39	1.23	2.00	-0.77	8.47	132.21	165252	12.09	32428	2.37
5		261.08	1.60	133.90	3.32	127.18	3.19	0.69	3.37	-5.36	-4.93	-0.43	0.44	880.29	161808	11.84	3100	0.23
6		332.98	12.96	53.79	23.44	279.19	32.05	-0.13	0.68	5.73	5.83	-0.09	1.06	800.61	4780	0.35	33308	2.44
7		289.03	4.36	101.38	8.43	187.65	12.06	-0.48	1.57	1.85	2.30	-0.45	0.98	1002.58	53064	3.88	85504	6.26
8		307.64	11.30	87.50	15.23	220.14	25.90	-0.34	0.56	4.02	4.49	-0.47	4.29	188.88	4780	0.35	206824	15.14
9		273.60	1.79	113.85	3.77	159.75	4.64	0.74	2.33	-2.35	-1.59	-0.76	2.05	356.05	259144	18.97	2944	0.22
10		258.27	4.82	138.29	7.41	122.55	7.79	0.53	2.28	-3.11	-2.87	-0.24	0.12	1904.37	36752	2.69	17144	1.25
11		272.03	2.62	120.12	4.43	151.90	5.35	0.83	2.31	-1.65	-1.17	-0.48	0.52	1135.10	116716	8.54	58352	4.27
12		279.15	4.74	116.47	7.18	163.04	9.73	0.05	1.92	-0.21	0.16	-0.37	0.10	1576.24	13872	1.02	13324	0.98
13		264.18	1.66	145.97	2.04	118.21	2.39	0.52	3.99	-6.50	-6.03	-0.46	7.73	125.66	129020	9.44	*Does not exist*	
14		341.26	18.34	49.84	42.33	291.42	54.37	-0.14	0.36	5.76	5.91	-0.15	5.31	210.66	68	0.00	121452	8.89
15		288.39	3.61	105.10	6.27	183.29	8.70	0.13	1.45	-0.71	1.88	-2.59	2.25	369.63	21848	1.60	19856	1.45
16		291.12	4.21	99.31	6.85	191.80	9.60	-0.17	1.55	1.53	1.92	-0.40	0.55	1135.02	32472	2.38	17004	1.24
17		277.81	4.48	130.79	3.08	147.02	5.08	0.79	2.86	-1.64	-1.05	-0.59	1.56	322.24	62128	4.55	126096	9.23
18		281.04	2.73	106.50	4.98	174.54	6.48	-0.68	2.42	0.88	1.52	-0.65	0.78	447.81	127440	9.33	83748	6.13
19		281.32	2.52	106.12	4.91	175.20	6.22	0.73	1.11	-0.84	0.01	-0.85	1.11	323.34	142208	10.41	67412	4.93
20		288.91	4.12	94.69	7.32	194.22	10.12	-0.56	1.10	2.33	3.09	-0.76	1.49	377.30	6720	0.49	354452	25.94

Table 6: Spatial correspondence between the 10 quantitatively defined MSTC Ecoregions and the eight dominantly associated Level 2 ecological groups consisting of the 32 ecoregions defined by Nowacki et al. [31].

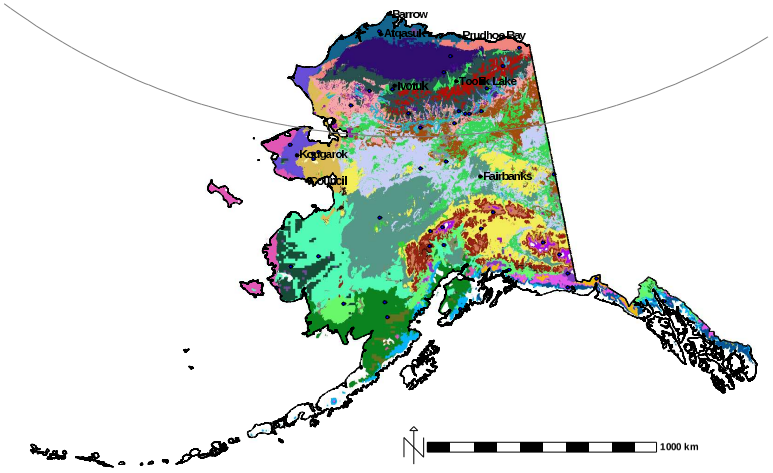
MSTC Ecoregion	**Nowacki Level 2 Ecological Group**	**% area overlap of MSTC on Nowacki**	**% area overlap of Nowacki on MSTC**
1	Coastal Rainforests	85.62	30.83
2	Bering Tundra	58.69	78.77
3	Arctic Tundra	95.75	93.44
4	Bering Taiga	47.66	70.63
5	Intermontane Boreal	78.70	81.58
6	Aleutian Mountains	41.31	22.23
7	Aleutian Mountains	64.18	2.94
8	Coastal Rainforests	96.56	27.46
9	Alaska Range Transition	59.99	35.23
10	Alaska Range Transition	64.38	9.19

Ecoregions defined quantitatively may or may not correspond well to expert-derived ecoregions [39]. Table 6 shows the spatial overlap or correspondence between the 10 quantitatively defined MSTC Ecoregions and the eight dominantly associated Level 2 ecological groups consisting of the 32 ecoregions defined by Nowacki et al. [31]. As expected, strongly distinctive or orographically constrained ecoregions, like Arctic Tundra, have a high degree of correspondence. As shown in Table 6, nearly 96% of MSTC Ecoregion 3 overlaps with the Arctic Tundra Level 2 ecological group defined by Nowacki et al. [31], and 93% of their Arctic Tundra group overlaps with MSTC Ecoregion 3. Meanwhile, MSTC Ecoregion 4 intersects multiple Level 2 ecological groups but most dominantly corresponds to the Bering Taiga group with less than 48% overlap. Because 10 MSTC Ecoregions are intersected with eight Level 2 ecological groups, MSTC Ecoregions appear to subdivide two Level 2 ecological groups and the percent area overlap of MSTC Ecoregions on Level 2 ecological groups is usually larger than the percent area overlap of Level 2 ecological groups on MSTC Ecoregions. A quantitative goodness-of-fit method that explicitly accounts for the degree of spatial correspondence between categorical maps with different numbers of categories [39] can be used to further explore this sort of correspondence analysis.

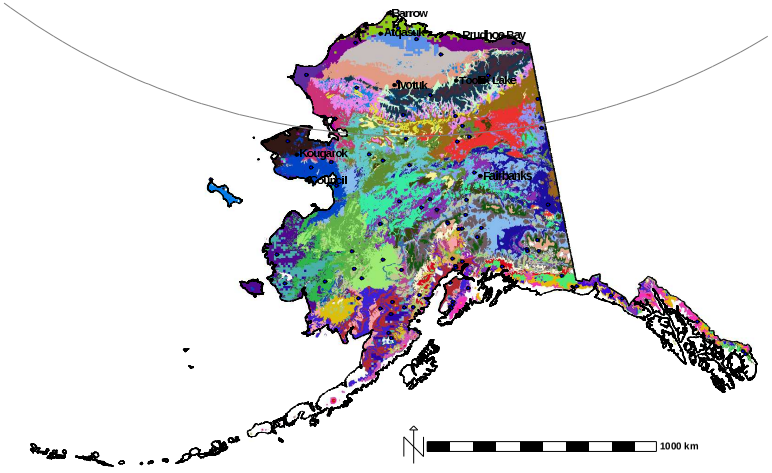
Alaska exhibits wide ranging heterogeneity in environmental conditions, which can be resolved by selecting larger numbers of clusters in the MSTC algorithm. While MSTC is a non-hierarchical procedure, inherently hierarchical relationships within the combinations of state variables automatically emerge when increasing the level of division. For example, at a level of division of $k = 10$, the North Slope of Alaska is represented by a single ecoregion (#3) corresponding to the Arctic Tundra Level 2 ecological group (Figure 3(a)). The North Slope is divided into two ecoregions (#5 and #13) corresponding to the Brooks Range and Beaufort Coastal Plains ecoregions defined by Nowacki et al. [31] at a level of division of $k = 20$ (Figure 3(b)). By further increasing the level of division to $k = 50$, the North Slope



