Measurements to Models: ILAMB and Representativeness/Scaling

Forrest M. Hoffman

James T. Randerson, Jitendra Kumar, William W. Hargrove, and Richard T. Mills

Computational Earth Sciences Group & Climate Change Science Institute (CCSI),
Oak Ridge National Laboratory

R_H Workshop

H. J. Andrews Experimental Forest, Blue River, Oregon
October 30–November 1, 2012

ILAMB Goals

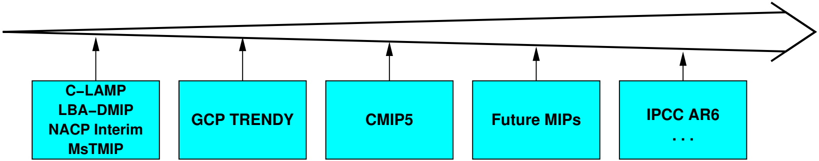
The goals of the International Land Model Benchmarking (ILAMB) project are to:

- Develop benchmarks for land model performance, with a focus on carbon cycle, ecosystem, surface energy, and hydrological processes. The benchmarks should be designed and accepted by the community.
- Apply the benchmarks to global models.
- Support the design and development of a new, open-source, benchmarking software system for either diagnostic or model intercomparison purposes.
- Strengthen linkages between experimental, monitoring, remote sensing, and climate modeling communities in the design of new model tests and new measurement programs.

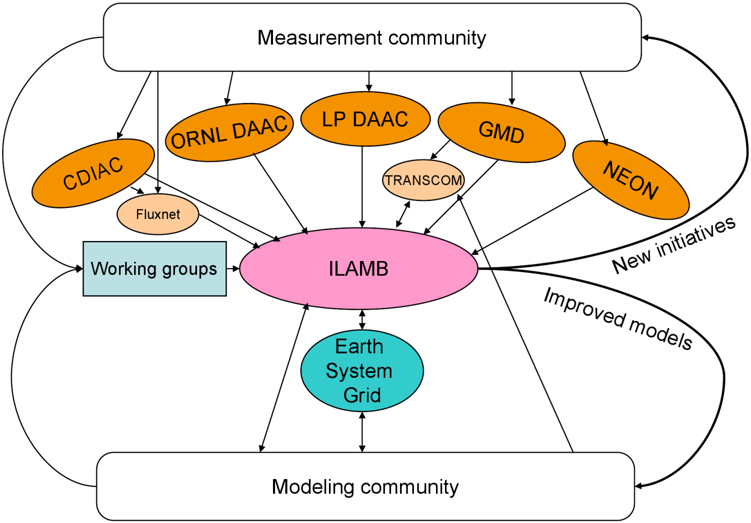
Why Benchmark?

- to show the broader science community and the public that the representation of the carbon cycle in climate models is improving;
- to provide a means, in Earth System models, to quantitatively diagnose impacts of model development in related fields on carbon cycle and land surface processes;
- to guide synthesis efforts, such as the Intergovernmental Panel on Climate Change (IPCC), in the review of mechanisms of global change in models that are broadly consistent with available contemporary observations;
- to increase scrutiny of key datasets used for model evaluation;
- to identify gaps in existing observations needed for model validation;
- to provide a quantitative, application-specific set of minimum criteria for participation in model intercomparison projects (MIPs);
- to provide an optional weighting system for multi-model mean estimates of future changes in the carbon cycle.

