NGEE Arctic
Next-Generation Ecosystem Experiments

Pan-Arctic Representativeness for Site Selection and Model Evaluation

Forrest M. Hoffman, Jitendra Kumar, W. Robert Bolton, Stan D. Wullschleger, and Colleen Iversen

Oak Ridge National Laboratory, Oak Ridge, TN
Tuesday, December 12, 2023

Next-Generation Ecosystem Experiments: Arctic
Ecosystem Processes
Modeling, Prediction
Systems Understanding

U.S. DEPARTMENT OF ENERGY | Office of Science

OAK RIDGE National Laboratory

Los Alamos NATIONAL LABORATORY

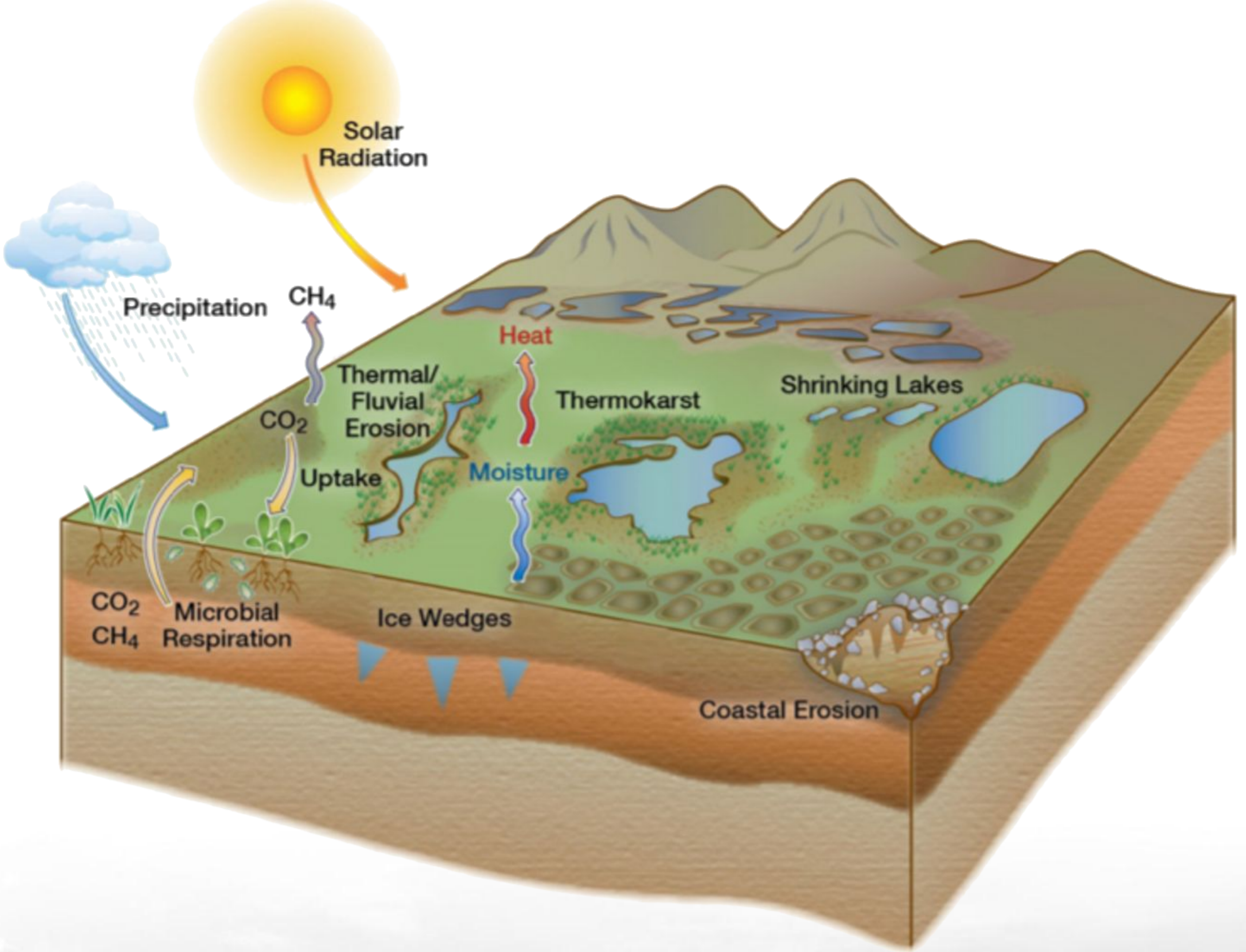
BERKELEY LAB

Brookhaven National Laboratory

UAF UNIVERSITY OF ALASKA FAIRBANKS

Next Generation Ecosystem Experiments (NGEE) - Arctic

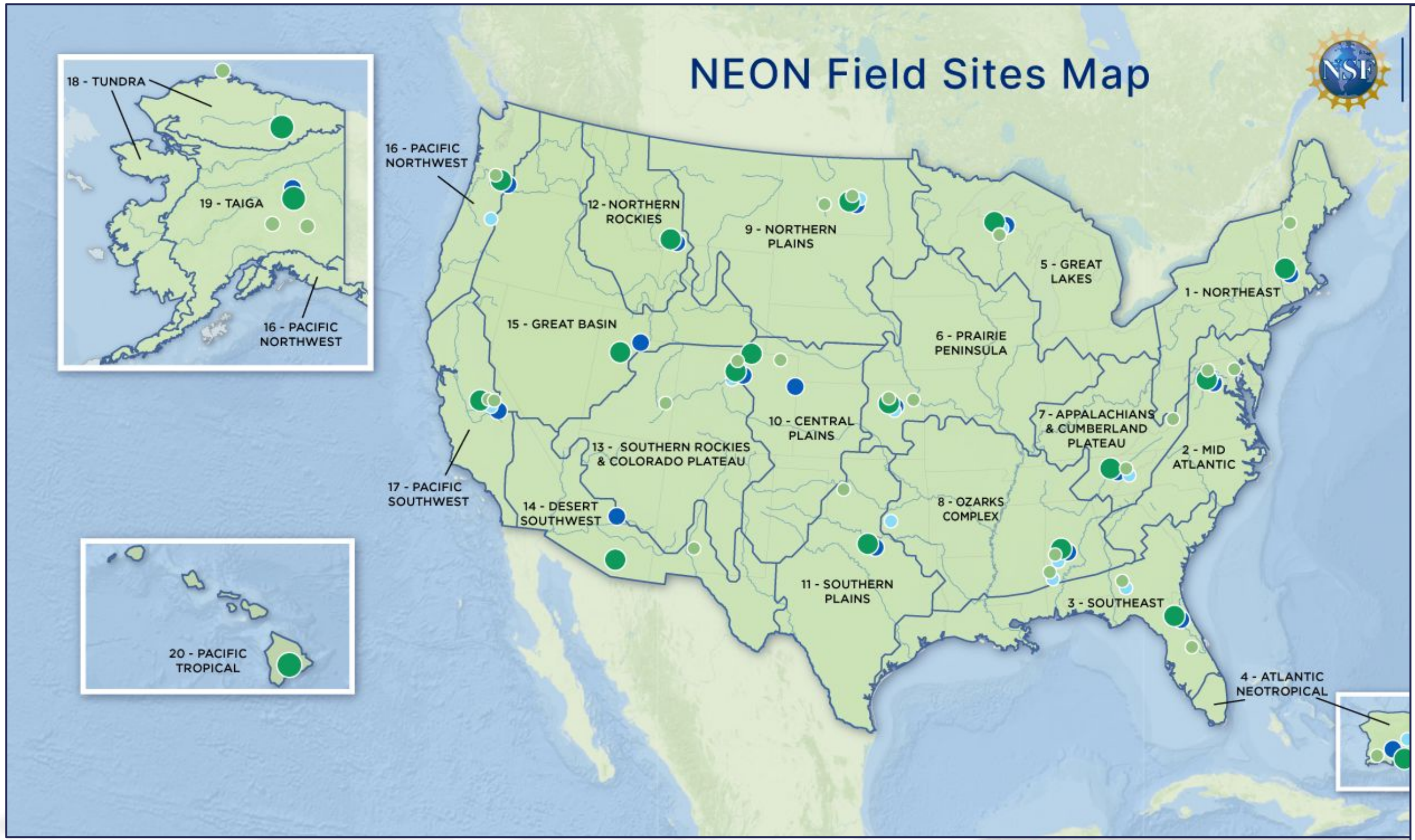
The goal of NGEE Arctic is to support the US Department of Energy (DOE) Biological and Environmental Research (BER) mission to advance a robust predictive understanding of Earth's climate and environmental systems by delivering a process-rich ecosystem model, extending from bedrock to the interface between the vegetative canopy and the atmosphere, that can simulate the evolution of Arctic ecosystems in a changing climate at the scale of a high-resolution grid cell in DOE's Energy Exascale Earth System Model (E3SM)



NGEE Arctic
Next-Generation Ecosystem Experiments

Site Selection, Sampling Design, and Data Synthesis

Site selection and sampling design should be informed by a quantitative multivariate analysis of important environmental gradients to optimize the collection of observations and field measurements and to upscale them to the larger landscape



COMMENTARY COMMENTARY COMMENTARY

282

A continental strategy for the National Ecological Observatory Network

One of the great realizations of the past half-century in both biological and Earth sciences is that, throughout geologic time, life has been shaping the Earth's surface and regulating the chemistry of its oceans and atmosphere (eg Berkner and Marshall 1964). In the present Anthropocene Era (Crutzen and Steffen 2003; Ruddiman 2003), humanity is directly shaping the biosphere and physical environment, triggering potentially devastating and currently unpredictable consequences (Doney and Schimel 2007). While subtle interactions between the Earth's orbit, ocean circulation, and the biosphere have dominated climate feedbacks for eons, now human perturbations to the cycles of CO_2, other trace gases, and aerosols regulate the pace of climate change. Accompanying the biogeochemical perturbations are the vast changes resulting from biodiversity loss and a profound rearrangement of the biosphere due to species movements and invasions. Scientists and managers of biological resources require a stronger basis for forecasting the consequences of such changes.

In this Special Issue of *Frontiers*, the scientific community confronts the challenge of research and environmental management in a human-dominated, increasingly connected world (Peters *et al.* p 229). Carbon dioxide, a key driver of climate change produced by a host of local and small-scale processes (eg clearing of forests, extraction and use of fossil fuels), affects the global energy balance (Marshall *et al.* p 273). Invasive species, though small from a large-scale perspective, nonetheless modify the continental biosphere (Crowl *et al.* p 238). Aquatic systems are tightly coupled to both terrestrial systems and the marine environment (Hopkinson *et al.* p 255). Flowing water not only intrinsically creates a highly connected system, but acts a transducer of climate, land-use, and invasive species effects, spreading their impacts from terrestrial and upstream centers of action downstream and into distant systems (Williamson *et al.* p 247). Human activities such as urbanization create new connections; materials, organisms, and energy flow into cities from globally distributed sources and waste products are exported back into the environment (Grimm *et al.* p 264).

