Artificial Intelligence for Exploring Climate Change Mitigation Strategies and Advancing Earth System Prediction

Forrest M. Hoffman
forrest@climatemodeling.org

Oak Ridge National Laboratory and University of Tennessee

Presentation to UTK Bredesen Center Recruits

Oak Ridge National Laboratory

February 16, 2023

Forrest M. Hoffman, Computational Earth System Scientist

- Group Leader for the ORNL Computational Earth Sciences Group
- 34 years at ORNL in Environmental Sciences Division, then Computer Science and Mathematics Division, and now Computational Sciences and Engineering Division
- Develop and apply Earth system models to study global biogeochemical cycles, including terrestrial & marine carbon cycle
- Investigate methods for reconciling uncertainties in carbon–climate feedbacks through comparison with observations
- Apply artificial intelligence methods (machine learning and data mining) to environmental characterization, simulation, & analysis
- Joint Faculty, University of Tennessee, Knoxville, Department of Civil & Environmental Engineering

Oak Ridge National Laboratory

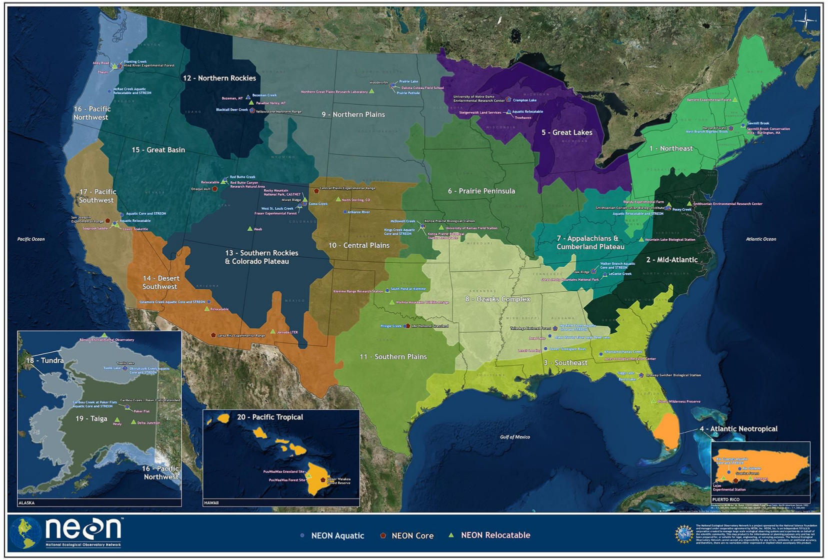
The University of Tennessee Knoxville

Introduction

- Observations of the Earth system are increasing in spatial resolution and temporal frequency, and will grow exponentially over the next 5–10 years
- With Exascale computing, simulation output is growing even faster, outpacing our ability to analyze, interpret and evaluate model results
- Explosive data growth and the promise of discovery through data-driven modeling necessitate new methods for feature extraction, change/anomaly detection, data assimilation, simulation, and analysis

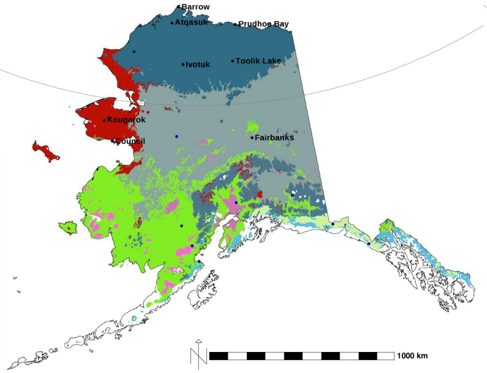
Frontier at Oak Ridge National Laboratory is the #1 fastest supercomputer on the [TOP500](TOP500) List and the first supercomputer to break the exaflop barrier (Nov 14, 2022).

Sampling Network Design

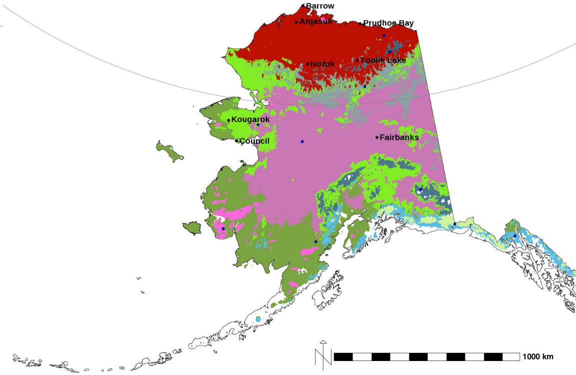


NSF's NEON Sampling Domains

Gridded data from satellite and airborne remote sensing, models, and synthesis products can be combined to design optimal sampling networks and understand representativeness as it evolves through time