An Open Source Benchmarking Software System



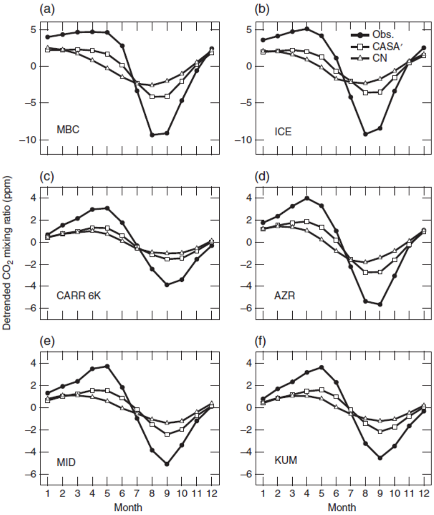
- Human capital costs of making rigorous model-data comparisons is considerable and constrains the scope of individual MIPs.
- Many MIPs spend resources "reinventing the wheel" in terms of variable naming conventions, model simulation protocols, and analysis software.
- **Need for ILAMB:** Each new MIP has access to the model-data comparison modules from past MIPs through ILAMB (*e.g.*, MIPs use one common modular software system). Standardized international naming conventions also increase MIP efficiency.



International Land Model Benchmarking project and diagnostic system

What is a Benchmark?

- A benchmark is a quantitative test of model function, for which the uncertainties associated with the observations can be quantified.
- Acceptable performance on benchmarks **is a necessary but not sufficient condition** for a fully functioning model.
- Since all datasets have strengths and weaknesses, an effective benchmark is one that draws upon a broad set of independent observations to evaluate model performance on multiple temporal and spatial scales.



From Randerson et al. (2009)

Temperature Dependence of Heterotrophic Respiration

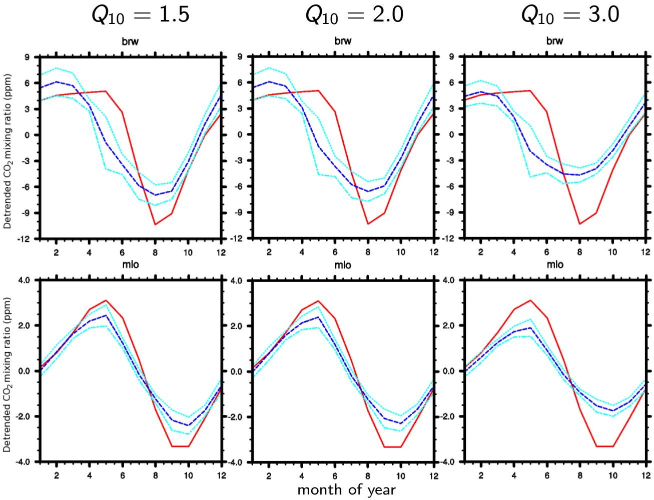
$$R_h = C_{bg}\, k_0\, Q_{10}^{\left(\frac{T - T_0}{10}\right)}$$

Barrow (BRW)

71.32°N, -156.61°E

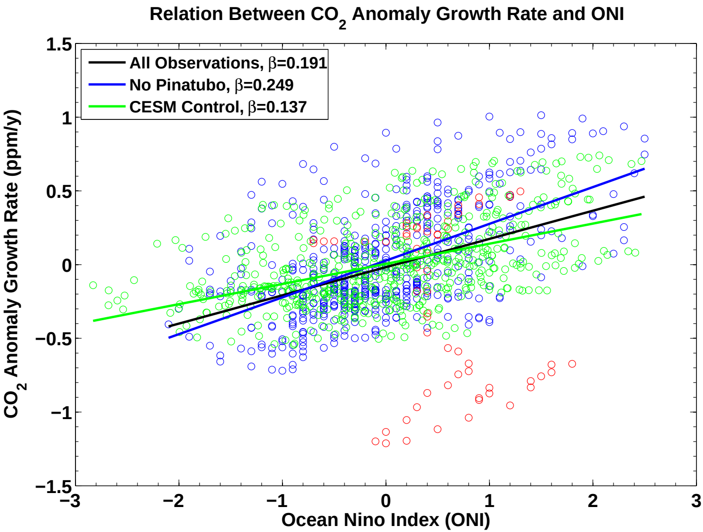
Mauna Loa (MLO)