All of the papers in this issue of *Frontiers* conclude that a new approach to studying the biosphere is required in the present era. In response to this challenge, with the support of the National Science Foundation (NSF), ecologists in the US are planning a National Ecological Observatory Network (NEON). The conceptual design of this network (Field *et al.* 2006) gives rise to several general questions:

(1) How will the ecosystems (of the US) and their components respond to changes in natural- and human-

www.frontiersinecology.org

GUEST EDITORIAL GUEST EDITORIAL GUEST EDITORIAL

59

NEON: a hierarchically designed national ecological network

In the past year, planning for the National Ecological Observatory Network (NEON; www.neoninc.org/) has made major advances. The Integrated Science and Education Plan (ISEP) was completed and reviewed, the Conceptual Design Review was held, a Request For Information (RFI) on prospective sites and integrative science questions was released, and responses were received. This February, NEON Inc released its site-specific design proposal, identifying core wildland sites and environmental gradients for study. This site-specific proposal will form the basis of the NEON Project Execution Plan, which will be submitted to the National Science Foundation for review as a Major Research Equipment and Facilities Construction project in the spring of 2007. Evaluating the RFI responses and selecting sites consistent with the ISEP was a complex and humbling challenge, the basis for which is detailed below.

The NEON Inc observing strategy and site selection process is based on systematic sampling across the largest scales of ecological variability to provide a basis for "scaling up" analyses across the nation. NEON has divided the continent into eco-climatic regimes called "domains". The conterminous US plus Puerto Rico has 17 domains, and Alaska and Hawaii add three more, to give a total of 20. The NEON network includes stable, fixed elements (core wildland sites), relocatable gradient sites, and mobile laboratories. The NEON domains have been chosen as elements of a continental-scale observing strategy, based on theory and informed by a wide range of datasets and statistical approaches.

The NEON Inc spatial sampling design is based on climatic, edaphic, and topographic attributes, delineated into domains using climate data and soil properties. The domains are mapped based on physical variables that effectively capture key biological aspects of US ecology. The NEON domains were based on an explicit multivariate statistical procedure that is transparent and repeatable, and provides a national stratification that shows how best to deploy our 20 core sites in order to maximize the coverage, coherence, and representativeness of the network. The domains can also be used to extrapolate measurements made at sites, facilitating upscaling (http://research.esd.ornl.gov/~hnw/neon/withindomainrep2). This sampling plan, based on 20 domains, makes NEON the largest and most comprehensive ecological network to have been statistically designed before deployment.

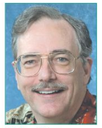
NEON's design can be viewed as a hierarchy of constraints. While we often think of climate as an independent variable relative to biology, at the continental scale climate variables are constrained by latitude, "continentality", and orography (the physical geography of mountains). The influence of the ocean basins that surround the continent also percolates into each domain, affecting temperature, precipitation, and their variability in time (eg via El Niño). Core wildland sites address these largest scales.

Characteristic patterns of land use, management, disturbances (such as fire or flooding), and recovery develop within the domains, constrained by the biophysical setting, the local biota, and historical effects. Different patterns of natural resource use and settlement tend to evolve in each domain (eg timber production in forested areas and agriculture in regions with ample fertility and precipitation or irrigation). These varying patterns of socio-environmental regimes allow the study of ecosystem interactions with human dynamics. Relocatable gradient sites sample these smaller scales of variation.

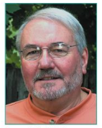
Ecologists have primarily exploited naturally occurring variability along gradients where only one factor varies, as in Hans Jenny's famous studies along soil age and climate gradients. The NEON domains do not represent sites chosen along single-factor axes; instead, they sample regimes in which climate, biota, soils, and land-management practices vary together. While this design does not allow for simple statistical inference, it provides diverse conditions across which hypotheses, questions, and models can be addressed. Modern statistical and process models are now sophisticated enough to be used with this complex and interactive regime design.

Considerable flexibility remains in the instruments and measurements to be deployed and the future sites of the relocatable facilities, so that the current NEON design can evolve as our science matures. The current network design balances the broad system of national coverage with the relevance of individual sites and sub-networks to local and regional questions. Many difficult decisions remain, and NEON Inc depends on the community's input. Everyone at NEON Inc, from the Project Office staff to the Board of Directors, is open to any and all communication and looks forward to hearing from you.


David Schimel, NEON Inc, Boulder, CO

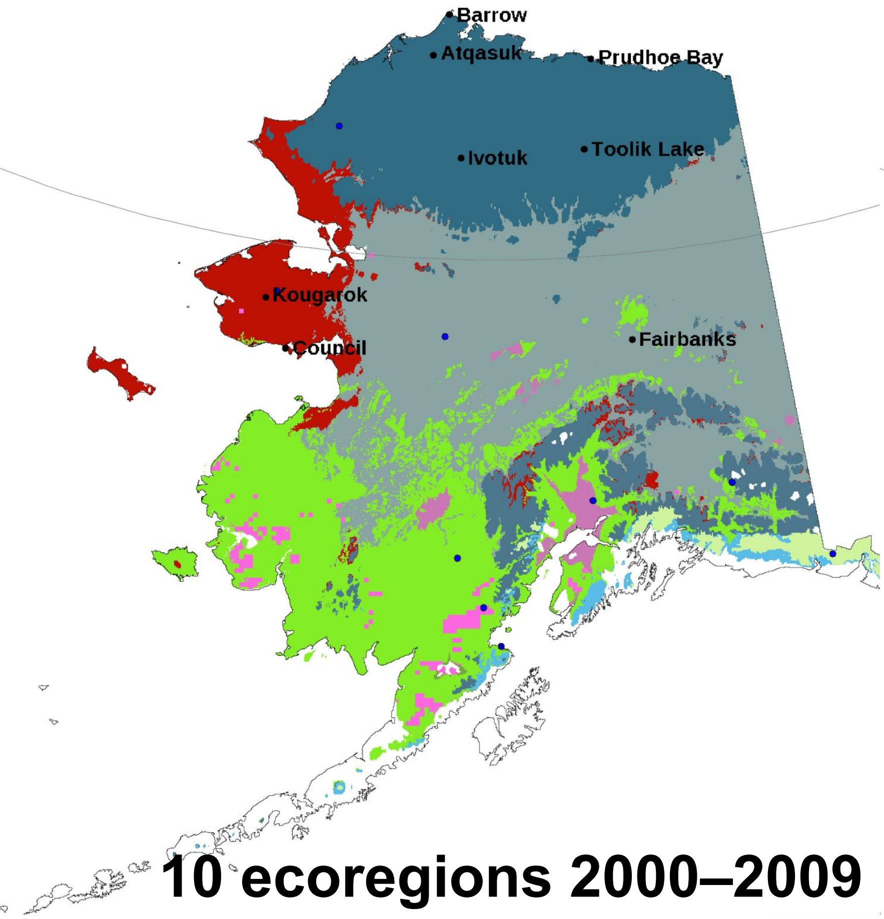

William Hargrove, USDA Forest Service, Asheville, NC


Forrest Hoffman, Oak Ridge National Laboratory, TN

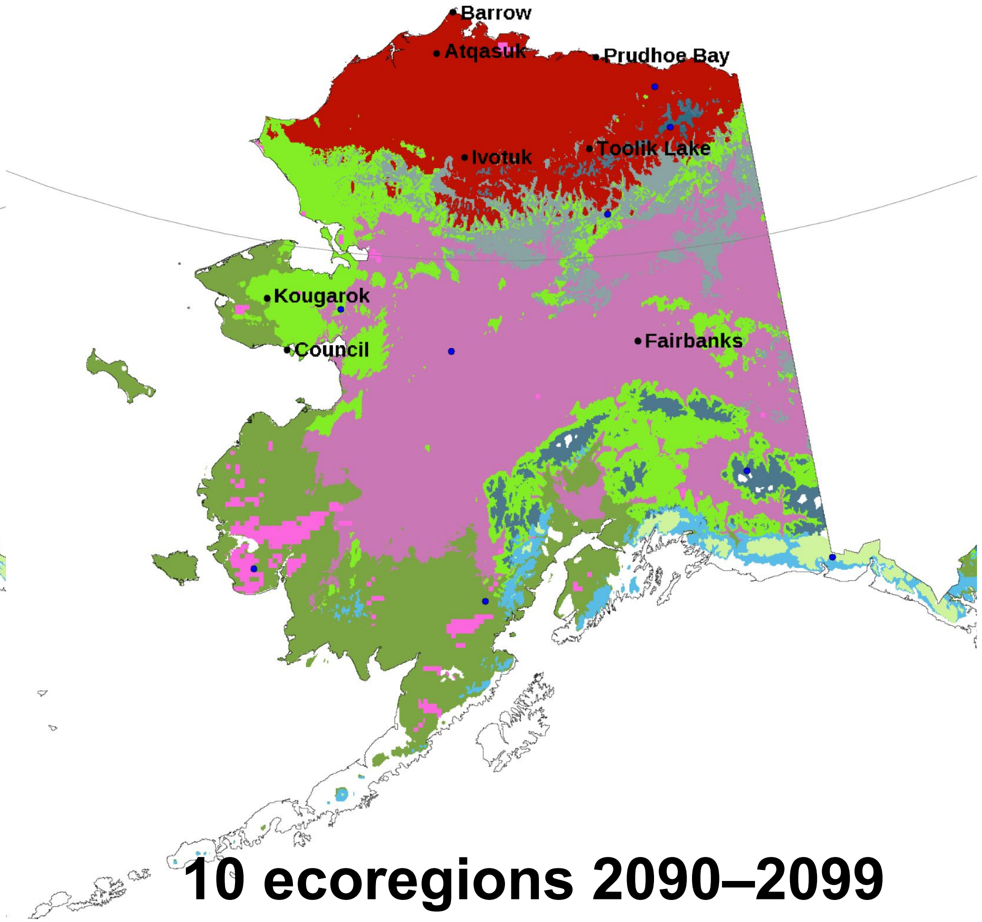

James MacMahon, Utah State University, UT

Ecoregions of Alaska

We previously developed a dynamic ecoregionalization for the State of Alaska that showed northward movement of ecoregions under projected climate change



10 ecoregions 2000–2009



10 ecoregions 2090–2099

Landscape Ecol
DOI 10.1007/s10980-013-9902-0

RESEARCH ARTICLE

Representativeness-based sampling network design for the State of Alaska

Forrest M. Hoffman · Jitendra Kumar · Richard T. Mills · William W. Hargrove

Received: 13 February 2013 / Accepted: 31 May 2013

Abstract Resource and logistical constraints limit the frequency and extent of environmental observations, particularly in the Arctic, necessitating the development of a systematic sampling strategy to maximize coverage and objectively represent environmental variability at desired scales. A quantitative methodology for stratifying sampling domains, informing site selection, and determining the representativeness of measurement sites and networks is described here. Multivariate spatiotemporal clustering was applied to down-scaled general circulation model results and data for the State of Alaska at 4 km^2 resolution to define multiple sets of ecoregions across two decadal time periods. Maps of ecoregions for the present (2000–2009) and future (2090–2099) were produced, showing how combinations of 37 characteristics are distributed and how they may shift in the future. Representative sampling locations are identified on present and future ecoregion maps. A representativeness metric was developed, and representativeness maps for eight candidate sampling locations were produced. This metric was used to characterize the environmental similarity of each site. This analysis provides model-inspired insights into optimal sampling strategies, offers a framework for up-scaling measurements, and provides a down-scaling approach for integration of models and measurements. These techniques can be applied at different spatial and temporal scales to meet the needs of individual measurement campaigns.