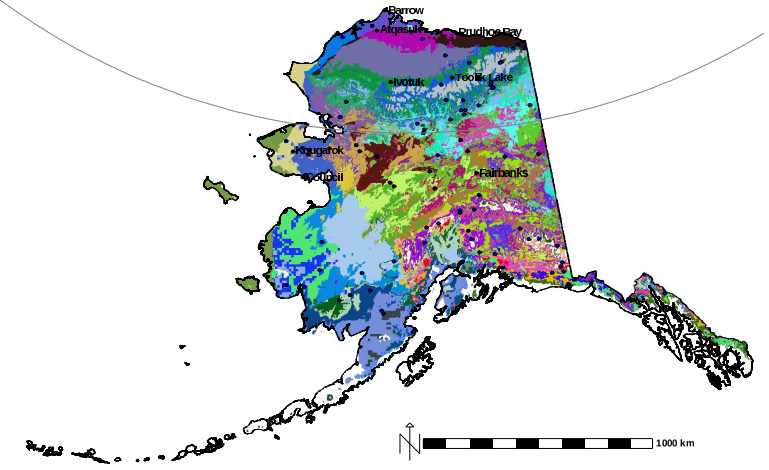
(a) 50 ecoregions, present (2000–2009)



(b) 50 ecoregions, future (2090–2099)

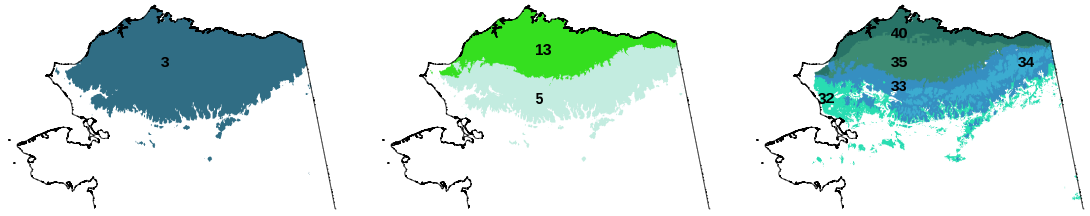


(c) 100 ecoregions, present (2000–2009)



(d) 100 ecoregions, future (2090–2099)

Figure 2: The 50 (a and b) and 100 (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.



(a) At $k = 10$, the North Slope is occupied by MSTC Ecoregion #3, which corresponds to the Arctic Tundra Level 2 ecological group.

(b) At $k = 20$, the North Slope is occupied by MSTC Ecoregion #5, corresponding to the Brooks Range ecoregion; and MSTC Ecoregion #13, corresponding to the Beaufort Coastal Plains ecoregion.

(c) At $k = 50$, the North Slope is occupied by MSTC Ecoregion #32, corresponding to the Intermontane Boreal ecological group; MSTC Ecoregions #33 and #34, corresponding to low- and high-elevation subsets of the Brooks Range ecoregion; MSTC Ecoregion #35, which corresponds to the Brooks Foothills ecoregion; and MSTC Ecoregion #40, which corresponds to the Beaufort Coastal Plains ecoregion.

Figure 3: A hierarchy of increasingly specific ecoregions for the North Slope of Alaska emerge by increasing the level of division in the MSTC algorithm. MSTC cluster numbers are shown and the spatially corresponding Level 2 ecological group or ecoregion defined by Nowacki et al. [31] is identified.

is divided into five different ecoregions (#32, 33, 34, 35, and 40) corresponding to the Intermontane Boreal ecological group, high- and low-elevation Brooks Range, Brooks Foothills, and Beaufort Coastal Plains ecoregions defined by Nowacki et al. [31] (Figure 3(c)). Even more specialized ecoregions can be resolved by further increasing the desired level of division in the MSTC algorithm (Figure 2).

3 Mapping Sensitive Environments

Evidence of environmental change in the Arctic and resulting impacts on aquatic productivity and biodiversity, terrestrial ecosystems, and local economies were highlighted by Anisimov et al. [3]. Increased shrub abundance has been observed in Alaska [40, 41, 42]. During the last 50 years, the tree line along the Arctic to sub-Arctic boundary has moved 10 km northward and 2% of Alaskan tundra on the Seward Peninsula has been replaced by forests. Ecoregions derived for the present and future (Figure 1) show a similar northward shift, indicating a dramatic change in environmental conditions due to a warming climate by the end of this century, as projected by models using the A1B emissions scenario [25]. By tracking changes in the spatial area and migration of ecoregions statistically derived from a hypervolume of environmental gradients [43], this objective approach for mapping landscapes undergoing environmental change can be applied to predict shifts in species ranges

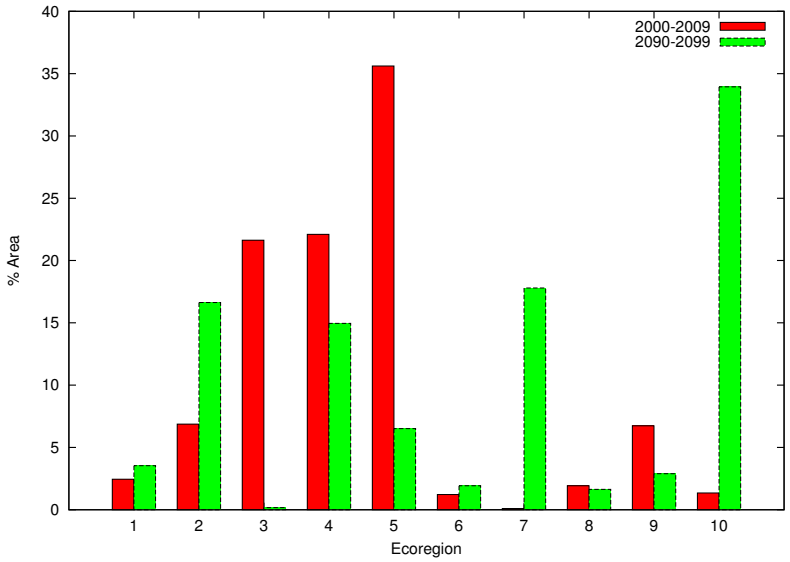


Figure 4: Percent area distribution of 10 ecoregions during the present (2000–2009) and future (2090–2099) periods. Mean values for the state variables of each ecoregion are contained in Tables 2 and 3.

and constrain estimates of changes in the carbon balance of sensitive environments.