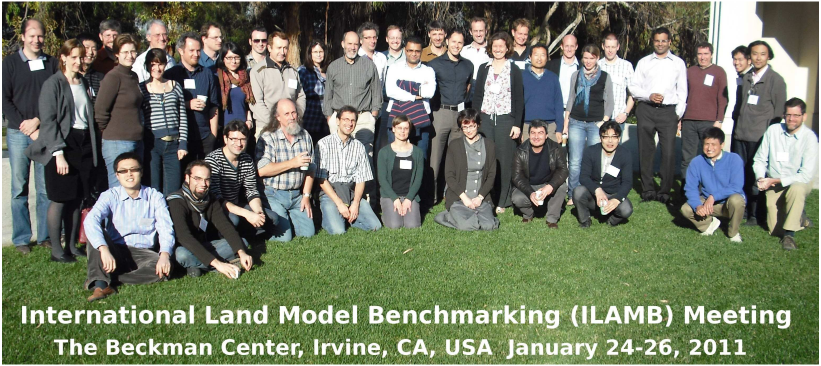
19.54°N, -155.58°E



GLOBALVIEW-CO2/TRANSCOM impulse response function (TIRF), CLM4/Mean TIRF, CLM4/TIRF bounds

CESM vs. Mauna Loa CO_2 Observations





- Meeting Co-organized by Forrest Hoffman (UC-Irvine and ORNL), Chris Jones (UK Met Office Hadley Centre), Pierre Friedlingstein (U. Exeter), and Jim Randerson (UC-Irvine).
- About 45 researchers participated from the United States, Canada, the United Kingdom, the Netherlands, France, Germany, Switzerland, China, Japan, and Australia.

ILAMB Meeting Goals

- Design the first set of ILAMB benchmarks for global models.
  - How many flavors (carbon cycle, LUC, hydrology, . . .)?
  - What datasets do we include?
  - What graphics and cost functions?
- Coordinate carbon cycle and land model evaluation analyses for TRENDY and CMIP5 results.
- Develop an implementation plan for application of the ILAMB 1.0 benchmarks to TRENDY and CMIP5 output over next year.
- Decide upon the approach for developing ILAMB code.
  - netCDF for datasets? Language for evaluation code?
  - Need to extend variable naming conventions beyond CMIP5.
- Decide upon a future schedule and means to secure funding.
  - Key deadline is July 2012 for submission of manuscripts for IPCC AR5 Working Group 1.
  - Should ILAMB meet once a year until AR6?

Example Benchmark Score Sheet from C-LAMP

Models →

BGC Datasets ↓

Metric	Metric components	Uncertainty of obs.	Scaling mismatch	Total score	Sub-score	CASA′		CN	
LAI	Matching MODIS observations			15.0		13.5		12.0	
	• Phase (assessed using the month of maximum LAI)	Low	Low		6.0		5.1		4.2
	• Maximum (derived separately for major biome classes)	Moderate	Low		5.0		4.6		4.3
	• Mean (derived separately for major biome classes)	Moderate	Low		4.0		3.8		3.5
NPP	Comparisons with field observations and satellite products			10.0		8.0		8.2	
	• Matching EMDI Net Primary Production observations	High	High		2.0		1.5		1.6
	• EMDI comparison, normalized by precipitation	Moderate	Moderate		4.0		3.0		3.4
	• Correlation with MODIS (r^2)	High	Low		2.0		1.6		1.4
	• Latitudinal profile comparison with MODIS (r^2)	High	Low		2.0		1.9		1.8
CO_2 annual cycle	Matching phase and amplitude at Globalview flash sites			15.0		10.4		7.7	
	• 60°–90°N	Low	Low		6.0		4.1		2.8
	• 30°–60°N	Low	Low		6.0		4.2		3.2
	• 0°–30°N	Moderate	Low		3.0		2.1		1.7
Energy & CO_2 fluxes	Matching eddy covariance monthly mean observations			30.0		17.2		16.6	
	• Net ecosystem exchange	Low	High		6.0		2.5		2.1
	• Gross primary production	Moderate	Moderate		6.0		3.4		3.5
	• Latent heat	Low	Moderate		9.0		6.4		6.4
	• Sensible heat	Low	Moderate		9.0		4.9		4.6
Transient dynamics	Evaluating model processes that regulate carbon exchange on decadal to century timescales			30.0		16.8		13.8	
	• Aboveground live biomass within the Amazon Basin	Moderate	Moderate		10.0		5.3		5.0
	• Sensitivity of NPP to elevated levels of CO_2: comparison to temperate forest FACE sites	Low	Moderate		10.0		7.9		4.1
	• Interannual variability of global carbon fluxes: comparison with TRANSCOM	High	Low		5.0		3.6		3.0
	• Regional and global fire emissions: comparison to GFEDv2	High	Low		5.0		0.0		1.7
			Total:	**100.0**		**65.9**		**58.3**	