Keywords Ecoregions · Representativeness · Network design · Cluster analysis · Alaska · Permafrost

F. M. Hoffman (✉)
Computer Science & Mathematics Division, Climate Change Science Institute (CCSI), Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: forrest@climatemodeling.org

F. M. Hoffman · J. Kumar · R. T. Mills
Environmental Sciences Division, Climate Change Science Institute (CCSI), Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: jkumar@climatemodeling.org

R. T. Mills
e-mail: rmills@ornl.gov

W. W. Hargrove
Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, Southern Research Station, Asheville, NC, USA
e-mail: hnw@geobabble.org

Introduction

The Arctic contains vast amounts of frozen water in the form of sea ice, snow, glaciers, and permafrost. Extended areas of permafrost in the Arctic contain soil organic carbon that is equivalent to twice the size of the atmospheric carbon pool, and this large stabilized

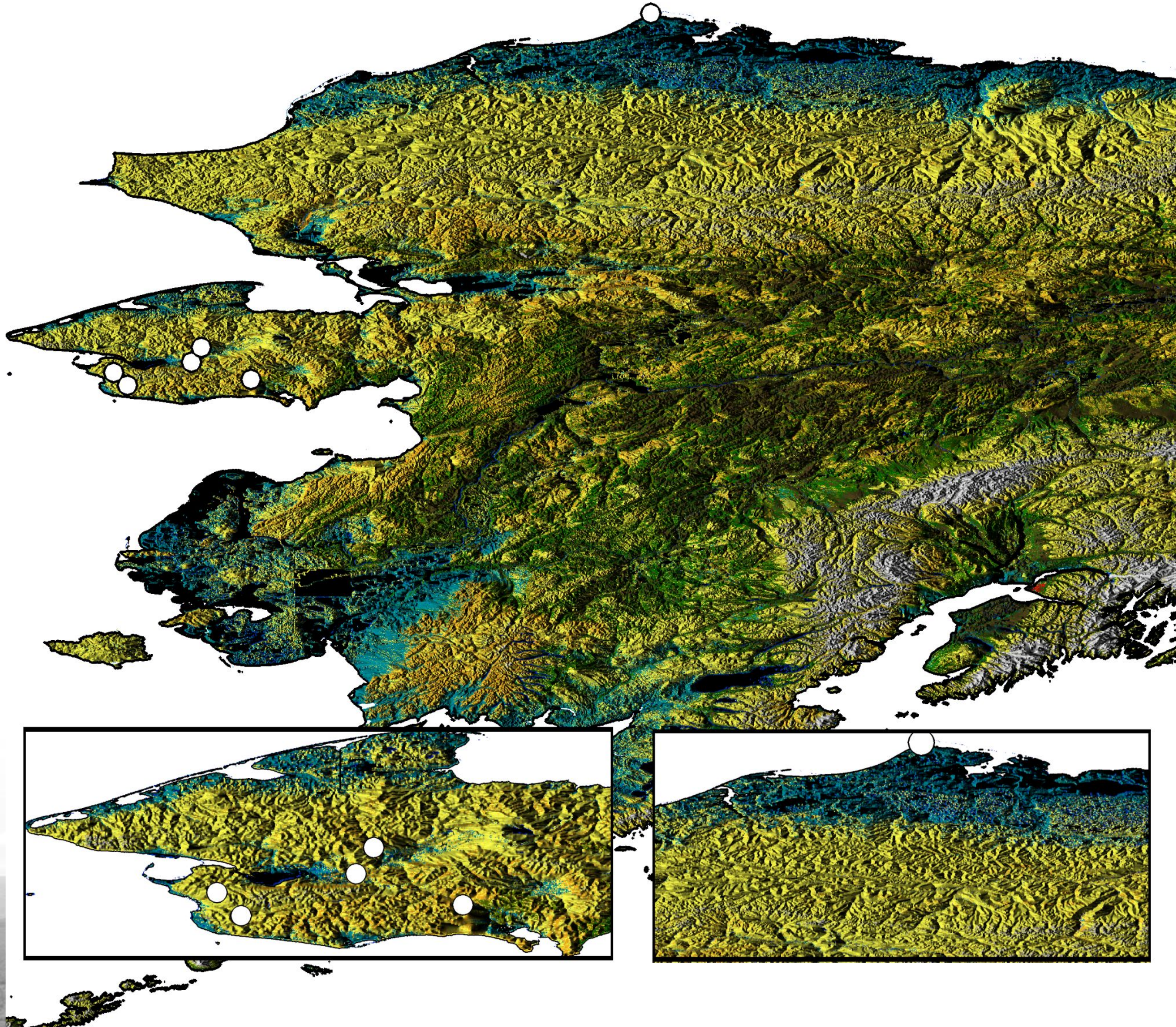
Published online: 20 June 2013
Springer

This study informed site selection for NGEE Arctic Phases 2 and 3

Site selection for NGEE Arctic Phases 1–3

Phase 1 research focused on polygonal ground tundra in the Barrow Environmental Observatory in Utqiaġvik, Alaska.

Phase 2 and 3 research was extended to sites, spanning three watersheds, on the Seward Peninsula of Alaska.



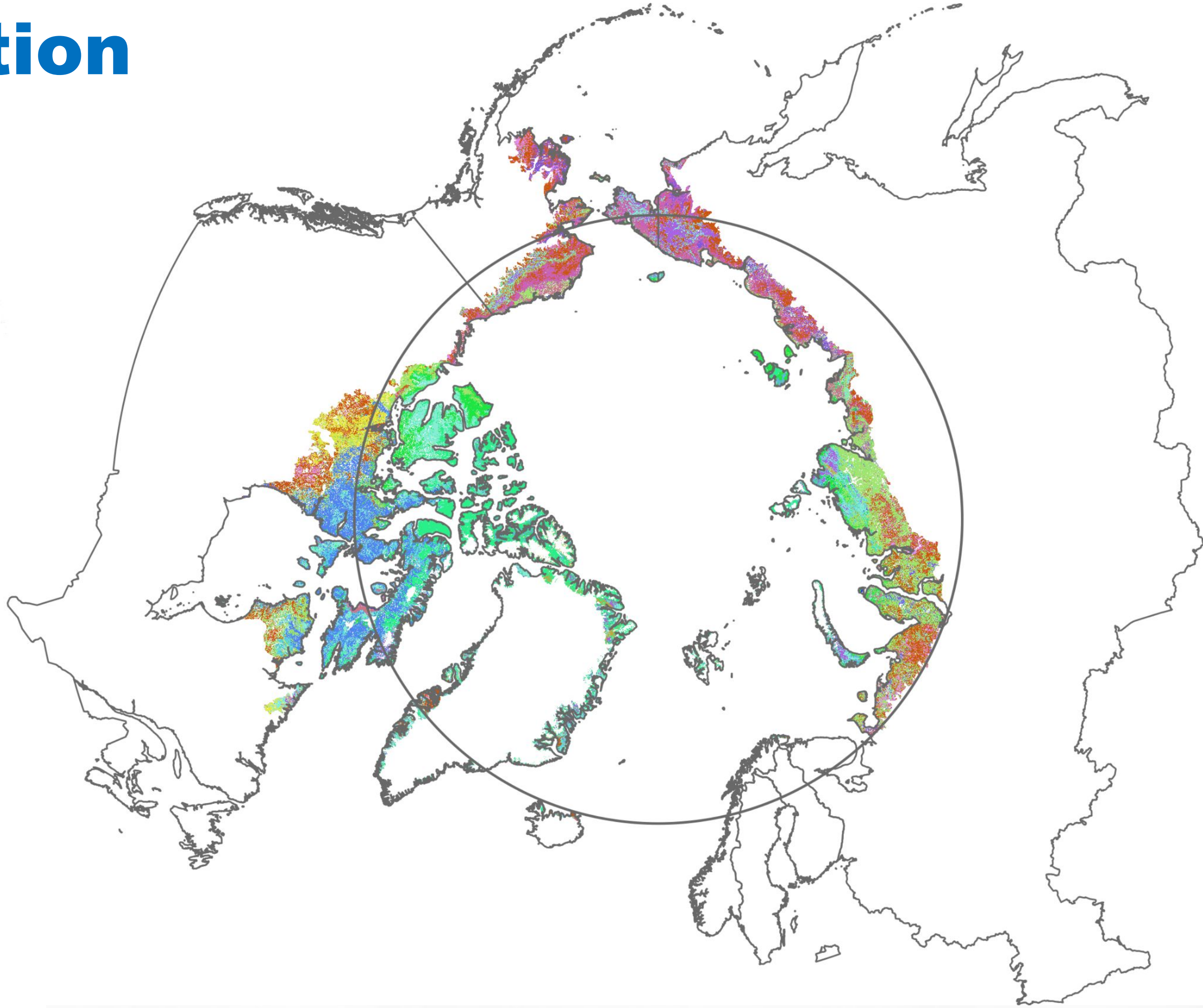
Circumpolar Arctic Vegetation

The circumpolar Arctic represents a large and heterogeneous region with a wide diversity of vegetation types.

NGEE Arctic aims to improve the process representation of the Arctic ecosystems in global scale models.

This requires data integration across many sparsely sampled regions to constrain model processes.