Figure 4 shows the percent area distribution of each ecoregion, at the $k = 10$ level of division, for the present and future time periods. Correspondence between these MSTC Ecoregions and Nowacki et al. [31] Level 2 ecological groups is shown in Table 6. A significant decrease in the area of Ecoregion #3, representing most of the North Slope of Alaska as shown in Figure 3(a), is observed. This contemporary Arctic Tundra environment is predicted to be reduced to about 0.78% of its present area by the end of the century. About 76% of the area will be replaced by conditions typical of the warmer Bering Tundra environment (Ecoregion #2). Meanwhile, the Bering Tundra (Ecoregion #2) environment moves northward by the end of the century and more than doubles in areal extent. About 70% of its current area, especially over the Seward Peninsula, will change to conditions similar to contemporary Bering Taiga (Ecoregion #4). In the future, the Bering Taiga (Ecoregion #4) environment decreases in extent by 32% and migrates northward. Under increased temperatures and reduced permafrost conditions, the present-day Aleutian Mountains (Ecoregion #7) environmental conditions are predicted to replace 65% of Bering Taiga (Ecoregion #4), and Alaska Range Transition (Ecoregion #10) environmental conditions are expected to replace 28% of Bering Taiga (Ecoregion #4). Aleutian Mountain (Ecoregion #7) and Alaska Range Transition (Ecoregion #10) environments, which exist in the southern coastal regions of Alaska, are expected to grow in extent northward and occupy a larger portion of Alaska. Alaska Range Transition (Ecoregion #10) environmental conditions are also expected to replace about 75% of the Intermontane Boreal (Ecoregion #5) environment in the future, which will be reduced to 18% of its current area by the end of the century. While similar trends of large scale northward migrations and changes in the areal extents of the environments discussed above are observed at 20 and higher levels of divisions, these ecoregion refinements highlight the changes that are occurring in

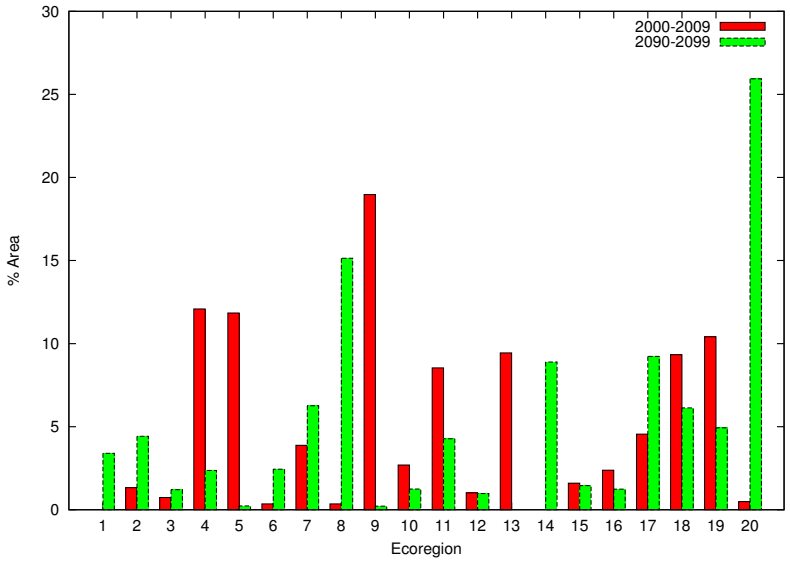
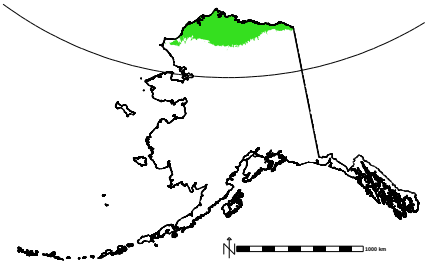


Figure 5: Percent area distribution of 20 ecoregions during the present (2000–2009) and future (2090–2099) periods. Mean values for the state variables of each ecoregion are contained in Tables 4 and 5.

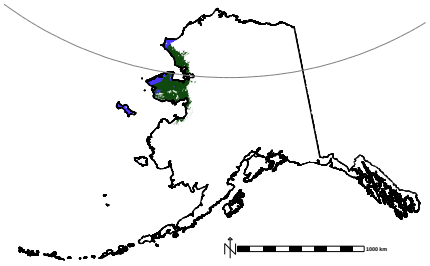
smaller, more uniquely defined environments.

Figure 5 shows the percent area distribution of $k = 20$ ecoregions for the present and future time periods. In addition to areal extent changes and geographic redistribution of ecoregions between the present and future, at this level of division one present-day ecoregion ceases to exist in the future (*i.e.*, becomes extinct) while another ecoregion exists only in the future (*i.e.*, is born) and has no analog in the present. Ecoregion #13 (Figure 6(a)), which represents the most northern portion of Arctic Tundra on the North Slope, becomes extinct in the future due to projected climate change. Ecoregions #2 and #17, which presently occupy the Seward Peninsula and nearby coasts (Figure 6(b)), replace Ecoregion #13 in the future (Figure 6(c)). Approximately 46% of the area of Ecoregion #13 is replaced by Ecoregion #2 and 53% is replaced by Ecoregion #17. Under this climate change scenario, the ecoregions replacing the extinct region in the future have characteristically higher precipitation, higher temperatures, earlier thaw dates, later freeze dates, a longer growing season, increased active layer depth, and higher ground surface temperatures (Tables 4 and 5). At the end of the century, much of the Seward Peninsula and nearby coasts are occupied by an entirely new combination of environmental conditions, defined by Ecoregion #1, which has no analog in the present (Figure 6(d)). This new ecoregion, which appears only in the future time period, represents an environment with higher precipitation and temperature, an increased growing season length, increased active layer depth, and higher soil temperatures (Tables 4 and 5).

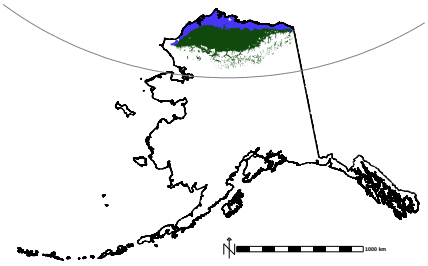
As the level of division is increased in the MSTC algorithm, more specialized ecoregions are delineated. As a result, the number of present-day ecoregions that become extinct and the number of non-analog future ecoregions will both increase. The MSTC procedure was applied for $k = 5$, 10, 20, 50, 100, 200, 500, and 1000 levels of division [38]. Identification of regions representing new combinations of environment conditions that did not previously occur together is important for forecasting



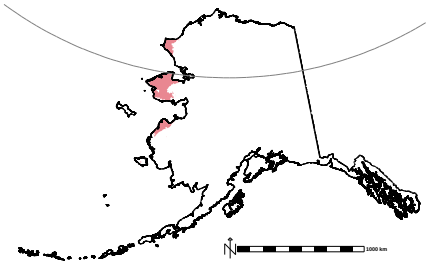
(a) Ecoregion #13 (present), the coldest Arctic Tundra on the North Slope, ceases to exist in the future.



(b) Ecoregions #2 and #17 (present) are limited to the Seward Peninsula and nearby coasts.



(c) Ecoregions #2 and #17 (future) have moved northward, displacing Ecoregion #13 (present) on the North Slope.



(d) A new Ecoregion #1 (future) occupies much of the Seward Peninsula and nearby coasts and has no analog in present.

Figure 6: At $k = 20$, MSTC Ecoregions migrate across the landscape, one becomes extinct, and one comes into existence between the present and future.

species range distributions, conservation planning, and climate change impacts on biodiversity [44].