From Randerson et al. (2009)

	Annual Mean	Seasonal Cycle	Interannual Variability	Trend	Data Source
Atmospheric CO_2					
Flask/conc. + transport		✓	✓	✓	NOAA, SIO, CSIRO
TCCON + transport		✓	✓	✓	Caltech
Fluxnet					
GPP, NEE, TER, LE, H, RN	✓	✓	✓		Fluxnet, MAST-DC
Gridded: GPP	✓	✓	?		MPI-BGC
Hydrology/Energy					
river flow	✓		✓		GRDC, Dai, GFDL
global runoff/ocean balance	✓				Syed/Famiglietti
albedo (multi-band)		✓	✓		MODIS, CERES
soil moisture		✓	✓		de Jeur, SMAP
column water		✓	✓		GRACE
snow cover	✓	✓	✓	✓	AVHRR, GlobSnow
snow depth/SWE	✓	✓	✓	✓	CMC (N. America)
T_{air} & P	✓	✓	✓	✓	CRU, GPCP and TRMM
Gridded: LE, H	✓	✓			MPI-BGC, dedicated ET
Ecosystem Processes & State					
soil C, N	✓				HWSD, MPI-BGC
litter C, N	✓				LIDET
soil respiration	✓	✓	✓	✓	Bond-Lamberty
FAPAR	✓	✓			MODIS, SeaWIFS
biomass & change	✓			✓	Saatchi, Pan, Blackard
canopy height	✓				Lefsky, Fisher
NPP	✓				EMDI, Luyssaert
Vegetation Dynamics					
fire — burned area	✓	✓	✓		GFED3
wood harvest	✓			✓	Hurtt
land cover	✓				MODIS PFT fraction

Meeting Summary

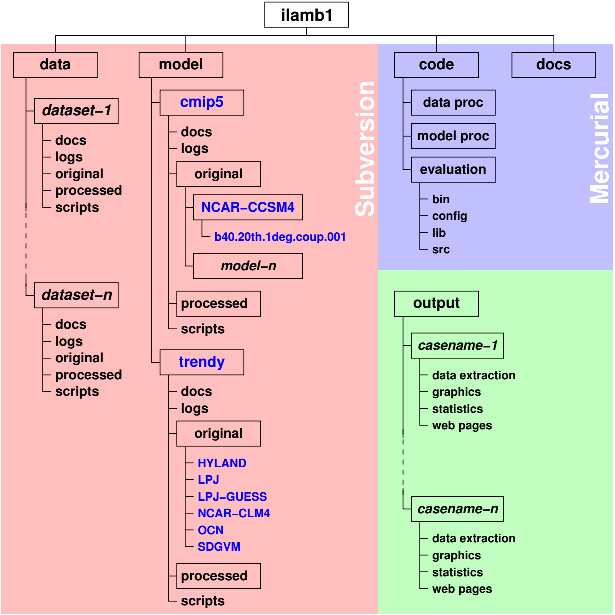
- Five break-out groups met, one for each benchmark category, to identify cost function metrics and graphics.
- Measurement and model uncertainty must be characterized and spatial scaling mismatch considered for evaluation.

- Key objectives are to use publicly available data and freely available software.
- The R package will be used for generating statistical results and diagnostics.
- Five initial benchmarks will be implemented to evaluate existing TRENDY and CMIP5 model results.

