Poster #1-12

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

Forrest Hoffman[1*], Jitendra Kumar[1], and Zachary L. Langford[2]

[1]Oak Ridge National Laboratory, Oak Ridge, TN
[2]University of Tennessee, Knoxville, TN

Contact: hoffmanfm@ornl.gov

BER Program: TES
Project: NGEE Arctic
Project Website: https://ngee-arctic.ornl.gov/

Understanding and modeling the evolution of ecosystem processes and land-atmosphere interactions across heterogeneous landscapes in the Arctic requires accurate characterization of vegetation and soil properties at appropriate spatial and temporal scales. However, resource and logistical constraints limit the frequency and extent of environmental observations, particularly in the Arctic, necessitating the use of high-resolution remote sensing imagery and the development of a systematic sampling strategy to maximize coverage and objectively represent environmental variability at desired scales. We have developed and applied data mining and machine learning methods for stratifying sampling domains, informing site selection, determining the representativeness of measurement sites and networks, and mapping vegetation from sparse samples and from multi-sensor remote sensing fusion. In the first study, multivariate spatiotemporal clustering was applied to down-scaled general circulation model results and data for the State of Alaska to define multiple sets of bioclimatic ecoregions across two decadal time periods. Maps of ecoregions for the present (2000-2009) and future (2090-2099) were produced, showing how combinations of 37 bioclimatic characteristics are distributed and how they may shift in the future. Representative sampling locations were identified on present and future ecoregion maps, a representativeness metric was developed, and representativeness maps for eight candidate sampling locations were produced to inform future site selection. In the second study, we applied data mining to scale up sparse vegetation samples by using multi-spectral remote sensing, and we demonstrated that including phenological information significantly improves the quality of vegetation type predictions. In a third study, we applied a convolutional neural network (CNN) approach to map Arctic vegetation distributions for the Kougarok area on the Seward Peninsula by fusing data from multiple sensors. We tested both supervised and unsupervised classification techniques over a 344 sq km area to generate vegetation classifications at 5 m and 12.5 m resolutions. We subsequently compared two CNN approaches to predict vegetation classes for every pixel within the study area in comparison with an existing vegetation classification map. In a fourth, ongoing study, we combined a digital elevation model and other data to stratify sampling domains across the entire Arctic region. We calculated pan- Arctic representativeness based on a suite of sampling locations operated by international research institutes. This presentation will highlight all of these research applications of large- scale data analytics methods and demonstrate their utility for coordinated measurement and modeling activities in the Next Generation Ecosystem Experiments (NGEE) in the Arctic.