4 Site Selection

Selection of sampling locations for long term monitoring of ecosystem properties and processes should be guided by an objective, quantitative, systematic, and defensible methodology. Instead, sampling locations in large-scale networks have often been established in opportunistic, political, or logistically-driven ways, resulting in unquantified representation of heterogeneity, biased sampling, uncharacterized uncertainty, and undirected network growth. Finite resources and logistical constraints limit the spatiotemporal frequency and extent of environmental observations, necessitating the development of a systematic sampling strategy to objectively represent environmental variability at the desired spatial scale. An appropriately designed observation strategy should be employed to quantitatively delineate sampling domains, sites, and frequencies. The National Science Foundation's (NSF's) National Ecological Observatory Network (NEON) adopted the objective, data-based methodology described above to define 20 optimal sampling domains across the conterminous United States [7, 8]. Accurate characterization of the landscape and translation of data collected in the field and laboratory into useful datasets, process algorithms, and model parameters requires classification of the landscape into discrete units

based on ecological, hydrological, and geological properties. In much the same way that ecologists develop ecoregions, geologists often classify landscape areas into geomorphological units based on their geophysical and hydrological features. For complex and evolving landscapes featuring interacting vegetation and geomorphological dynamics responding to changes in climate, such as in the Arctic, these stratification concepts may be unified to produce *biogeomorphic units* at relevant spatial scales for landscape characterization, identification of ecological and geomorphological processes, assessing the representativeness of measurements, and providing a framework for scaling measurements and model parameters to larger domains.

An important aspect of site selection and the up- and down-scaling approach to integration of models, observations, and process studies is the estimation of *representativeness*. The MSTC methodology described above for landscape characterization offers useful metrics for indicating the representativeness of sites, measurements, and model parameters. Hargrove et al. [45] described this technique for understanding the representativeness of a sampling network based on a suite of environmental gradients considered to be useful proxies for the characteristics being measured. Maps identifying poorly represented regions can be produced, suggesting where new measurements should be taken to maximize sampling network coverage. As discussed in §2.4, since the cluster centroid represents the mean value of all the state variables in an ecoregion, the realized centroid for an ecoregion is the location that best represents the combination of environmental conditions of the entire ecoregion. Therefore, statistically defined realized centroids, indicated by blue dots in each ecoregion in Figures 1 and 2, are the optimal sampling locations for each ecoregion. Logistical constraints—including accessibility, availability of electric power and telecommunications infrastructure, and geologic stability—may prevent establishment of sampling sites at such optimal locations, particularly in an Arctic environment. Nevertheless, the MSTC Ecoregion framework provides a means for quantifying the representativeness of measurements taken at sub-optimal locations, either within an ecoregion or across any larger domain for which the desired state variables are available.

5 Quantifying Representativeness

While most *in situ* field measurements are made at relatively small, individual geographic points, ecosystem processes operate at many scales. In order 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. A useful representativeness metric is one that can inform the selection of sampling locations, up-scaling of point measurements, down-scaling of remote sensing data, and extrapolation of measurements to unsampled domains. The representativeness metric described by Hargrove et al. [45] provides a unit-less, relative measure of the dissimilarity between the ecoregion of interest, which may contain a sampling site, and any other ecoregion. It is calculated as the Euclidean distance between two ecoregion centroids within the standardized

n-dimensional state space. Ecoregions with similar combinations of environmental conditions will have centroids located near to each other in state space. Therefore, the Euclidean distance between those centroids will be small, representing a low dissimilarity or high representativeness measure. Meanwhile, ecoregions with very different combinations of environmental conditions will have centroids located far from each other in state space, resulting in a large Euclidean distance between them. Such ecoregions will have a high dissimilarity or low representativeness measure. To best capture the natural heterogeneity at the scale of interest, this *ecoregion-based representativeness* should be calculated using MSTC Ecoregions with a large number of divisions (*i.e*, a large value of k).

While Hargrove et al. [45] calculated representativeness in the context of ecoregions, this same approach can be applied to every map cell projected individually onto the n-dimensional state space used to perform the cluster analysis that produced MSTC Ecoregions. This *point-based representativeness* metric captures the full range of heterogeneity in the combinations of environmental conditions, providing a continuously varying measure of dissimilarity for every map cell with respect to a map cell of interest, which may contain a sampling location. When a single ecoregion centroid or map cell of interest is considered, a map of *site representativeness* can be produced. However, multiple ecoregions or map cells of interest may be considered simultaneously, for instance, to provide a quantitative measure of the representativeness of an array or network of sampling sites. The result is a map of *network representativeness* for which the dissimilarity measure for every ecoregion centroid or map cell is the Euclidean distance between that point and the nearest ecoregion centroid or map cell of interest (*i.e.*, the minimum value from a stack of site representativeness maps, one for each ecoregion centroid or map cell containing a measurement site). This representativeness metric, whether ecoregion- or point-based, can be calculated not only between different geographic points in space, but also between different (or the same) geographic points through time. For example, the Euclidean distance between the present combination of environmental conditions and those of the future for any single map cell represents a measure of the magnitude of environmental change over time. Therefore, with this metric it is possible to calculate not only the present-day representativeness of measurements from a site, but also the future representativeness of those present-day measurements, based on future projections of the state variables used in the analysis.

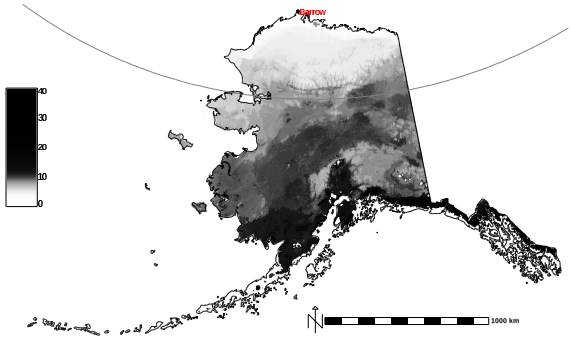
5.1 Site Representativeness

Due to significant logistical constraints when working in the Arctic, a set of eight potential sites were identified as candidates for measurements, long term monitoring and potential manipulative experiments for the U.S. Department of Energy's Next Generation Ecosystem Experiment (NGEE) Arctic project in the State of Alaska: Barrow, Council, Atqasuk, Ivotuk, Kougarok, Prudhoe Bay, Toolik Lake, and Fairbanks. Because of available support infrastructure, Barrow was selected as an initial location for collecting field measurements. To adequately capture the heterogeneity of environmental gradients, an ecoregion-based representativeness analysis employed ecoregion maps at the $k = 1000$ level of division. Figure 7(a) shows the present-day

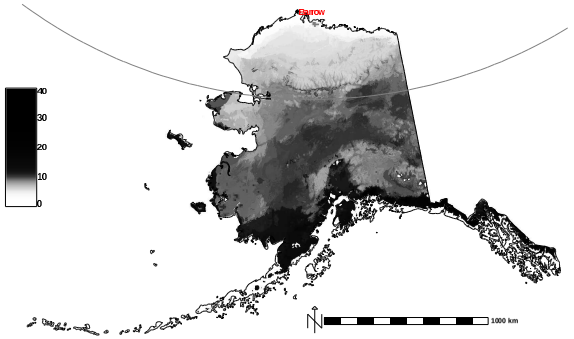
representativeness of the monitoring site at Barrow for the present period. In this map, white to light gray land areas are well-represented by the Barrow location, while dark gray to black land areas are poorly represented by Barrow. The Arctic Tundra of the North Slope is well represented by the Barrow site, but the representativeness drops rapidly at the Brooks Range, which experiences different climate conditions driven by high topography. If a field researcher were attempting to select one additional sampling location in order to provide optimal coverage of the environments within the state of Alaska, that next site should be chosen within the darkest land areas shown in the map. Once a new candidate site has been selected, a new map of representativeness can be generated with simultaneous consideration of both sites. Using this relative representativeness metric, optimal sampling locations can be chosen to maximize the coverage of environmental conditions for any domain at any scale for which sufficient state variable data are available.

