

Mapping Arctic Representativeness and Vegetation using Data Mining and Machine Learning Techniques



Forrest M. Hoffman¹ (forrest@climatemodeling.org), Jitendra Kumar¹, and Zachary Langford²

¹Oak Ridge National Laboratory, Oak Ridge, TN; and ²University of Tennessee, Knoxville, TN



Introduction

The Arctic ecosystem is a large permafrost-dominated region with regional climate, ecohydrology and geomorphology that drive a diverse and heterogeneous distribution of vegetation on the landscape. Using a coordinated modeling and field/laboratory observations approach, NGEET-Arctic is working to improve our understanding of Arctic ecosystems and how they may evolve in the future.

- While resource and logistical constraints limit the extent and frequency of measurements, we developed and employed a systematic sampling strategy to objectively represent environmental variability at appropriate spatial and temporal scales.
- Using machine learning techniques we combined high resolution satellite data with field observations to develop high resolution maps of Arctic vegetation.

Representativeness-based Design of Observation Network

NGEE-Arctic Sampling Network Design in Alaska

Our analysis using a suite of 37 climate, edaphic and permafrost characteristics for the present and projected future conditions (from downscaled GCMs) delineates key ecoregions in Alaska, and also suggests a northward shift in the environmental conditions under various climate change scenarios.

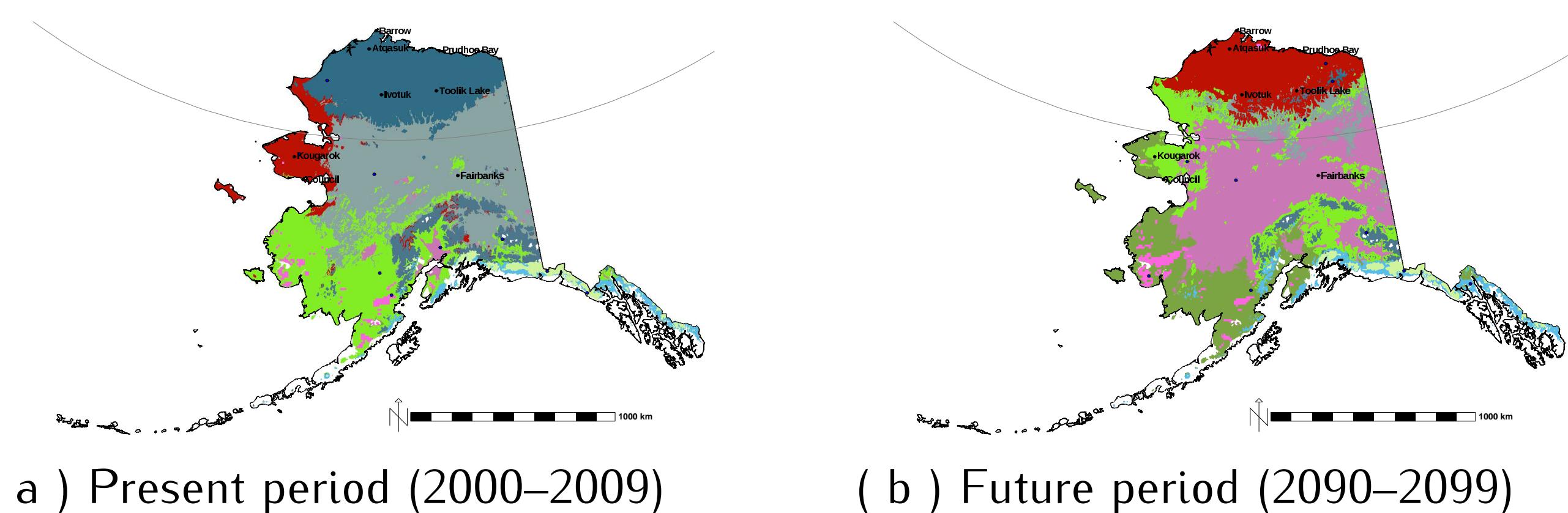


Figure 1: Environmental conditions and ecoregions in Alaska are expected to shift northwards under warming climate

With a combination of a northern site at Barrow, Alaska, and a set of southern sites at Seward Peninsula, Alaska, NGEET-Arctic is employing a space-for-time approach to sample the regions to best understand the current status of vegetation and implications of a warming climate on these sensitive ecosystems.