A team was identified to begin software architecture design.

A developmental hierarchy for data, model results, code, and docs is established.

Server-based and distributed version control systems will be used for handling data and code, respectively.



Next Steps

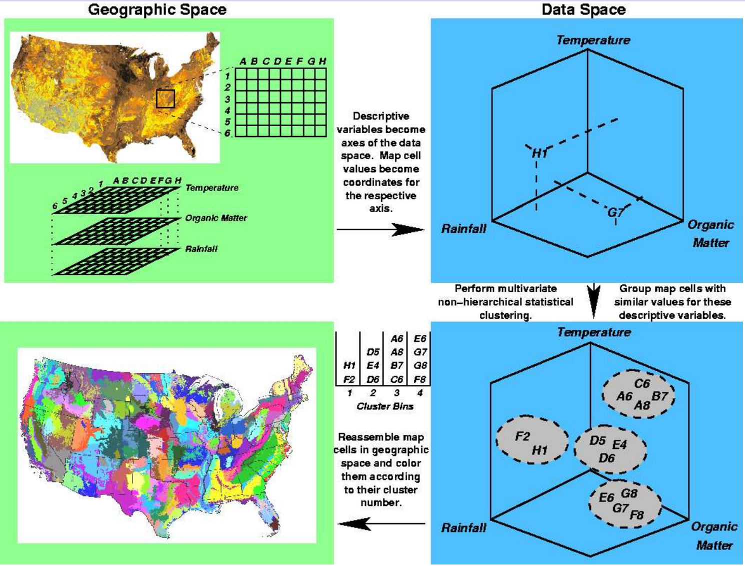
- Common model output
  - A draft document proposing additional new netCDF Climate and Forecast (CF) conventions, beyond those created for CMIP5, is available for comment.
  - To assist the modeling community, a translator between ALMA and CF standards may be created.
- Future: New protocols and forcing data comparisons?
- An ILAMB Meeting was held at the AGU Fall Meeting last December.
- Future meetings are planned after IPCC publication deadlines have passed.

International Land Model Benchmarking (ILAMB) Project
http://www.ilamb.org/

Representativeness and Scaling

- Resource and logistical constraints limit the frequency and extent of observations, necessitating the development of a systematic sampling strategy that objectively represents environmental variability at the desired spatial scale.
- Required is a methodology that provides a quantitative framework for informing site selection and determining the representativeness of measurements.
- Multivariate spatiotemporal clustering (MSTC) was applied at the landscape scale (4 km^2) for the State of Alaska to demonstrate its utility for representativeness and scaling.
- An extension of the method applied by Hargrove and Hoffman for design of National Science Foundation's (NSF's) National Ecological Observatory Network (NEON) domains.

Multivariate Spatiotemporal Clustering (MSTC)