Since climate model projections for the future were included in the MSTC procedure, the future representativeness of the present-day Barrow-containing ecoregion can also be mapped (Figure 7(b)). Since the climate is projected to change significantly, the future representativeness of the present-day ecoregion is relatively lower, which is indicated by darker colors in Figure 7(b) as compared with Figure 7(a). Such changes in representativeness are especially large in the Northern Arctic Coastal Plains since this Arctic Tundra is projected to warm significantly and has been identified as a sensitive environment (§3). Similarly, Figures 8(a) and 8(b) contain maps of the present and future representativeness of present-day Barrow, respectively, calculated using the point-based representativeness method. As expected, the large-scale pattern of maps in Figure 8 is the same as that of the maps in Figure 7, but the maps in Figure 8 show more detail and are less generalized than those in Figure 7. Point-based site representativeness maps for each of the eight candidate sites for the present time period are shown in Figure 9.

Since the representativeness metric—or measure of dissimilarity—can be computed between any two map locations, a table quantitatively characterizing dissimilarity of the eight individual candidate sampling locations may be useful for site selection purposes. Table 7 shows point-to-point dissimilarity values for the eight candidate sampling locations for the present time period. Of those locations, Barrow and Fairbanks are the most dissimilar, having a dissimilarity value of 12.16. Atqasuk and Prudhoe Bay are the most similar of the sites. Both Atqasuk and Prudhoe Bay are near-coastal sites at the northern extent of the North Slope; therefore, the environmental conditions are expected to be similar. In addition, according to Table 7, the Prudhoe Bay site is most similar to Barrow, while the Council site is the most dissimilar to Barrow, ignoring Fairbanks. This example analysis suggests that if Barrow were the first sampling site selected, Council may be a strong candidate for a second site in the northern half of the State of Alaska because of its dissimilarity to Barrow. Similarly, Table 8 shows point-to-point dissimilarity values for the eight candidate sampling locations for the future time period. While the dissimilarity values for the future are similar to those of the present, it is apparent that some sites become more similar while others become less similar. For example, Barrow and Council become less dissimilar in the future (*i.e.*, their dissimilarity value of 9.13 in the present changes to 8.87 in the future), indicating that the environmental

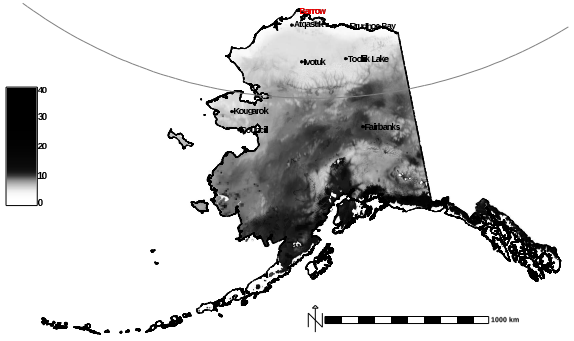


(a) Ecoregion-based representativeness of present-day Barrow for the present period (2000–2009)

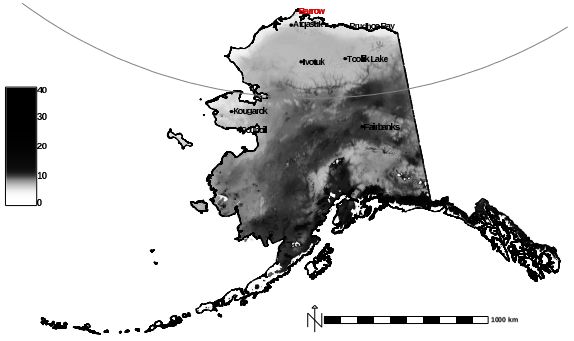


(b) Ecoregion-based representativeness of present-day Barrow for the future period (2090–2099)

Figure 7: Ecoregion-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.

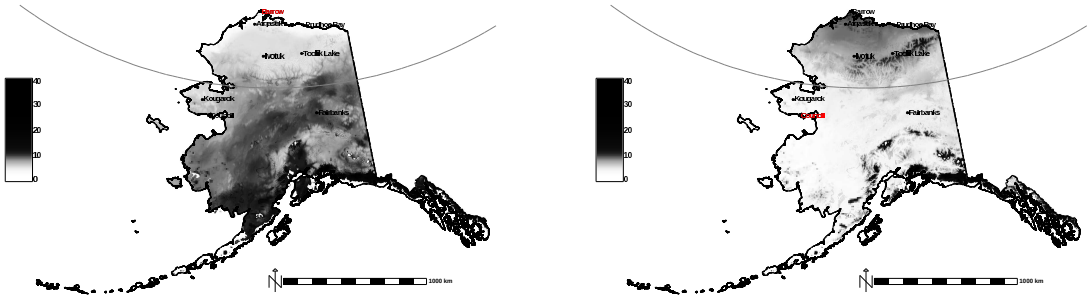


(a) Point-based representativeness of present-day Barrow for the present period (2000–2009)



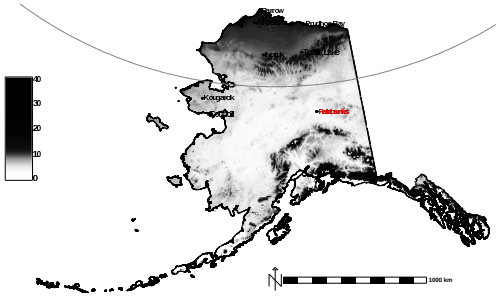
(b) Point-based representativeness of present-day Barrow for the future period (2090–2099)

Figure 8: 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.

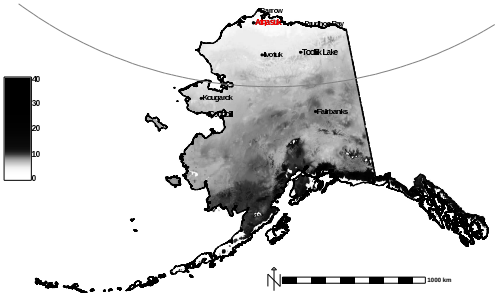


(a) Barrow

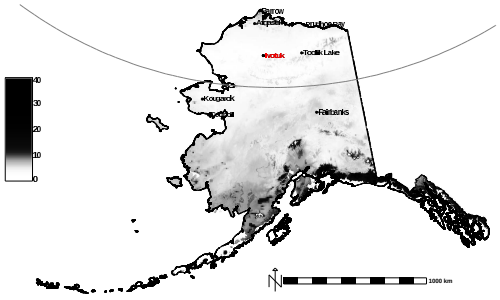
(b) Council



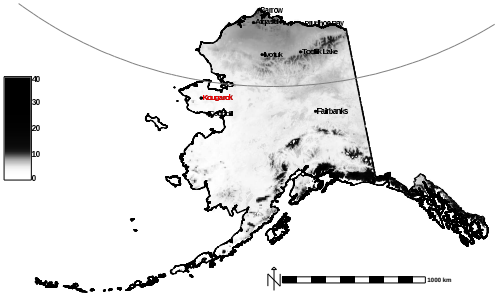
(c) Fairbanks



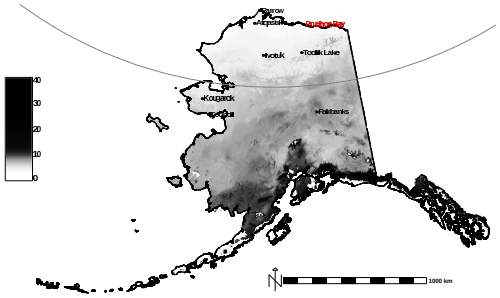
(d) Atqasuk



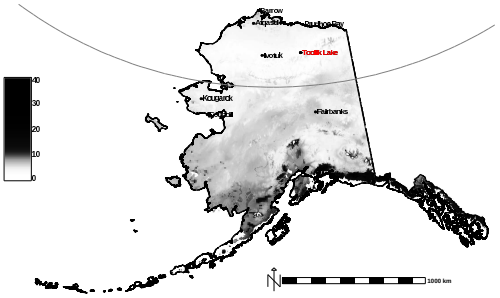
(e) Ivotuk



(f) Kougarok



(g) Prudhoe Bay



(h) Toolik Lake

Figure 9: Point-based representativeness for eight 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.