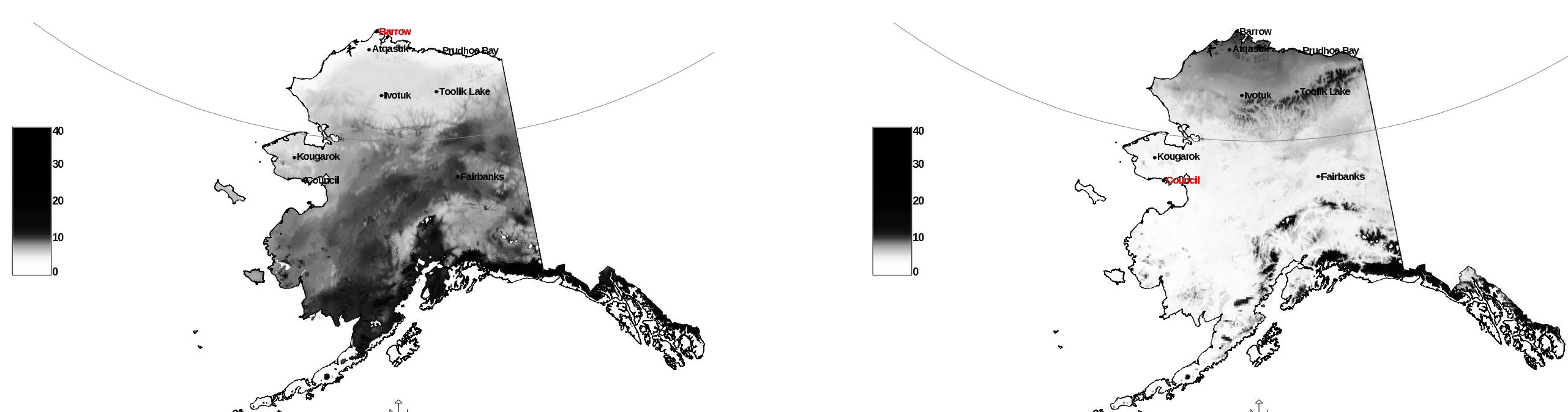


Figure 2: Quantifying representativeness of observations collected at NGEET-Arctic cores sites in the context of the broader landscape

Our representativeness analysis allows for quantitative assessment of the optimality of the observations and samples being conducted and provides a statistical scaling framework for extrapolation of these observations to the larger landscape.

Hoffman, F. M., J. Kumar, R. T. Mills, and W. W. Hargrove (2013), Representativeness-based Sampling Network Design for the State of Alaska, *Landscape Ecol.*, 28(8):1567–1586, doi:10.1007/s10980-013-9902-0.