- [B1] Cryptogam herb barren
- [B2a] Cryptogam barren complex
- [B3] Non-carbonate mountain complex
- [B4] Non-carbonate mountain complex
- [B2b] Cryptogam dwarf-shrub barren complex
- [G1] Rush/grass forb cryptogam tundra
- [G2] Graminoid prostrate dwarf-shrub forb tun
- [G3] Non-tussock sedge dwarf-shrub moss tun
- [G4] Tussock-sedge dwarf-shrub moss tundra
- [P1] Prostrate dwarf-shrub herb tundra
- [P2] Prostrate/hemi-prostrate dwarf-shrub tun
- [S1] Erect dwarf-shrub moss tundra
- [S2] Low-shrub tundra
- [W1] Sedge/grass moss wetland
- [W2] Sedge moss dwarf-shrub wetland
- [W3] Sedge moss low-shrub wetland
- [FW] Lakes
- [SW] Lagoons
- [GL] Glacier
- [NA] Non-arctic



Circumpolar Arctic Vegetation Map (CAVM)

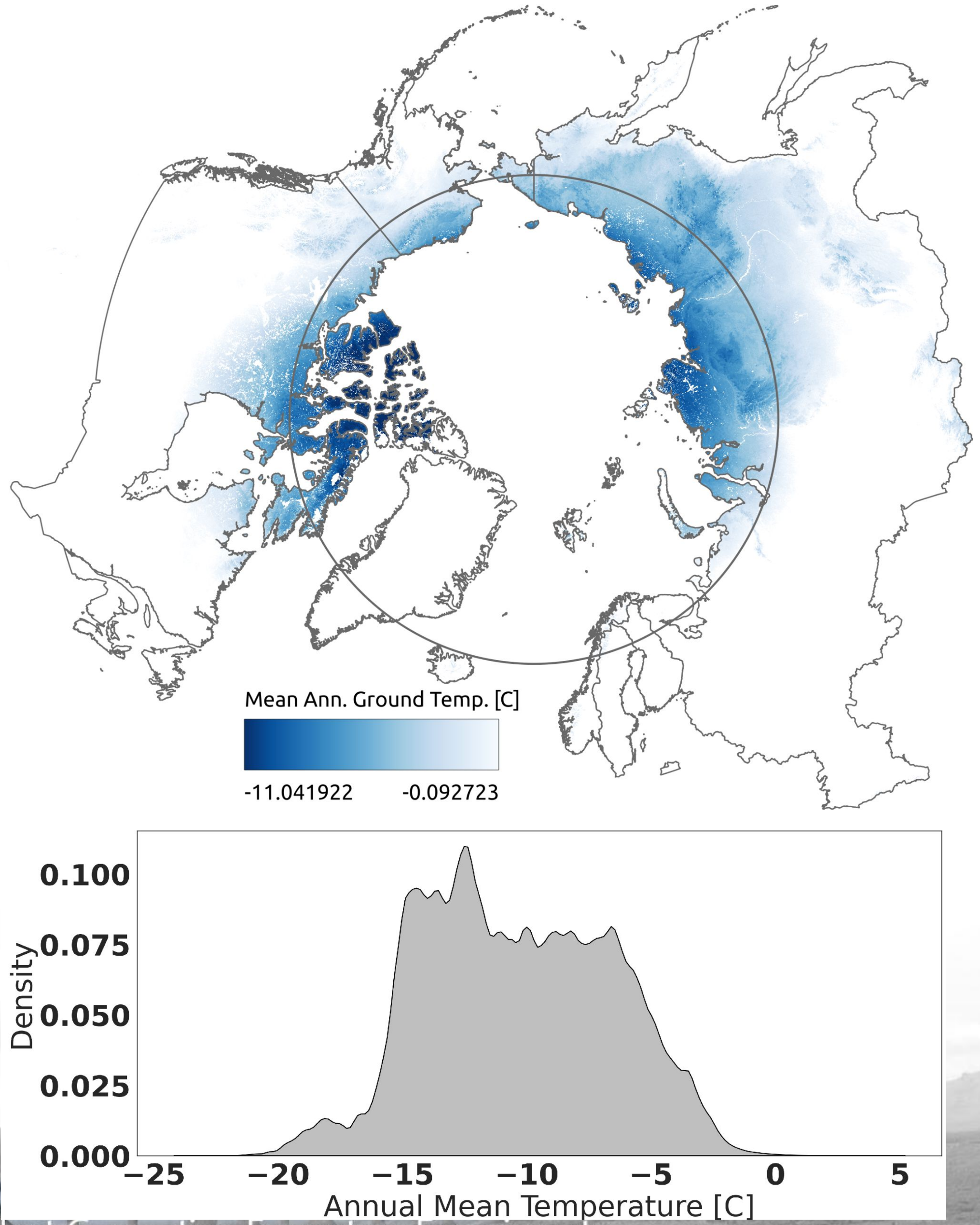
Multivariate Characterization of Pan-Arctic Environment

To inform strategies for pan-Arctic model evaluation and site selection, we conducted a comprehensive quantitative multivariate assessment that leveraged and integrated best-available observations.

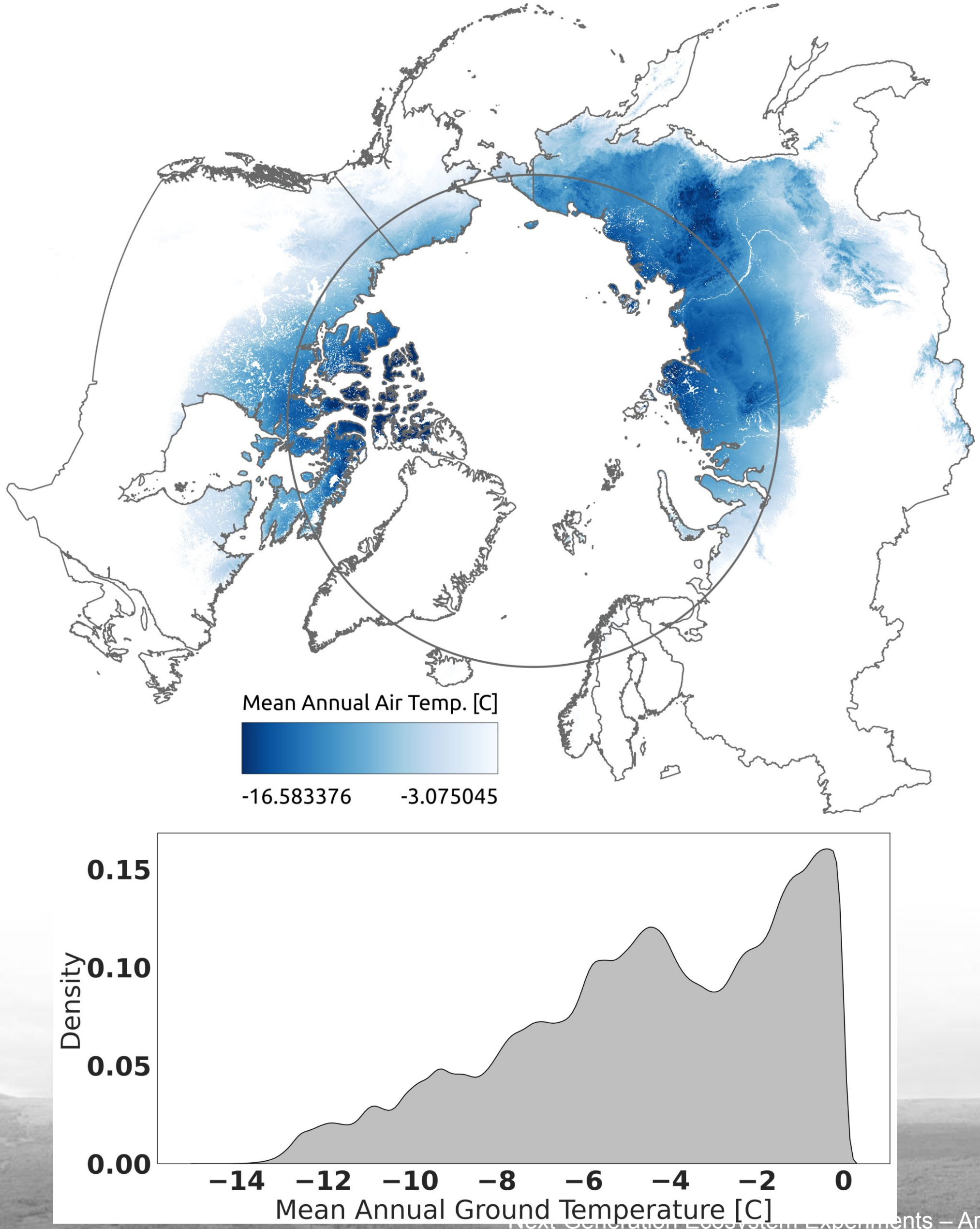
NGEE Arctic will scale up an Arctic-informed version of E3SM Land Model (ELM) to the pan-Arctic region, and our analysis provides a framework for synthesis and evaluation of this pan-Arctic ELM.

#	Variable	Source	Resolution
1	Annual mean temperature	WorldClim 2.0	1km
2	Isothermality	WorldClim 2.0	1km
3	Temperature seasonality	WorldClim 2.0	1km
4	Mean temperature of warmest quarter	WorldClim 2.0	1km
5	Mean temperature of coldest quarter	WorldClim 2.0	1km
6	Annual precipitation	WorldClim 2.0	1km
7	Precipitation seasonality	WorldClim 2.0	1km
8	Precipitation of warmest quarter	WorldClim 2.0	1km
9	Precipitation of coldest quarter	WorldClim 2.0	1km
10	Available water capacity	Soilgrids250m 2.0	250m [1km]
11	Soil bulk density	Soilgrids250m 2.0	250m [1km]
12	Soil carbon content	Soilgrids250m 2.0	250m [1km]
13	Soil nitrogen content	Soilgrids250m 2.0	250m [1km]
14	pH	Soilgrids250m 2.0	250m [1km]
15	Compound topographic index	HYDRO1k	1km
16	Active Layer Thickness	Ran et. al. 2022, ESSD	1km
17	Mean Annual Ground Temperature	Ran et. al. 2022, ESSD	1km
18	Permafrost Probability	Ran et. al. 2022, ESSD	1km
19	MODIS GPP Cumulative	DHI U. Wisconsin	1km
20	MODIS GPP Coeff variation	DHI U. Wisconsin	1km
21	MODIS NDVI Cumulative	DHI U. Wisconsin	1km
22	MODIS NDVI Coeff variation	DHI U. Wisconsin	1km