Table 7: 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

Table 8: 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

conditions in Barrow and Council are more different in the present than they are projected to be in the future.

Table 9 shows a full matrix of point-to-point dissimilarity values for the eight candidate sites between the present and the future. This table quantifies the dissimilarity of present-day sites to those same sites in the future. For this list of widely dispersed locations, the environmental conditions for any single site in the present will be most like the environmental conditions for that same site in the future. Therefore, the smallest dissimilarity values are along the diagonal in Table 9. The largest value on the diagonal is for the Barrow site, indicating that environmental conditions at Barrow are projected to change more than at any other candidate site. In addition, this table shows that environmental conditions at Barrow in the future are more similar to those at Council in the present (8.38) than are the conditions at Barrow in the present to Council in the future (9.67). This result is consistent with the MSTC Ecoregion migration shown in Figure 6. This point-to-point analysis through time is a novel method for quantifying relationships between sampling locations and how those relationships evolve over time due to environmental change.

5.2 Network Representativeness

A monitoring network often consists of a geographically distributed constellation of measurement sites or may be locations where samples are collected for further analysis in the laboratory. Quantifying the representativeness of the network as a

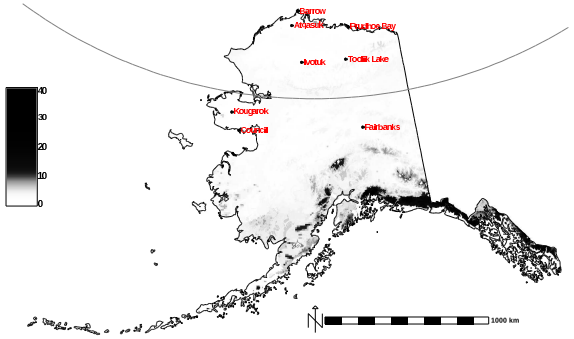
Table 9: Site state space distances between the present (2000–2009) and the future (2090–2099) with DEM

		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

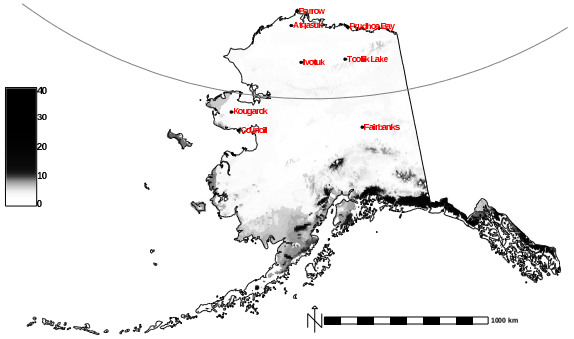
whole is important for optimal network design to avoid unnecessary duplication and to maximize the coverage of the monitoring network. By combining multiple maps of site representativeness for every sampling location, and calculating the minimum value for every map cell, maps of network representativeness are produced. Figures 10(a) and 10(b) contain maps of ecoregion-based network representativeness for all eight candidate sampling sites for the present and future time periods, respectively. Similarly, Figures 11(a) and 11(b) contain maps of point-based network representativeness for the same eight candidate sampling sites for the present and future time periods, respectively. 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. If the objective were to maximize the coverage of all environments in the State of Alaska, the next sampling location should be chosen within the darkest land areas shown in the map. Most of Alaska is well represented by this network of eight sampling locations.

6 Conclusions

Systematic sampling strategies are essential for understanding ecosystem responses to climate change and informing model development. In the harsh Arctic environment—where climate change appears to be most rapidly affecting sensitive ecosystems and vulnerable, carbon-rich permafrost—filling critical gaps in observations is expensive and technically challenging. To fully explore the regional and global implications of climate change in the Arctic, global Earth System Models must capture the important processes and feedbacks. Such models must be developed based on a rich body of observational data as representative as possible of multiple spatial and temporal scales. Meanwhile, finite resources and logistical constraints place restrictions on the number of sampling sites, spatial extent, frequency, and types of measurements that can be collected. This study proposes a quantitative, data-based methodology for stratifying sampling domains, informing site selection, and determining the representativeness of measurement sites and

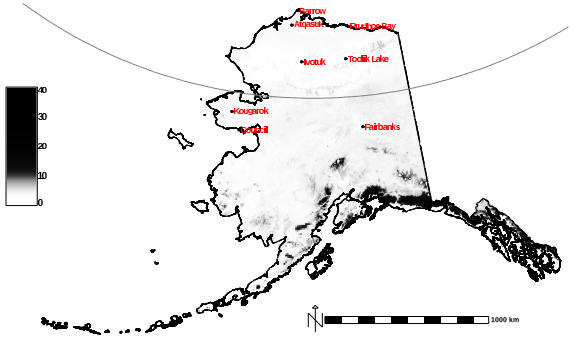


(a) Ecoregion-based network representativeness of eight sites for the present period

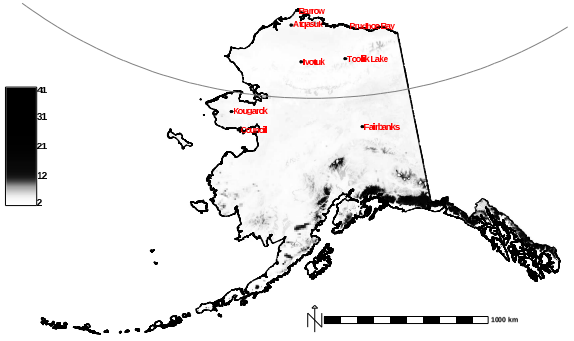


(b) Ecoregion-based network representativeness of eight sites for the future period

Figure 10: 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.



(a) Point-based network representativeness of eight sites for the present period



(b) Point-based network representativeness of eight sites for the future period

Figure 11: 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.

sampling networks.

Multivariate spatiotemporal clustering (MSTC), based on k-means cluster analysis, was applied to down-scaled general circulation model (GCM) results and observational data for the State of Alaska at a nominal resolution of 4 km^2 to define a set of ecoregions at multiple levels of division across two decadal time periods. Maps of ecoregions for the present (2000–2009) and future (2090–2099) were produced, showing how combinations of 37 environmental conditions are distributed across Alaska and how these combinations shift as a result projected climate change in the 21st century. Using this statistical approach, optimal sampling locations, called realized centroids, were identified for each ecoregion at every level of division. In addition, the resulting geographic shifts and changes in areal distribution of ecoregions suggested that some environments may disappear, many will be redistributed, and new ones will appear in the coming century. This analysis provides insights into the identification of the most sensitive and potentially vulnerable Arctic ecosystems. The Euclidean distance within the 37-dimensional state space used for MSTC provides a metric for representativeness. Gray-scale maps of representativeness, showing the similarity of every map cell to a list of eight candidate samples locations near town sites in Alaska, were produced for each site. Tables quantitatively characterizing the similarity of candidate sampling locations to each other across space and through time were generated. These tables are useful for understanding the strength of the environmental gradients between sites and how those gradients may change based on model projections of the future. Taken together, these analysis products provide model-inspired insights into optimal sampling strategies across space and through time, and these same techniques can be applied at different spatial and temporal scales to meet the needs of individual measurement or monitoring campaigns.