Scaling Understanding to the Pan-Arctic Region

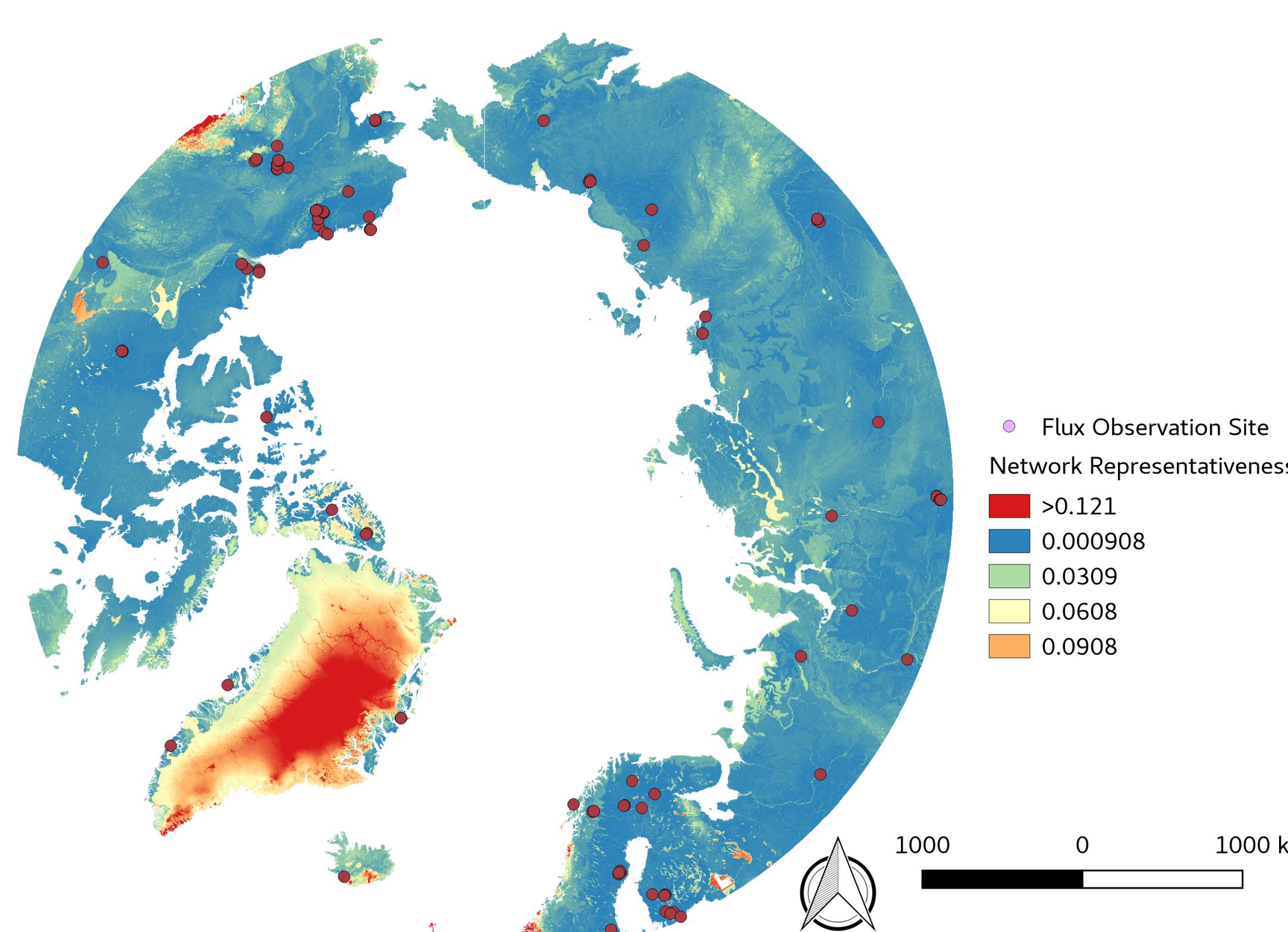


Figure 3: Distributed network of flux observation sites (131) across the Arctic provides potential for scaling our observations and modeling studies to understand the broader Pan-Arctic region.

While core field observation campaigns within NGEET-Arctic are focused at a set of cores sites in Alaska, leveraging the work being conducted by other research groups across the broader Pan-Arctic region allows scaling of observations and models to the larger landscape.