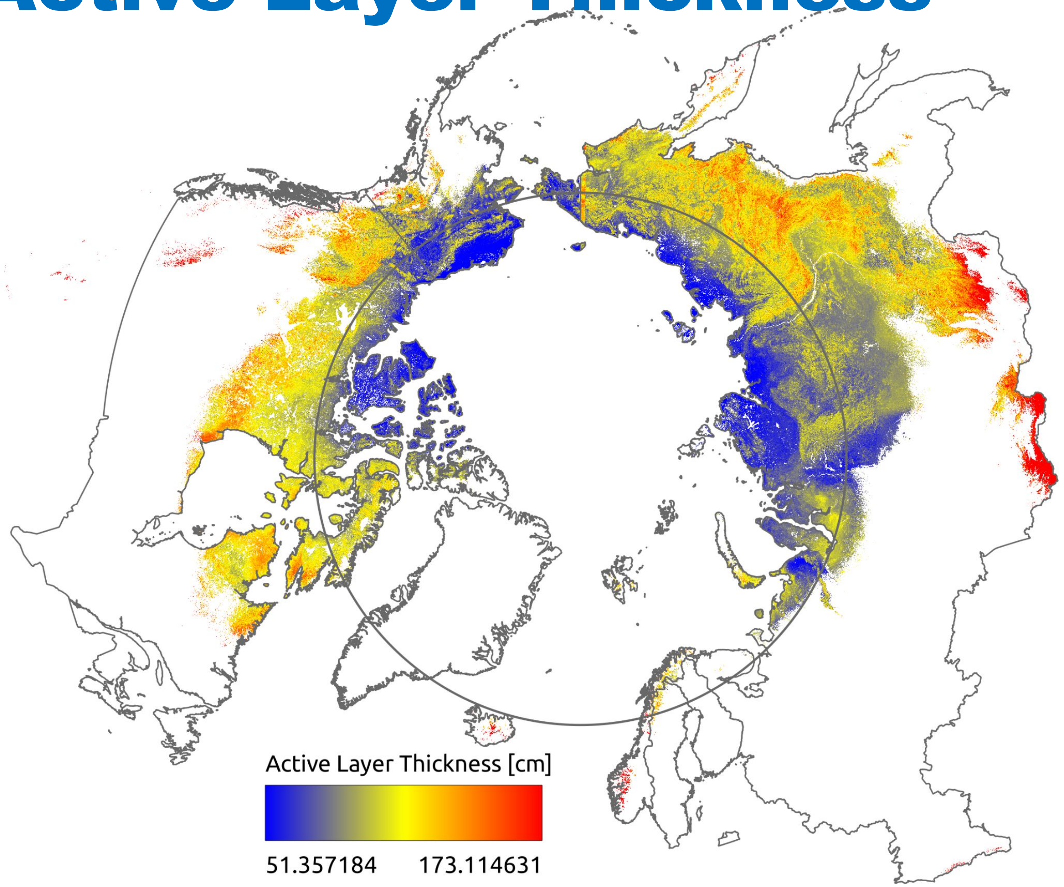
Mean Annual Air Temperature

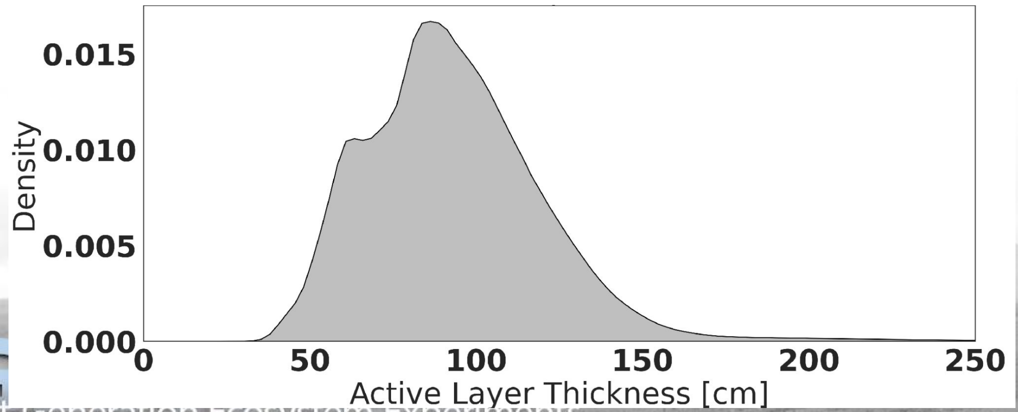


Mean Annual Ground Temperature

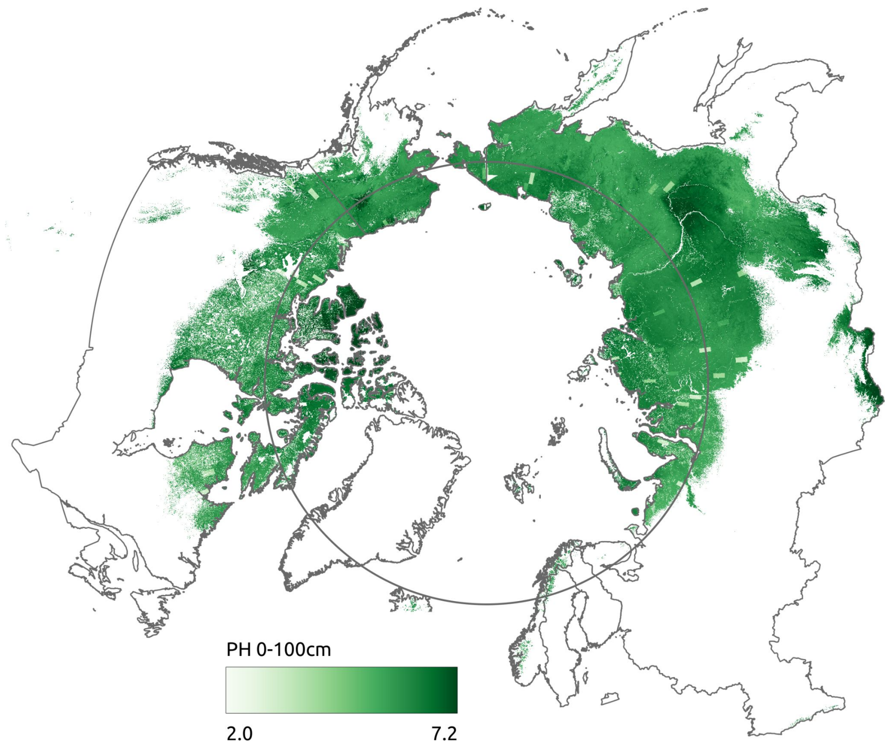


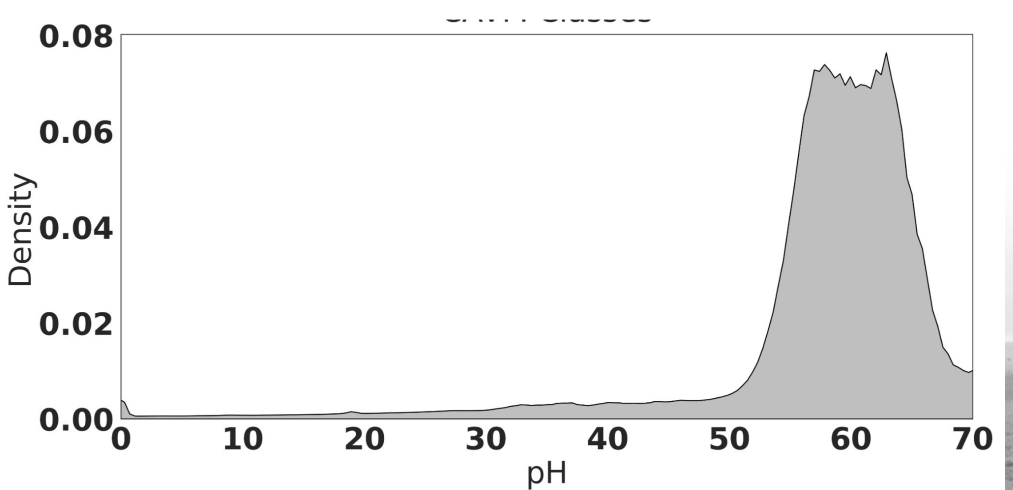
Active Layer Thickness



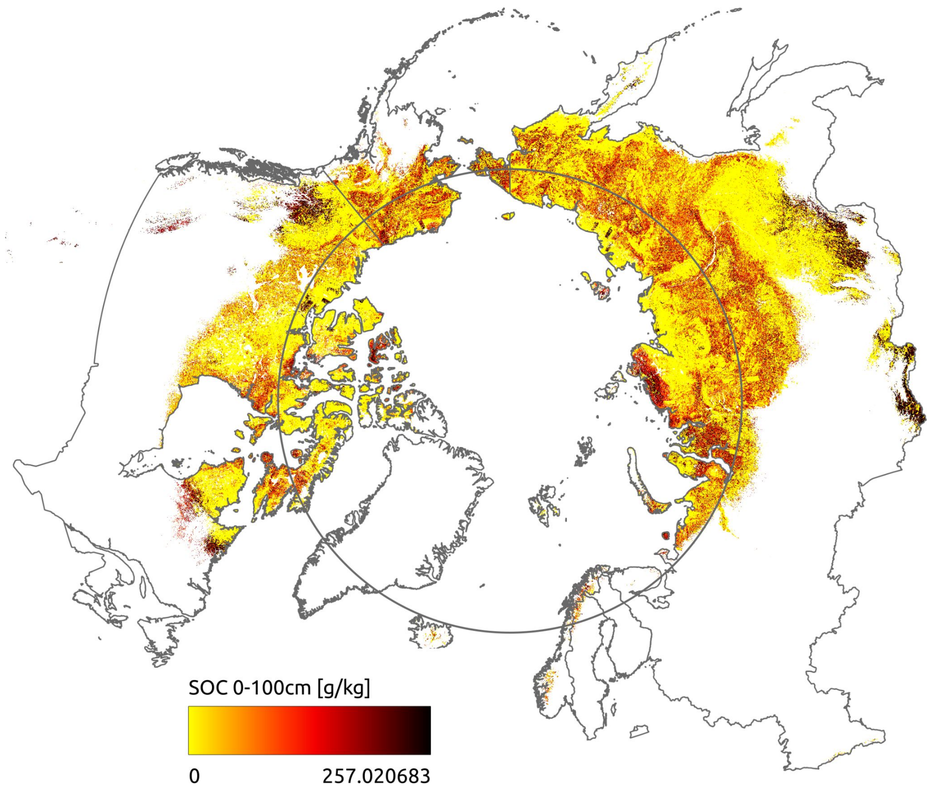


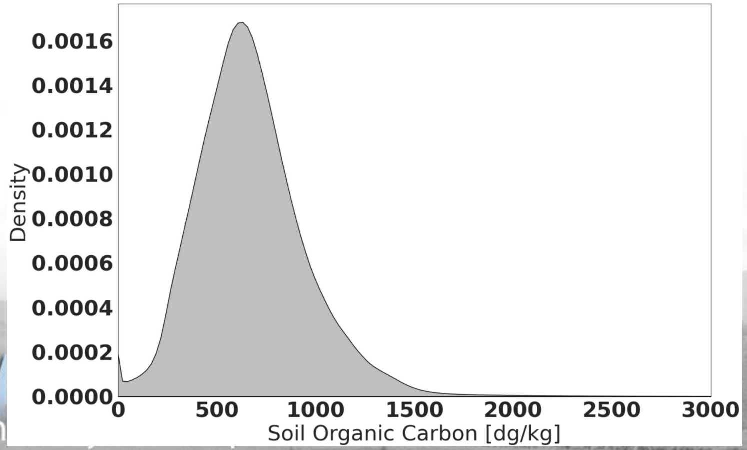
pH (0–100cm)