The representativeness of a sampling network is best maximized before the network is deployed. Even if additional "optimized" sites are added to an existing network, it will require many more additions to approach the theoretical maximum representativeness for a given number of initial sites. It is difficult, with only the sequential addition of new optimized sites, to achieve the same representativeness once some sampling sites have been established. Representativeness resulting from such network "repairs" rarely ever equal the representativeness of a network initially designed *de novo* with that same number of sampling sites. Even if the network is to be constructed in stages, it is best to design site placement using the final, ultimate complement of sites and to operate sub-optimally until the full network can be completed. Otherwise, many more sites will have to be added to the existing network in order to achieve the same representativeness than could otherwise have been designed in initially.

Cluster analysis and n-dimensional data space regressions offer quantitative methods for up-scaling and extrapolating measurements to land areas within and beyond the sampling domain and provide a down-scaling approach to the integration of models, observations, and process studies. The accuracy of the up-scaled data will be higher for areas represented well by the monitoring network and lower for areas that are poorly represented. At a large scale, these techniques are useful for delineating distinct, broad regions and optimal measurement sites. However, this methodology can also be applied at finer spatiotemporal scales, with inclusion of

other geophysical characteristics and remote sensing data, to inform measurement frequency and site selection within these broader ecoregions.

Acknowledgments

This research was sponsored by the Climate and Environmental Sciences Division (CESD) of the Office of Biological and Environmental Research (BER) within the U.S. Department of Energy (DOE) Office of Science. Additional support was provided by the U.S. Department of Agriculture (USDA) Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC). This research used resources of the Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

References

[1] Edward A. G. Schuur, James Bockheim, Josep G. Canadell, Eugenie Euskirchen, Christopher B. Field, Sergey V. Goryachkin, Stefan Hagemann, Peter Kuhry, Peter M. Lafleur, Hanna Lee, Galina Mazhitova, Frederick E. Nelson, Annette Rinke, Vladimir E. Romanovsky, Nikolay Shiklomanov, Charles Tarnocai, Sergey Venevsky, Jason G. Vogel, and Sergei A. Zimov. Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle. *Bioscience*, 58(8):701–714, September 2008. ISSN 0006-3568. doi:10.1641/B580807.

[2] IPCC. Summary for Policymakers. In Susan Solomon, Dahe Qin, Martin Manning, Zhenlin Chen, Melinda Marquis, Kristen B. Averyt, Melinda Tignor, and Henry L. Miller, editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, United Kingdom and New York, NY, USA, 2007. Cambridge University Press. ISBN 978-0-521-88009-1 hardback; 978-0-521-70596-7 paperback.

[3] O. A. Anisimov, D. G. Vaughan, T. V. Callaghan, C. Furgal, H. Marchant, T. D. Prowse, H. Vilhjálmsson, and J. E. Walsh. Polar regions (Arctic and Antarctic). In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, editors, *Climate Change 2007: Impacts, Adaptation and Vulnerability*, pages 653–685. Cambridge University Press, Cambridge, 2007.

[4] Larry Hinzman, Neil Bettez, W. Bolton, F. Chapin, Mark Dyurgerov, Chris Fastie, Brad Griffith, Robert Hollister, Allen Hope, Henry Huntington, Anne

Jensen, Gensuo Jia, Torre Jorgenson, Douglas Kane, David Klein, Gary Kofinas, Amanda Lynch, Andrea Lloyd, A. McGuire, Frederick Nelson, Walter Oechel, Thomas Osterkamp, Charles Racine, Vladimir Romanovsky, Robert Stone, Douglas Stow, Matthew Sturm, Craig Tweedie, George Vourlitis, Marilyn Walker, Donald Walker, Patrick Webber, Jeffrey Welker, Kevin Winker, and Kenji Yoshikawa. Evidence and implications of recent climate change in Northern Alaska and other Arctic regions. *Clim. Change*, 72(3):251–298, 2005. ISSN 0165-0009. doi:10.1007/s10584-005-5352-2.

[5] The Arctic Cimate Impact Assessment (ACIA). *Arctic Climate Impact Assessment*. Cambridge University Press, 2005. ISBN 9780521865098.

[6] National Research Council Committee on Designing an Arctic Observing Network. *Towards an Integrated Arctic Observing Network*. The National Academies Press, 2006. ISBN 9780309100526.

[7] Michael Keller, David Schimel, William Hargrove, and Forrest 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.

[8] David Schimel, William Hargrove, Forrest Hoffman, and James McMahon. NEON: A hierarchically designed national ecological network. *Front. Ecol. Environ.*, 5(2):59, March 2007. doi:10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2.

[9] James M. Omernik. Ecoregion of the conterminuous United States. *An. Assoc. Amer. Geog.*, 77(1):118–125, 1987. doi:10.1111/j.1467-8306.1987.tb00149.x.

[10] David M. Olson and Eric Dinerstein. The global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89(2):199–224, April 2002. ISSN 00266493.

[11] Robert G. Bailey and Howard C. Hogg. A world ecoregions map for resource reporting. *Environ. Conserv.*, 13(3):195–202, September 1986. doi:10.1017/S0376892900036237.

[12] Robert G. Bailey. Ecoregions of the United States. In *Ecosystem Geography*, Statistics for Social and Behavioral Sciences, pages 93–114. Springer New York, 2009. ISBN 978-0-387-89516-1. doi:10.1007/978-0-387-89516-1_7.

[13] William W. Hargrove and Forrest M. Hoffman. Using multivariate clustering to characterize ecoregion borders. *Comput. Sci. Eng.*, 1(4):18–25, July 1999. doi:10.1109/5992.774837.

[14] 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.

[15] Forrest M. Hoffman, William W. Hargrove, David J. Erickson, and Robert J. Oglesby. Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models. *Earth Interact.*, 9(10):1–27, August 2005. doi:10.1175/EI110.1.

[16] William B. Krohn, Randall B. Boone, and Stephanie L. Painton. Quantitative delineation and characterization of hierarchical biophysical regions of Maine. *Northeastern Naturalist*, 6(2):139–164, 1999.

[17] M. E. Jensen, I. A. Goodman, P. S. Bourgeron, N. L. Poff, and C. K. Brewer. Effectiveness of biophysical criteria in the hierarchical classification of drainage basins. *J. Am. Water Resour. Assoc.*, 37:1155–1167, 2001.

[18] J. A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, 1975.

[19] Paul S. Bradley and Usama M. Fayyad. Refining initial points for k-means clustering. In *ICML '98: Proceedings of the Fifteenth International Conference on Machine Learning*, pages 91–99, San Francisco, CA, USA, July 1998. Morgan Kaufmann Publishers Inc. ISBN 1-55860-556-8.

[20] 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.

[21] William W. Hargrove, Forrest M. Hoffman, and Thomas Sterling. The do-it-yourself supercomputer. *Sci. Am.*, 265(2):72–79, August 2001.