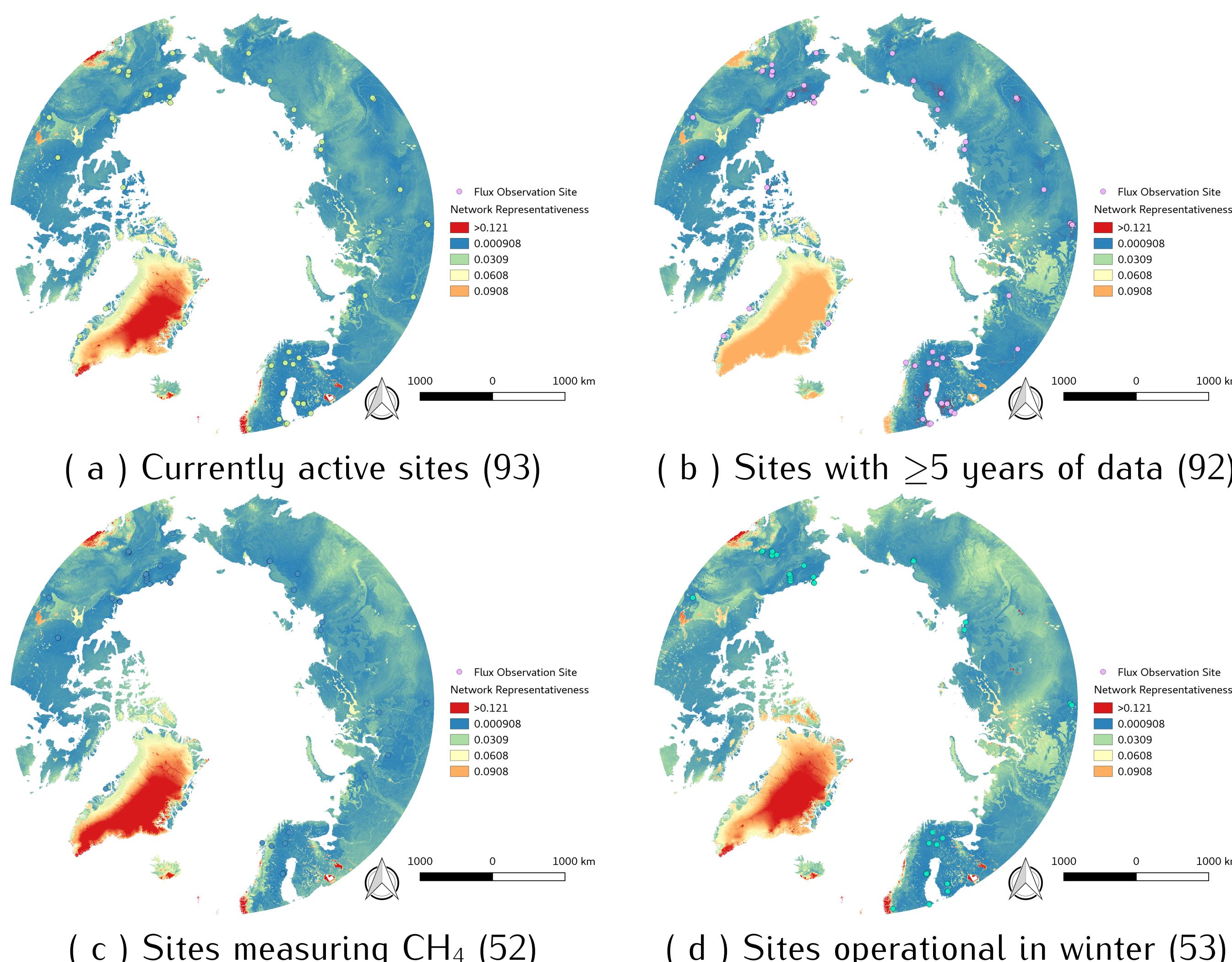


Figure 4: Representativeness of Pan-Arctic network of flux measurement sites

Understanding the inventory of observations available through international partner networks in the Arctic enables development of new synthesis data products and models.

Mapping PFT Distributions at BEO



Figure 5: Vegetation community and distribution show wide variations across polygonal tundra microtopography

Microtopography in continuous permafrost polygonal tundra at the Barrow Environmental Observatory govern the surface and subsurface hydrology and regulate soil moisture and soil temperature, which has direct implications for the vegetation community and distributions at sub-meter scales.