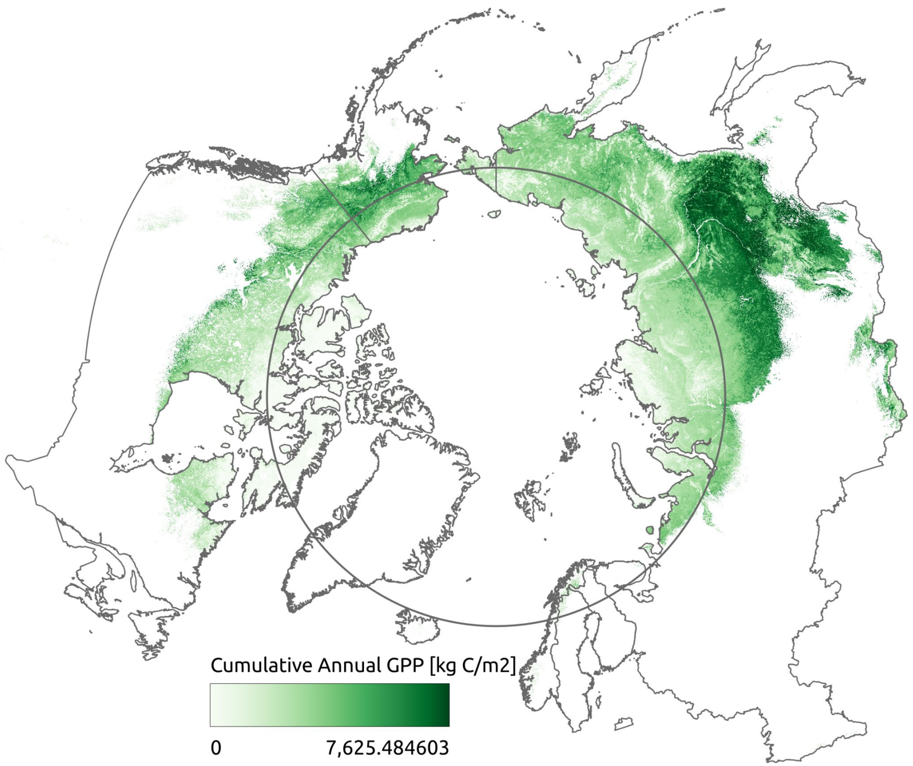


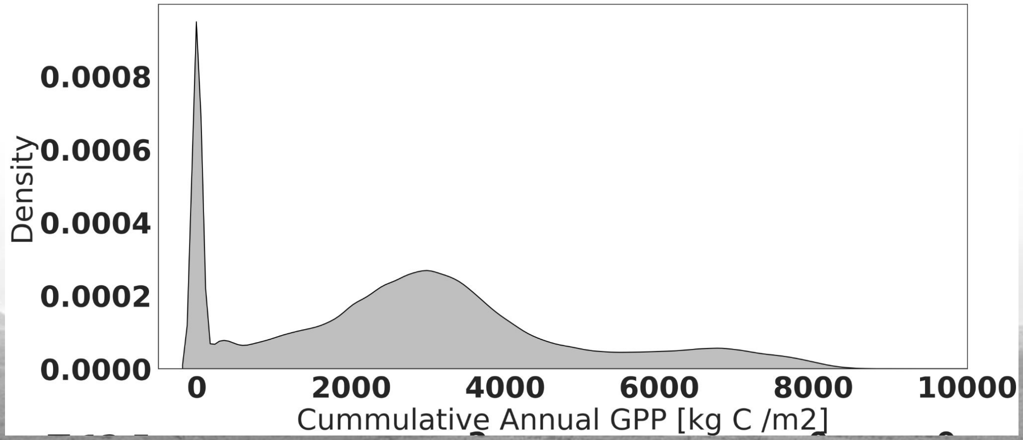
Soil Organic Carbon (0–100cm)





Annual GPP

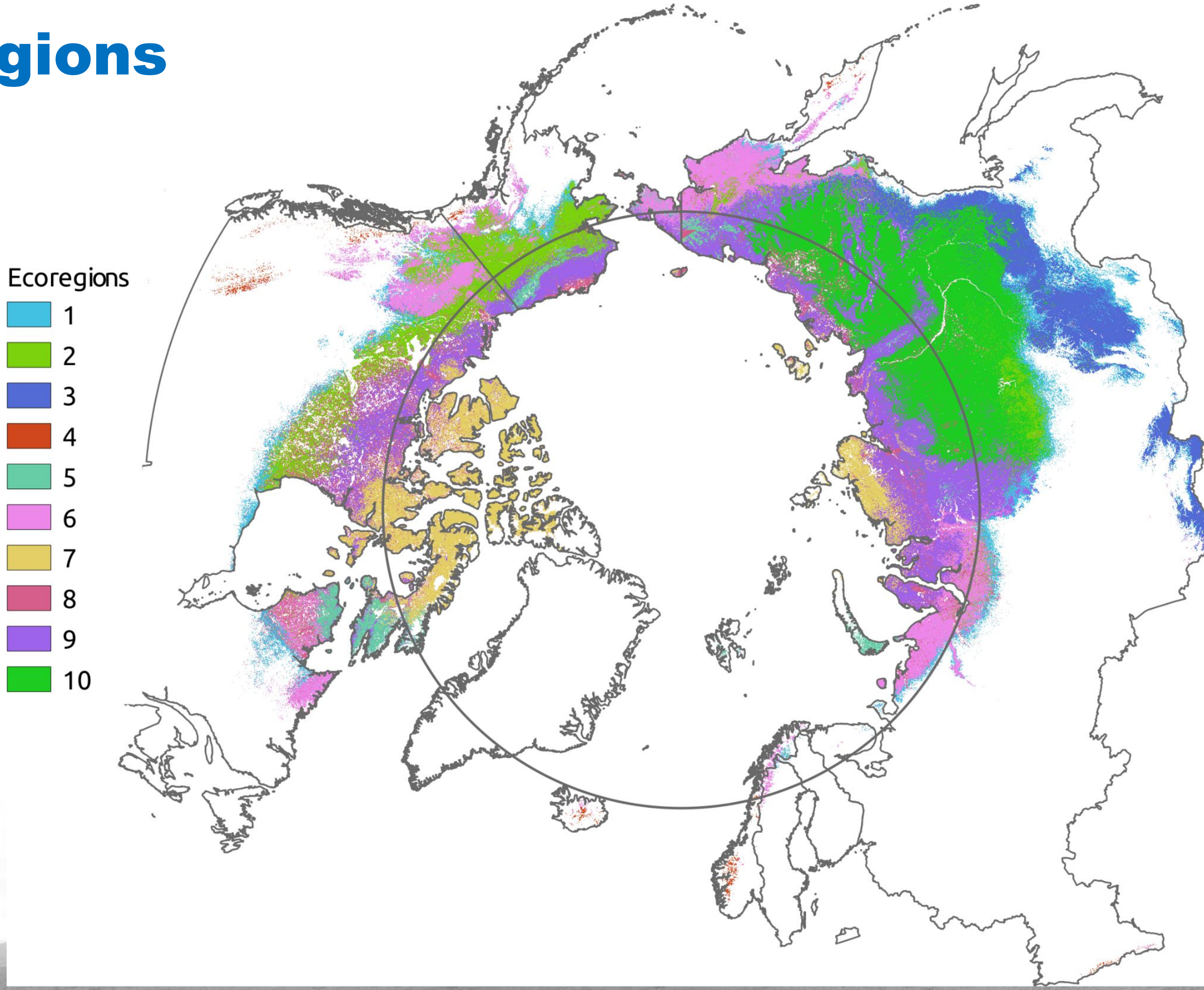




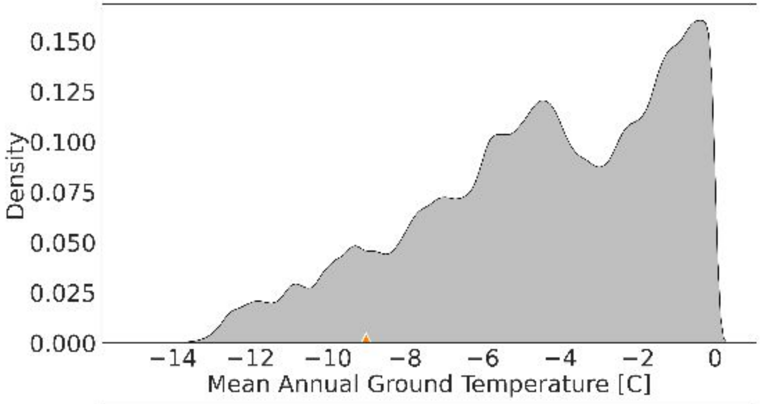
10 Pan-Arctic Ecoregions

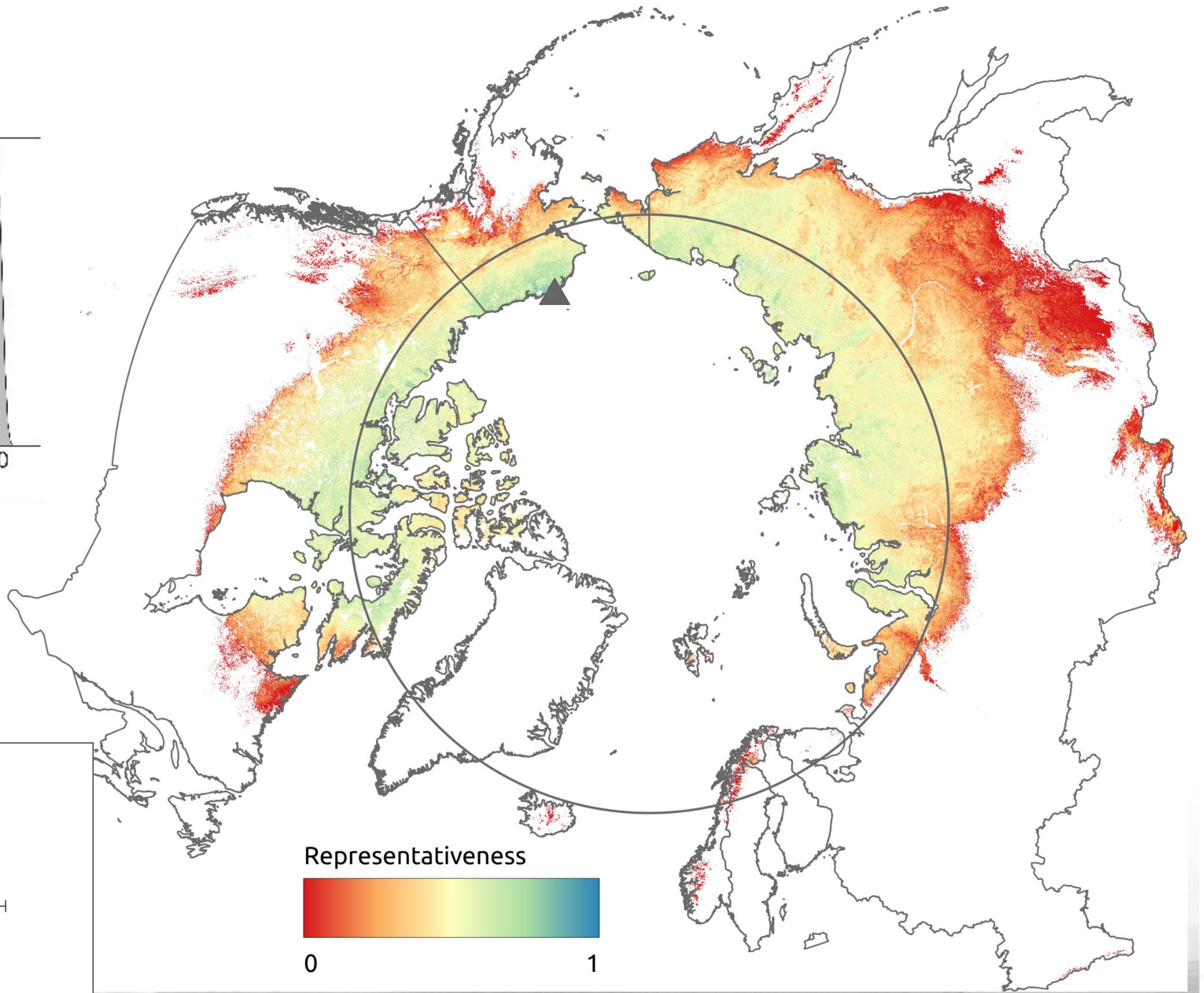
We used all 22 data layers across the pan-Arctic in a *k*-means cluster analysis to equitably partition the data variance into 10 clusters that represent ecoregions.