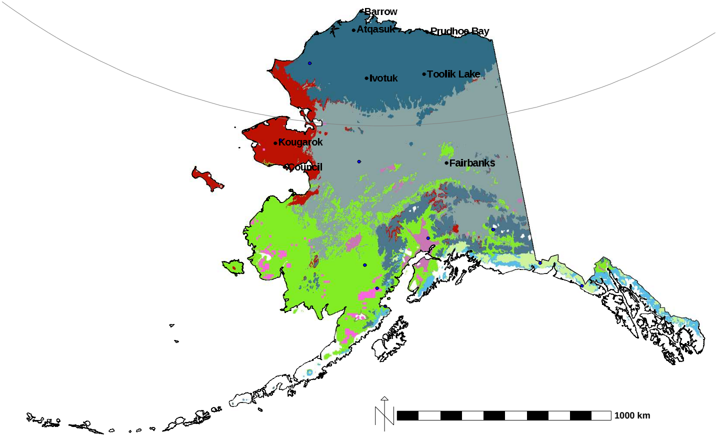


Data Layers

Table: 37 variables averaged for 2000–2009 and 2090–2099

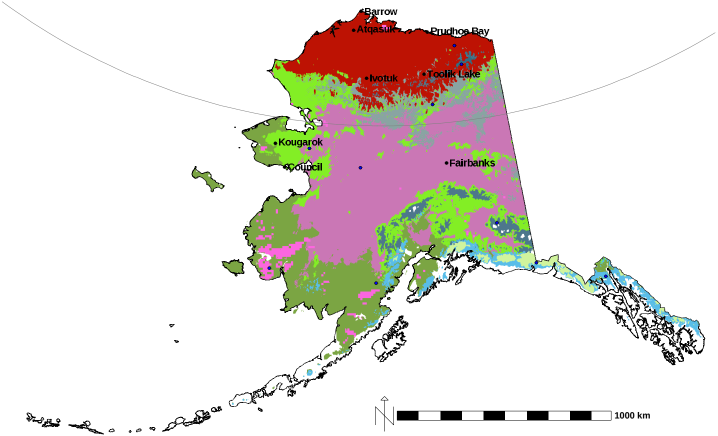
Description	Number/Name	Units	Source
Monthly mean air temperature	12	°C	GCM
Monthly mean precipitation	12	mm	GCM
Day of freeze	mean standard deviation	day of year days	GCM
Day of thaw	mean standard deviation	day of year days	GCM
Length of growing season	mean standard deviation	days days	GCM
Maximum active layer thickness	1	m	GIPL
Warming effect of snow	1	°C	GIPL
Mean annual ground temperature at bottom of active layer	1	°C	GIPL
Mean annual ground surface temperature	1	°C	GIPL
Thermal offset	1	°C	GIPL
Limnicity	1	%	NHD
Elevation	1	m	SRTM

10 Alaska Ecoregions (2000–2009)



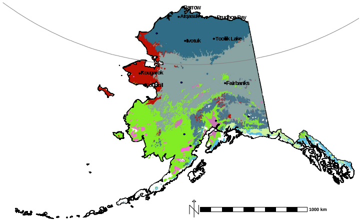
Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

10 Alaska Ecoregions (2090–2099)

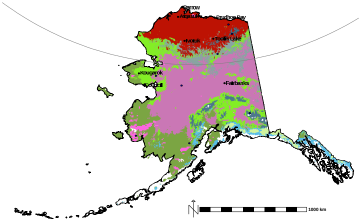


Each ecoregion is a different random color. Blue filled circles mark locations most representative of mean conditions of each region.

10 Alaska Ecoregions, Present and Future



2000–2009

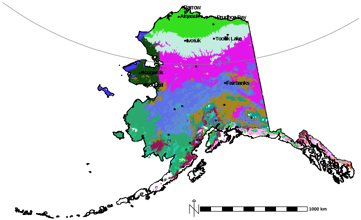


2090–2099

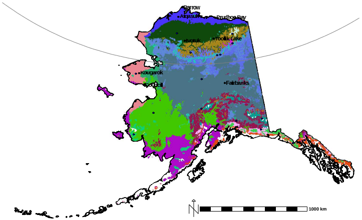
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At this level of division, the conditions in the large boreal forest become compressed onto the Brooks Range and the conditions on the Seward Peninsula migrate to the North Slope.

20 Alaska Ecoregions, Present and Future



2000–2009

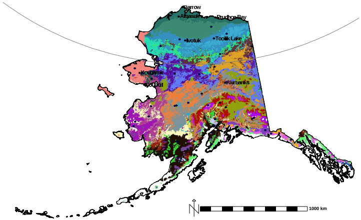


2090–2099

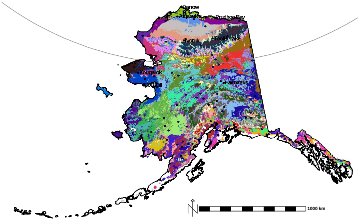
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At this level of division, the two primary regions of the Seward Peninsula and that of the northern boreal forest replace the two regions on the North Slope almost entirely.