Table 1: Remote sensing data used for vegetation mapping with and without phenology.

Variables	Number of Variables (without phenology, with phenology)	Platform
Elevation	1, 1	LiDAR
TOA Red Band	1, 6	WorldView-2
TOA Blue Band	1, 6	WorldView-2
TOA Green Band	1, 6	WorldView-2
TOA NIR Band	1, 6	WorldView-2
NDVI	1, 6	WorldView-2

We employed high resolution spectral remote sensing and digital elevation models to develop machine learning models using vegetation community data and develop high resolution maps of vegetation community distribution. Capturing vegetation phenology using repeat imagery exploits variations in the timing of green up for different vegetation types, allowing improved accuracy in resulting data products.

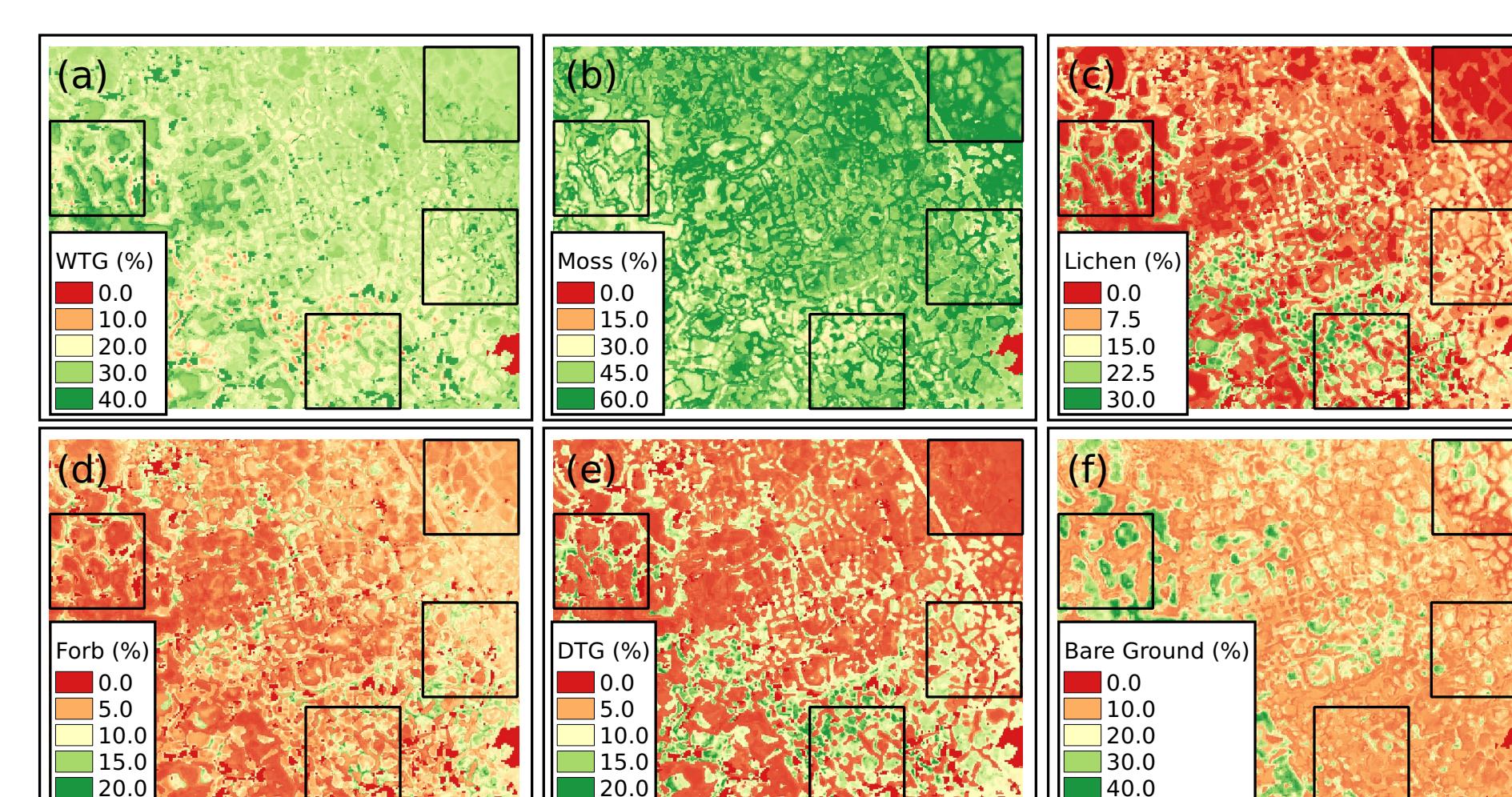


Figure 6: High resolution vegetation maps captures vegetation community and distribution across polygon types

Langford, Z. L., J. Kumar, F. M. Hoffman, R. J. Norby, S. D. Wullschleger, V. L. Sloan, and C. M. Iversen (2016), Mapping Arctic Plant Functional Type Distributions in the Barrow Environmental Observatory Using WorldView-2 and LiDAR Datasets, *Remote Sens.*, 8(9):733, doi:10.3390/rs8090733.

Arctic Vegetation Mapping at Seward Peninsula

Availability of high resolution remote sensing data in the Arctic is often limited due to frequent cloud cover and polar darkness. However, many satellites collect data using different sensor types, at different resolutions and return frequencies. We developed a multi-sensor data fusion approach to combine the information content from a range of sensor types and resolutions available at the Seward Peninsula for improved mapping of vegetation at high resolution.

Table 2: Spectral and topographic variables used in multi-sensor fusion based vegetation classifications

Sensor Group	Predictor Variable	Unit	Date
ALOS-1	HH	γ^0	29 August 2007
PALSAR	HV	γ^0	
SPOT-5	Green, Red, NIR (0.5–0.9 μ m)	DN	June–September 2009 – 2012
IIRSAR	Elevation	m	July 2012
EO-1	198 spectral bands (0.4 to 2.5 μ m)	DN	24 June 2015
Landsat 8	9 spectral bands (0.4 – 2.29 μ m)	DN	17 August 2016

Different satellite sensors are sensitive to different aspects of diverse vegetation on the landscape, and when combined provides a wealth of information.