[22] Gnanamanika Mahinthakumar, Forrest M. Hoffman, William W. Hargrove, and Nicolas T. Karonis. Multivariate geographic clustering in a metacomputing environment using Globus. In *Supercomputing '99: Proceedings of the 1999 ACM/IEEE conference on Supercomputing (CDROM)*, Supercomputing '99, New York, NY, USA, November 1999. ACM Press. ISBN 1-58113-091-0. doi: 10.1145/331532.331537.

[23] Forrest M. Hoffman, William W. Hargrove, Richard T. Mills, Salil Mahajan, David J. Erickson, and Robert J. Oglesby. Multivariate Spatio-Temporal Clustering (MSTC) as a data mining tool for environmental applications. In Miquel Sànchez-Marrè, Javier Béjar, Joaquim Comas, Andrea E. Rizzoli, and Giorgio Guariso, editors, *Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software Society (iEMSs 2008)*, pages 1774–1781, July 2008. ISBN 978-84-7653-074-0.

[24] Jitendra Kumar, Richard Tran Mills, Forrest M. Hoffman, and William W. Hargrove. Parallel k-means clustering for quantitative ecoregion delineation using large data sets. In Mitsuhisa Sato, Satoshi Matsuoka, Peter M. Sloot,

G. Dick van Albada, and Jack Dongarra, editors, *Proceedings of the International Conference on Computational Science (ICCS 2011)*, volume 4 of *Procedia Comput. Sci.*, pages 1602–1611. Elsevier, Amsterdam, June 2011. doi: 10.1016/j.procs.2011.04.173.

[25] Nebojša Nakićenović, Joseph Alcamo, Gerald Davis, Bert de Vries, Joergen Fenhann, Stuart Gaffin, Kenneth Gregory, Arnulf Grübler, Tae Yong Jung, Tom Kram, Emilio Lebre La Rovere, Laurie Michaelis, Shunsuke Mori, Tsuneyuki Morita, William Pepper, Hugh Pitcher, Lynn Price, Keywan Riahi, Alexander Roehrl, Hans-Holger Rogner, Alexei Sankovski, Michael Schlesinger, Priyadarshi Shukla, Steven Smith, Robert Swart, Sascha van Rooijen, Nadejda Victor, and Zhou Dadi. Special report on emissions scenarios. In Nebojša Nakićenović and Robert Swart, editors, *A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, page 570. Cambridge University Press, Cambridge, United Kingdom, July 2000. ISBN 92-9169-113-5.

[26] 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.

[27] Vladimir E. Romanovsky and Sergei Marchenko. The GIPL permafrost dynamics model. Technical report, University of Alaska, Fairbanks, Alaska, May 2009.

[28] Christopher D. Arp and Benjamin M. Jones. Geography of Alaska lake districts: Identification, description, and analysis of lake-rich regions of a diverse and dynamic state. Scientific Investigations Report 2008-5215, U.S. Geological Survey, 4210 University Dr., Anchorage, Alaska 99508, January 2009.

[29] Gregory Nowacki and Terry Brock. Ecoregions and Subregions of Alaska, EcoMap Version 2.0. map, USDA Forest Service, Alaska Region, Juneau, AK, 1995. Scale 1:5,000,000.

[30] A. L. Gallant, E. F. Binnian, J. M. Omernik, and M. B. Shasby. Ecoregions of Alaska. Professional paper 1567, U.S. Geological Survey, 1995.

[31] Gregory Nowacki, Page Spencer, Michael Fleming, Terry Brock, and Torre Jorgenson. Ecoregions of Alaska: 2001. Open-file report 02-297 (map), U.S. Geological Survey, 2001.

[32] Berman D. Hudson. The soil survey as paradigm-based science. *Soil Sci. Soc. Am. J.*, 56(3):836–841, 1992. doi:10.2136/sssaj1992.03615995005600030027x.

[33] Yingchun Zhou. An ecological regionalization model based on NOAA/AVHRR data. *International Archives of Photogrammetry and Remote Sensing*, XXXI, Part B4:1001–1006, 1996.

[34] Gerard McMahon, Steven M. Gregonis, Sharan W. Waltman, James M. Omernik, Thor D. Thorson, Jerry A. Freeouf, Andrew H. Rorick, and James E. Keys. Developing a spatial framework of common ecological regions for the conterminous united states. *Environ. Manage.*, 28(3):293–316, 2001. ISSN 0364-152X. doi:10.1007/s0026702429.

[35] Earl Saxon, Barry Baker, William Hargrove, Forrest Hoffman, and Chris Zganjar. Mapping environments at risk under different global climate change scenarios. *Ecol. Lett.*, 8:53–60, 2005. doi:10.1111/j.1461-0248.2004.00694.

[36] A. David McGuire, Leif G. Anderson, Torben R. Christensen, Scott Dallimore, Laodong Guo, Daniel J. Hayes, Martin Heimann, Thomas D. Lorenson, Robie W. Macdonald, and Nigel Roulet. Sensitivity of the carbon cycle in the Arctic to climate change. *Ecol. Monogr.*, 79(4):523–553, November 2009. doi: 10.1890/08-2025.1.

[37] F. S. Chapin, A. D. McGuire, R. W. Ruess, T. N. Hollingsworth, M. C. Mack, J. F. Johnstone, E. S. Kasischke, E. S. Euskirchen, J. B. Jones, M. T. Jorgenson, K. Kielland, G. P. Kofinas, M. R. Turetsky, J. Yarie, A. H. Lloyd, and D. L. Taylor. Resilience of Alaska's boreal forest to climatic change. *Can. J. Forest Res.*, 40(7):1360–1370, July 2010. doi:10.1139/X10-074.

[38] Forrest M. Hoffman, Jitendra Kumar, Richard T. Mills, William W. Hargrove, Peter E. Thornton, and Stan D. Wullschlegar. A geospatiotemporal analysis for site selection for the Next Generation Ecosystem Experiment (NGEE) Arctic project. Technical Memorandum ORNL/TM-XXXXX, Oak Ridge National Laboratory, 2012.

[39] William W. Hargrove, Forrest M. Hoffman, and Paul 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.

[40] Matthew Sturm, Charles Rachine, and Kenneth Tape. Climate change: Increasing shrub abundance in the Arctic. *Nature*, 411(6837):546–547, May 2001. doi:10.1038/35079180.

[41] Matthew Sturm, Tom Douglas, Charles Racine, and Glen E. Liston. Changing snow and shrub conditions affect albedo with global implications. *J. Geophys. Res.*, 110(G1):G01004, September 2005. ISSN 0148-0227. doi: 10.1029/2005JG000013.

[42] Ken Tape, Matthew Sturm, and Charles Racine. The evidence for shrub expansion in Northern Alaska and the Pan-Arctic. *Global Change Biol.*, 12(4): 686–702, 2006. ISSN 1365-2486. doi:10.1111/j.1365-2486.2006.01128.x.

[43] G. E. Hutchinson. Concluding remarks. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 22, pages 415–427, 1957. Reprinted in 1991: Classics in Theoretical Biology, *Bull. Math. Biol.* 53:193–213.

[44] Matthew Fitzpatrick and William Hargrove. The projection of species distribution models and the problem of non-analog climate. *Biodivers. Conserv.*, 18 (8):2255–2261, July 2009. doi:10.1007/s10531-009-9584-8.

[45] William W. Hargrove, Forrest M. Hoffman, and Beverly E. Law. New analysis reveals representativeness of the AmeriFlux Network. *Eos Trans. AGU*, 84(48): 529, 535, December 2003. doi:10.1029/2003EO480001.