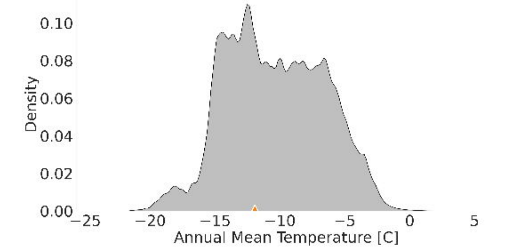
The objective is to select sites that provide maximum coverage across those ecoregions and to use the ecoregions as a framework for upscaling data to the full pan-Arctic region.

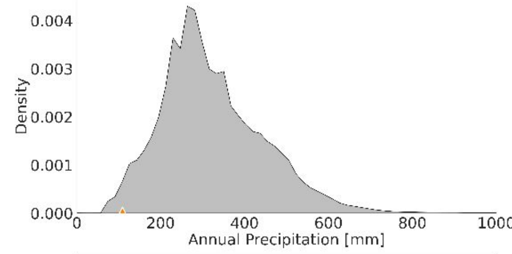


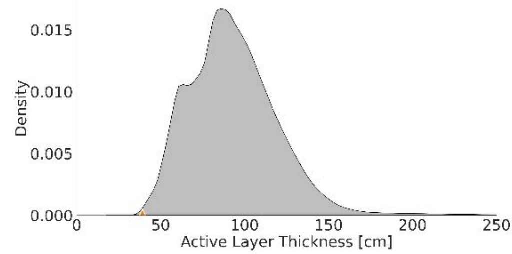
Phase 1: Representativeness (Barrow)