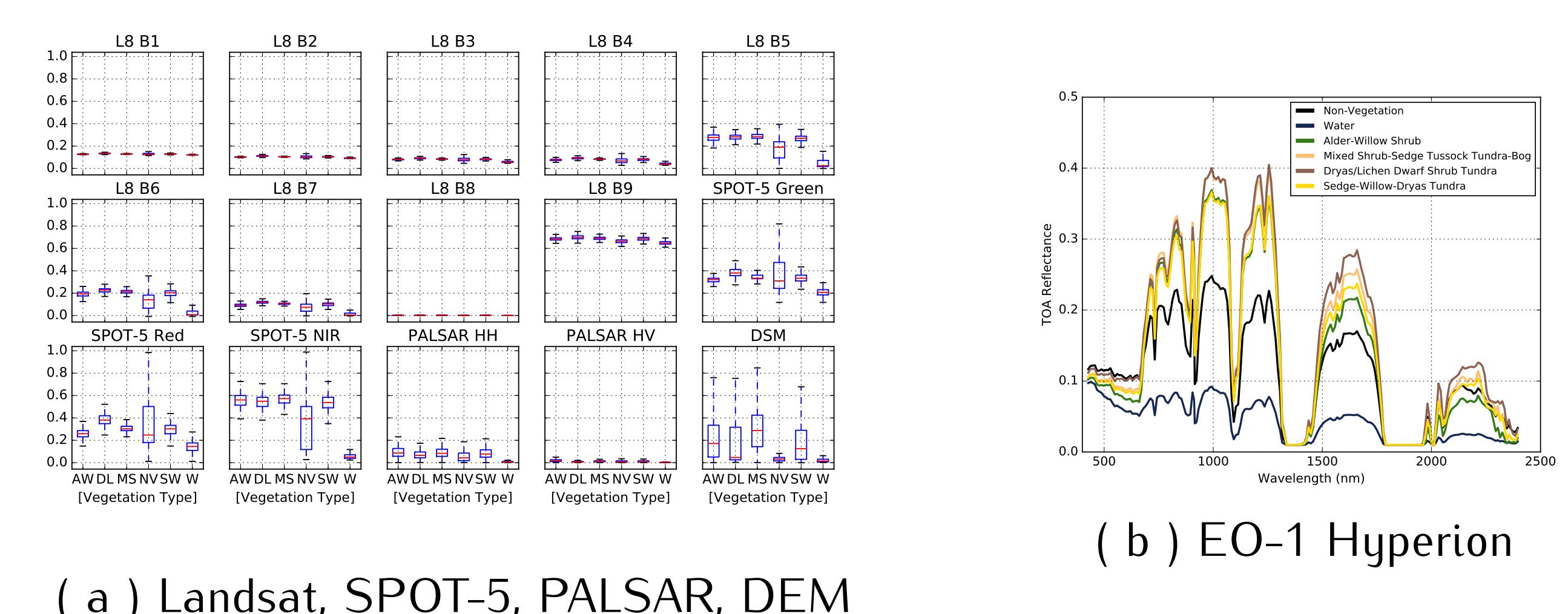


Figure 7: Response of various sensors to vegetation characteristics

We developed Deep Neural Network models and trained them with existing vegetation maps and field based vegetation community data for accurate high resolution maps of vegetation.

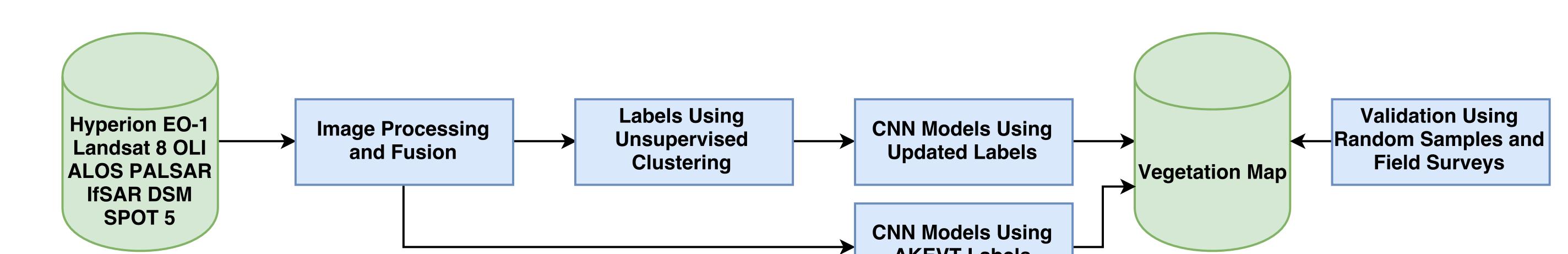


Figure 8: Deep learning-based multi-sensor fusion approach for vegetation mapping