50 and 100 Alaska Ecoregions, Present



$k = 50$, 2000–2009

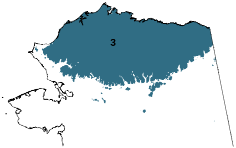


$k = 100$, 2000–2009

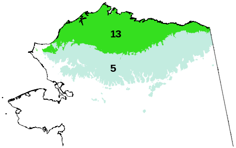
Since the random colors are the same in both maps, a change in color represents an environmental change between the present and the future.

At high levels of division, some regions vanish between the present and future while other region representing new combinations of environmental conditions come into existence.

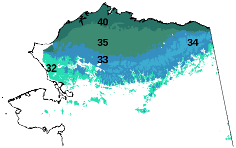
A Hierarchy of Ecoregions



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



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

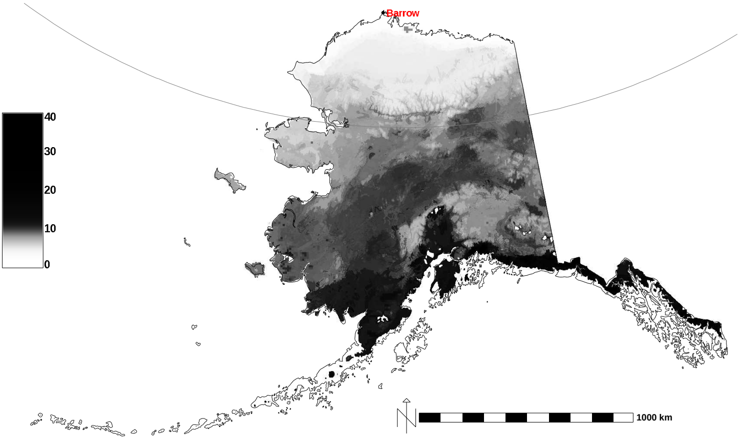


(c) At $k = 50$, the North Slope is occupied by Ecoregion #32, corresponding to the Intermontane Boreal ecological group; Ecoregions #33 and #34, corresponding to mid- and high-elevation of the Brooks Range ecoregion; Ecoregion #35, corresponding to the Brooks Foothills ecoregion; and Ecoregion #40,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

NGEE Arctic Site Representativeness

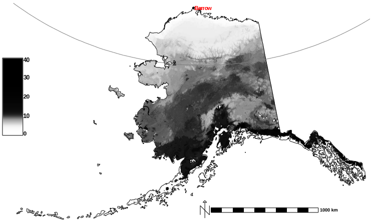
- This representativeness analysis uses the standardized *n*-dimensional data space formed from all input data layers.
- In this data space, the Euclidean distance between a sampling location (like Barrow) and every other point is calculated.
- These data space distances are then used to generate grayscale maps showing the similarity, or lack thereof, of every location to the sampling location.
- In the subsequent maps, white areas are well represented by the sampling location or network, while dark and black areas as poorly represented by the sampling location or network.
- This analysis assumes that the climate surrogates maintain their predictive power and that no significant biological adaptation occurs in the future.

Present Representativeness of Barrow or “Barrow-ness”

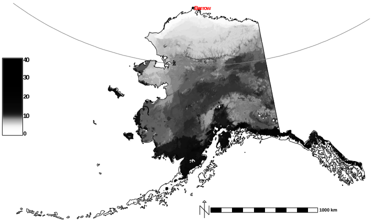


Light-colored regions are well represented and dark-colored regions are poorly represented by the sampling location listed in **red**.