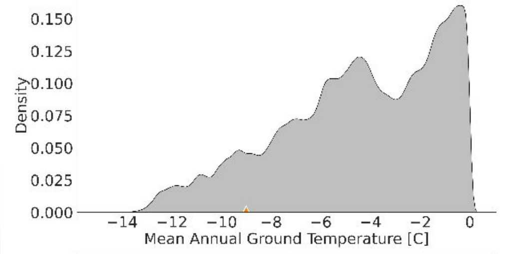


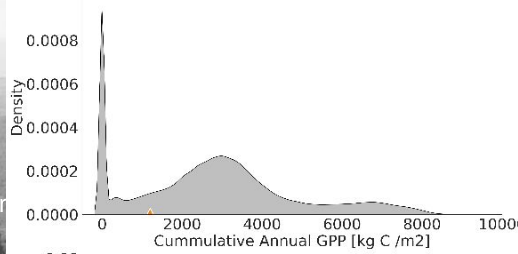


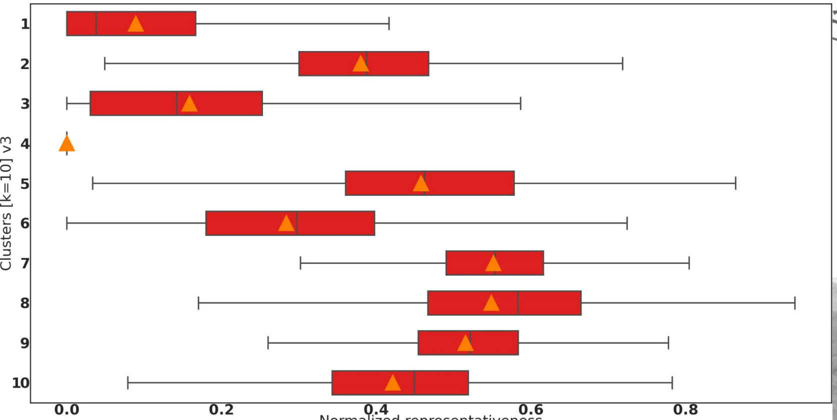




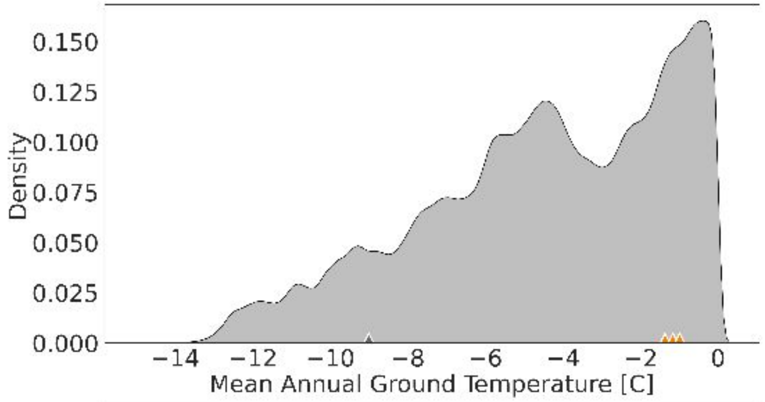


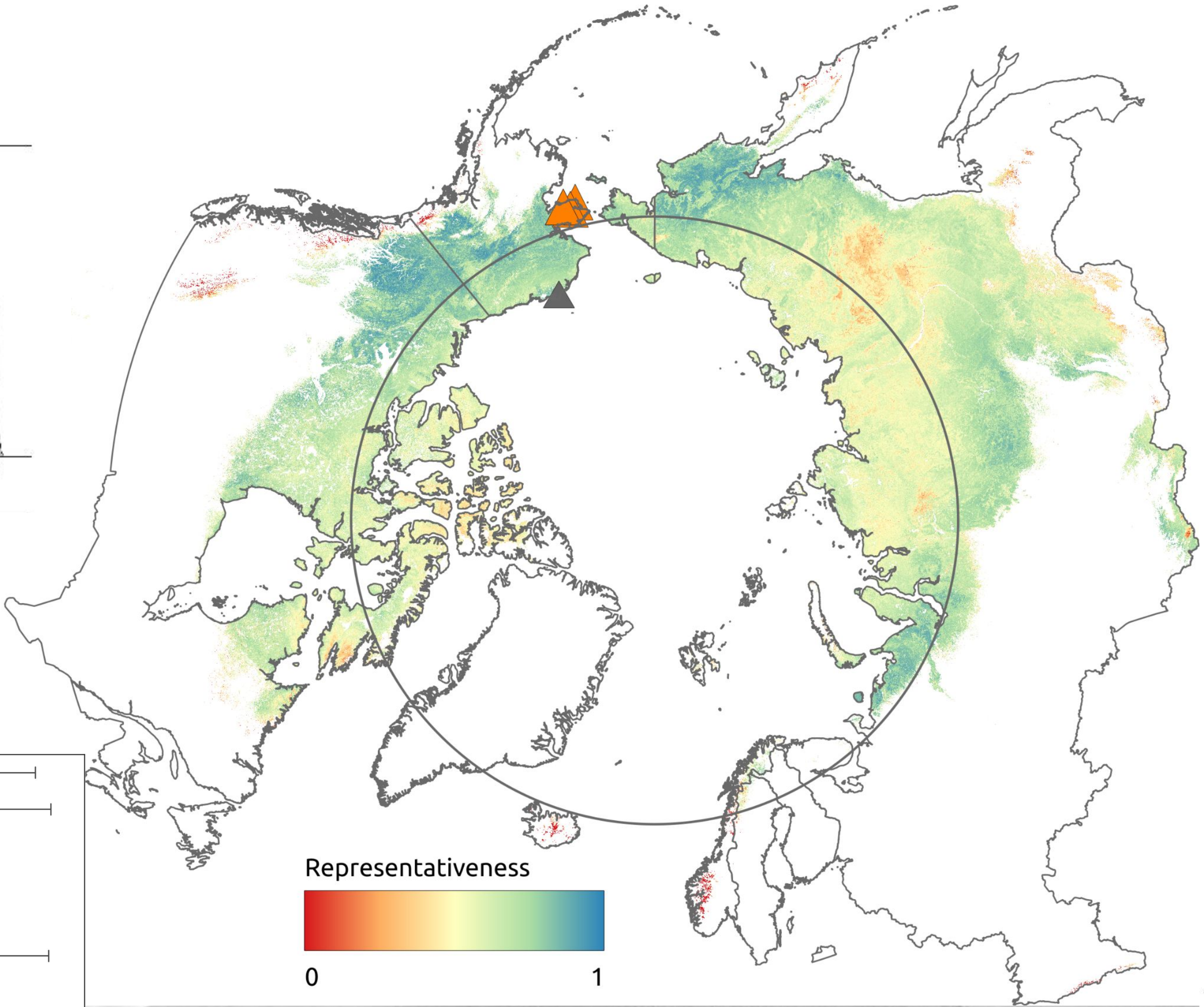


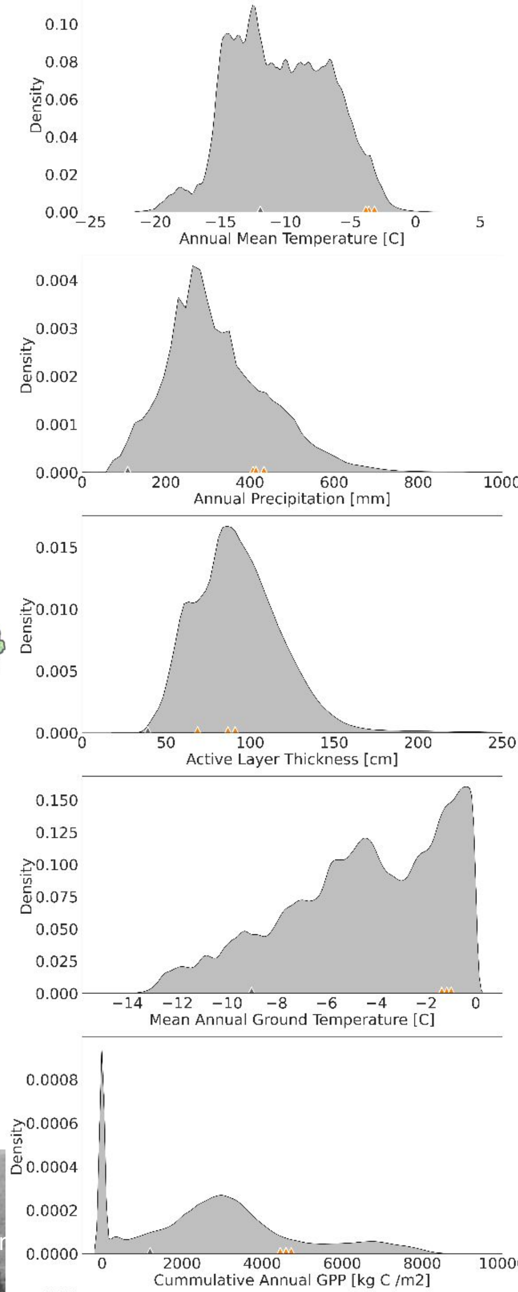


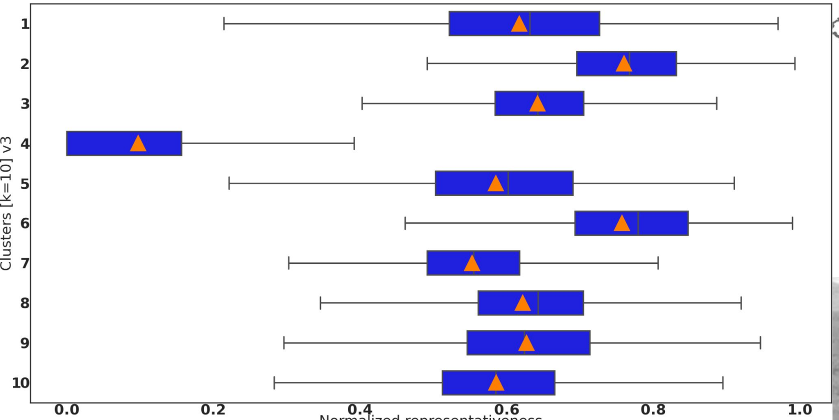


Phases 1–3: Representativeness (Barrow & Seward)

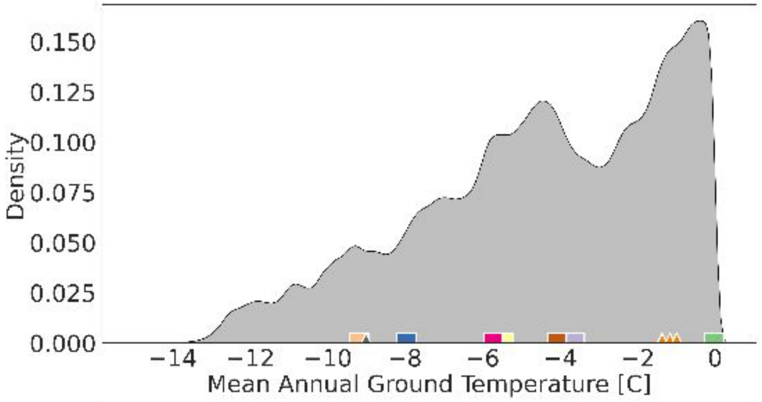


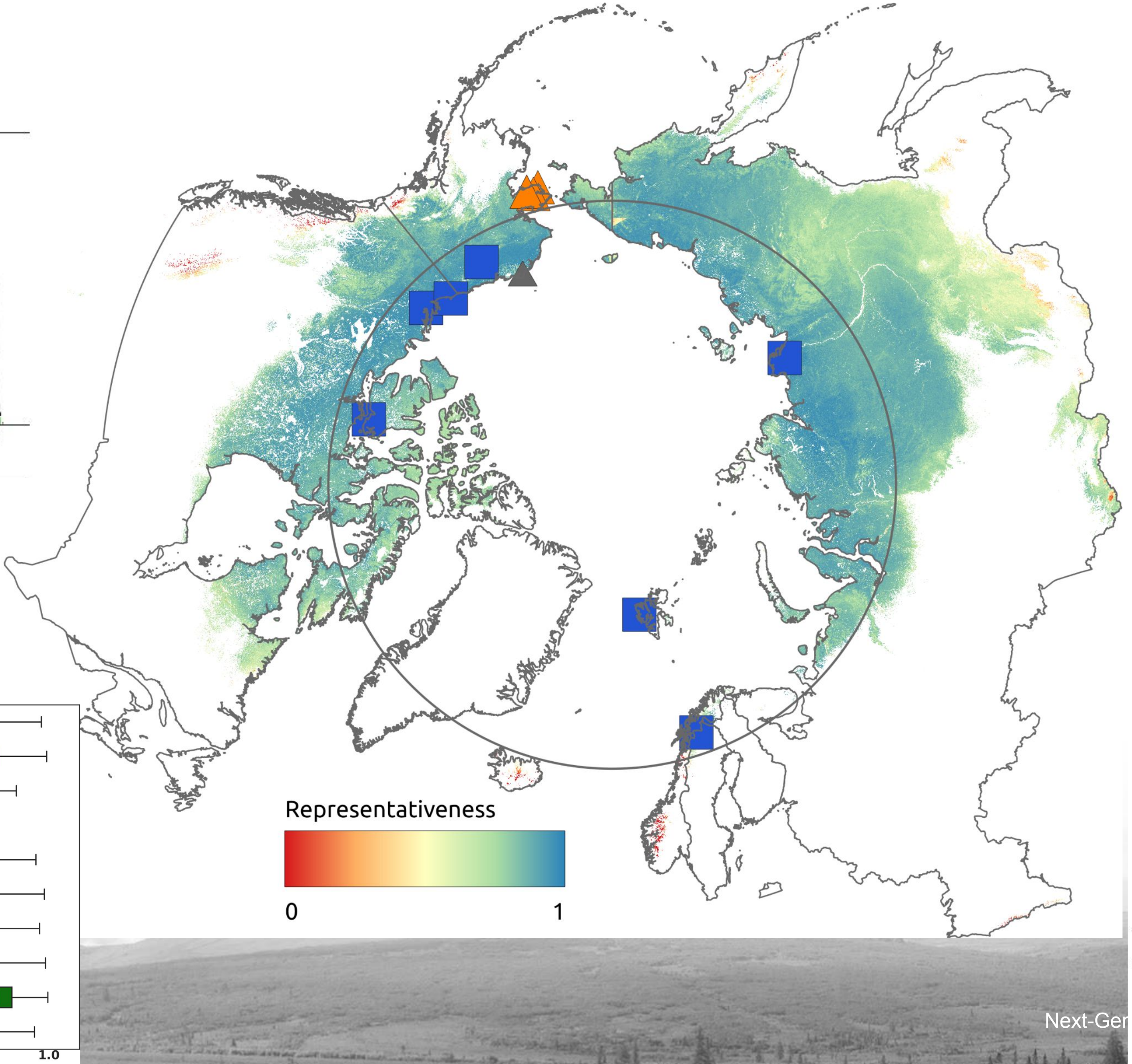


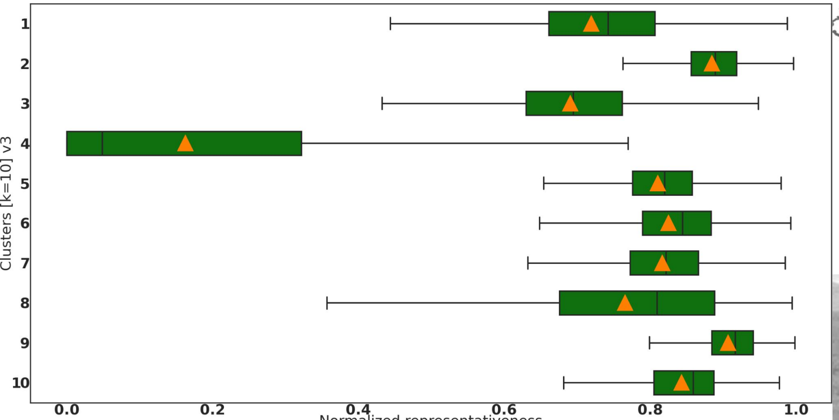


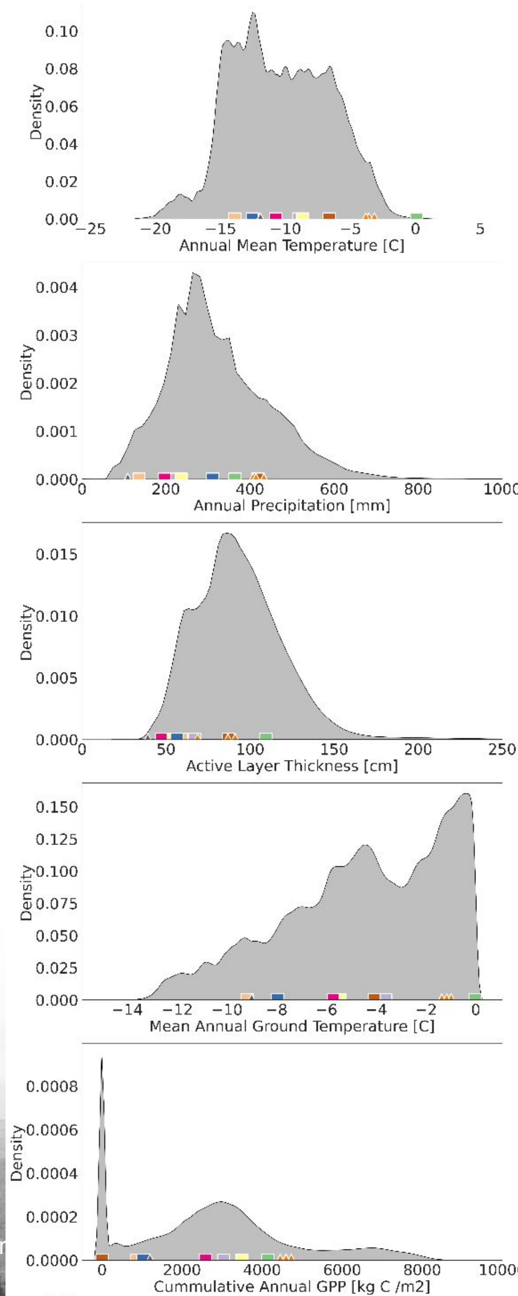


Phases 1–4: Barrow, Seward & Evaluation Sites









Summary

- We employed a cluster analysis method to partition a 22-dimensional environmental phase space spanning the pan-Arctic to create 10 ecoregions
- We selected seven additional “model evaluation sites” to improve the sampling of these environments
- This ecoregion framework, along with data from these partner sites, will be used to provide initialization and validation data for the Arctic-informed E3SM Land Model (ELM)
- A data synthesis cross-cut activity is proposed for Phase 4 to integrate observations across the pan-Arctic