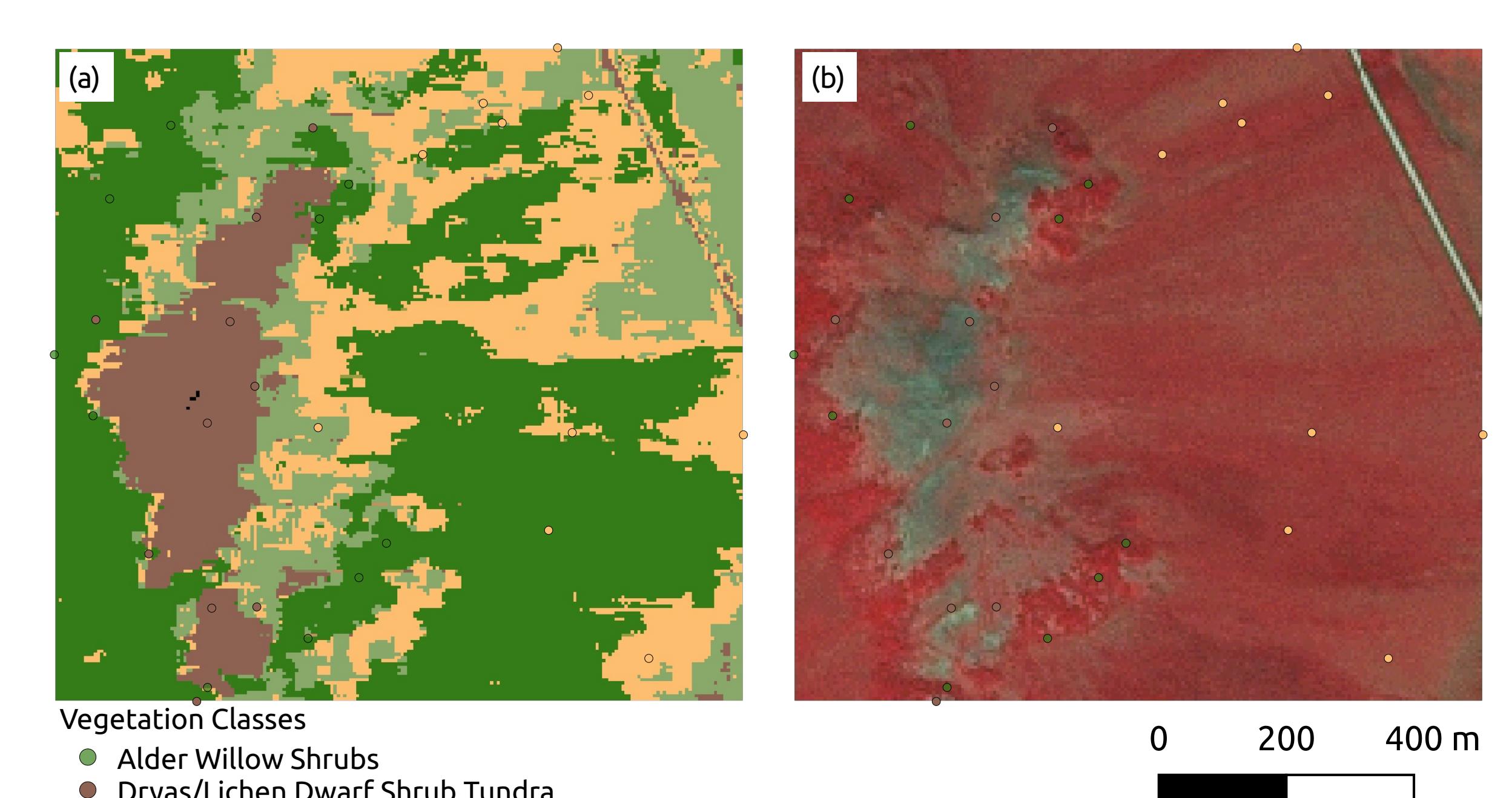


Figure 9: Deep learning-based vegetation map for Kougarok Watershed at Seward Peninsula

Langford, Z. L., J. Kumar, and F. M. Hoffman (2017), Convolutional Neural Network Approach for Mapping Arctic Vegetation using Multi-Sensor Remote Sensing Fusion, *Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW 2017)*, Institute of Electrical and Electronics Engineers (IEEE), Conference Publishing Services (CPS), doi:10.1109/ICDMW.2017.48, data doi:10.5440/1418854.

Acknowledgments

This research was supported by the Next-Generation Ecosystem Experiments (NGEE Arctic) which is sponsored by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the US Department of Energy Office of Science. Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the US Department of Energy under Contract No. DE-AC05-00OR22725.