Present vs. Future Barrow-ness



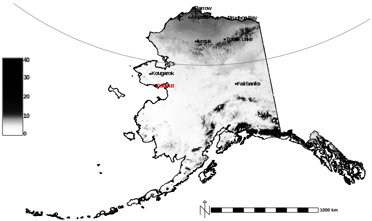
2000–2009



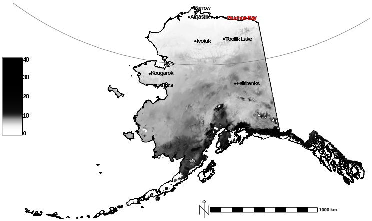
2090–2099

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

Council and Prudhoe Bay Representativeness



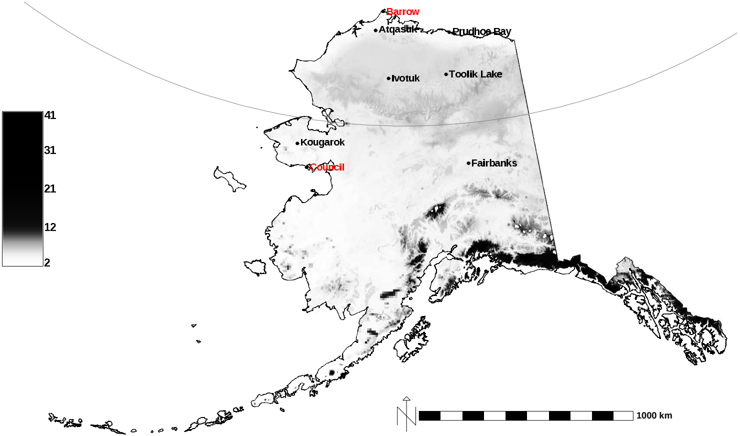
Council



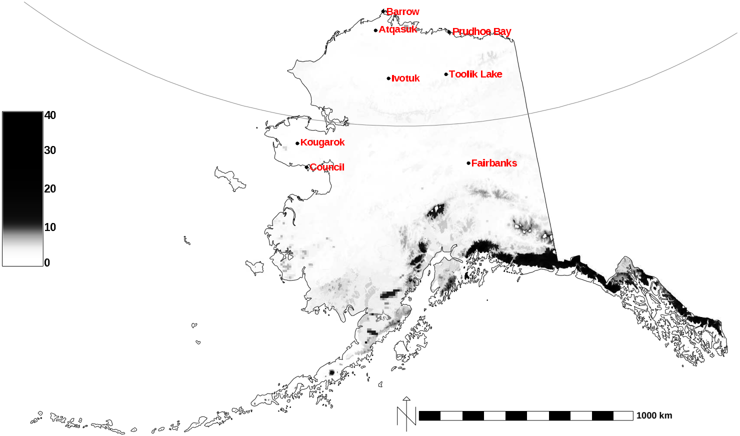
Prudhoe Bay

Representativeness analysis was performed for sites at Barrow, Council, Atqasuk, Ivotuk, Kougarok, Prudhoe Bay, Toolik Lake, and Fairbanks.

Network Representativeness: Barrow + Council



Network Representativeness: All 8 Sites



State Space Dissimilarity: 8 Sites, Present (2000–2009)

Table: 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

State Space Dissimilarity: 8 Sites, Future (2090–2099)

Table: 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

State Space Dissimilarity: 8 Sites, Present and Future

Table: 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

Representativeness: A Quantitative Approach for Scaling

- MSTC provides a quantitative framework for stratifying sampling domains, informing site selection, and determining representativeness of measurements.
- Representativeness analysis provides a systematic approach for up-scaling point measurements to larger domains.
- Methodology is independent of resolution, thus can be applied from site/plot scale to landscape/climate scale.
- It can be extended to include finer spatiotemporal scales, more geophysical characteristics, and remote sensing data.
- First paper describing the methodology has been submitted:

Hoffman, F. M., J. Kumar, R. T. Mills, and W. W. Hargrove (2012) "Representativeness-Based Sampling Network Design for the Arctic." *Landscape Ecol.*, submitted.

References

J. T. Randerson, F. M. Hoffman, P. E. Thornton, N. M. Mahowald, K. Lindsay, Y.-H. Lee, C. D. Nevison, S. C. Doney, G. Bonan, R. Stöckli, C. Covey, S. W. Running, and I. Y. Fung. Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biol.*, 15(9):2462–2484, Sept. 2009. doi:10.1111/j.1365-2486.2009.01912.x.