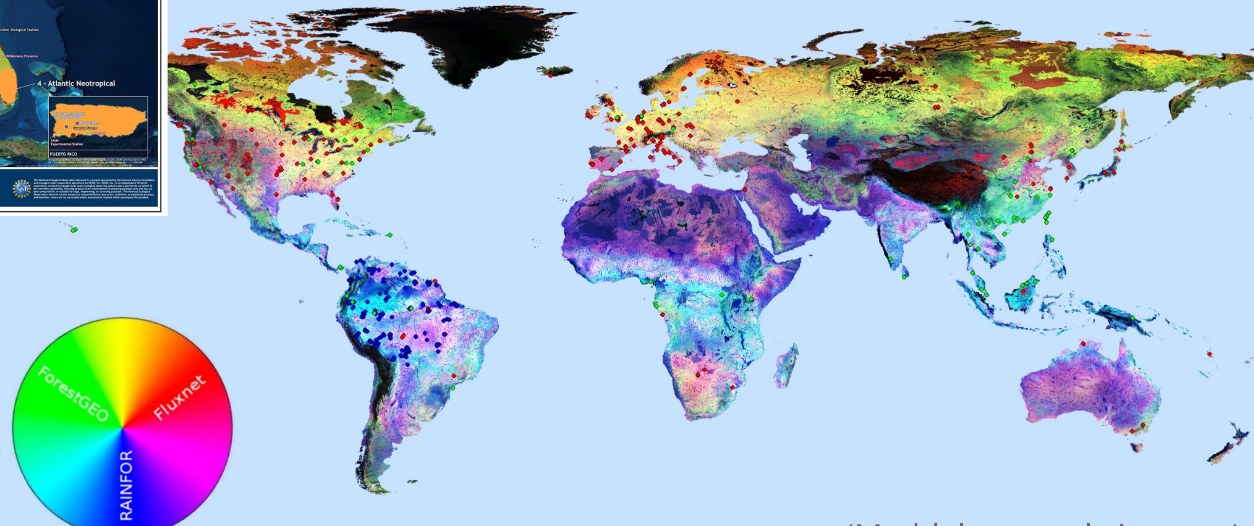
2000–2009



2090–2000

(Hoffman et al., 2013)

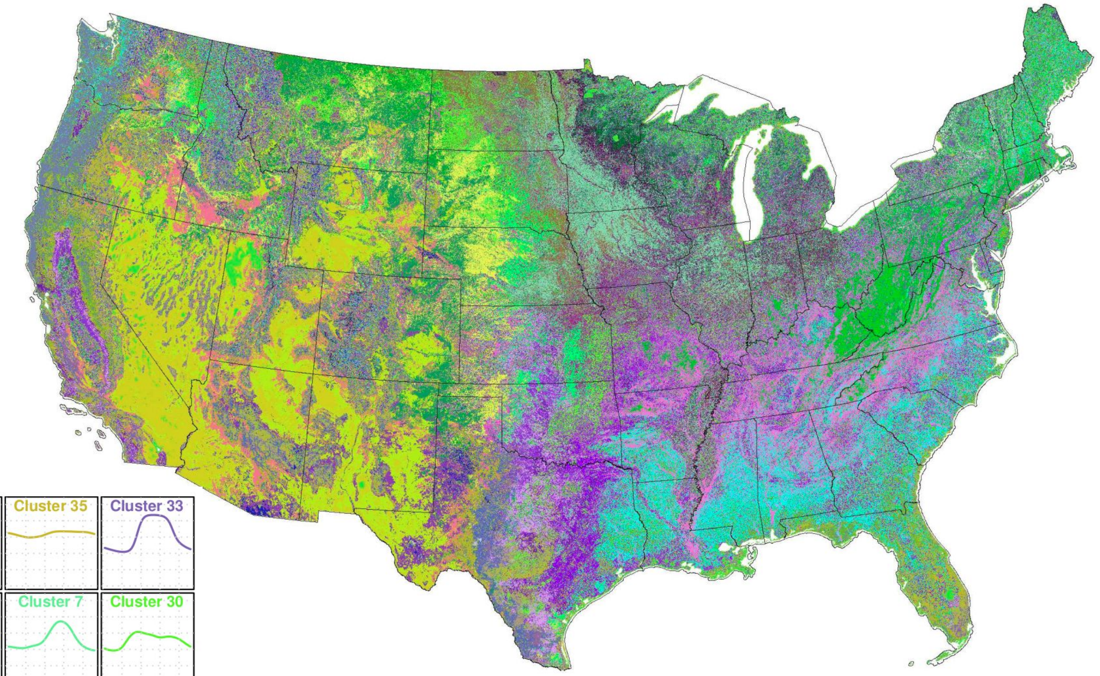
Triple-Network Global Representativeness

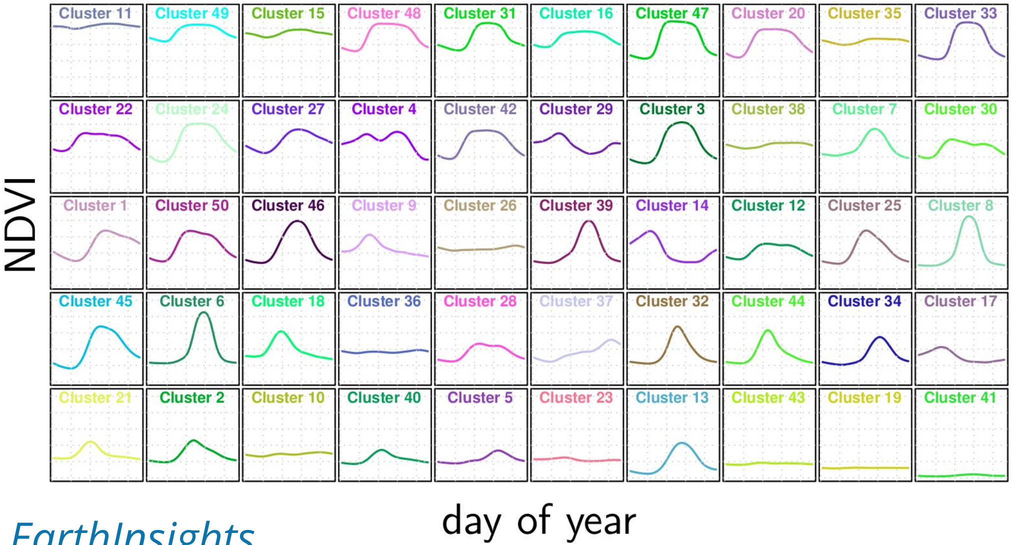


(Maddalena et al., in prep.)

50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI
Every 8 days (46 images/year)
Clustered from year 2000 to present



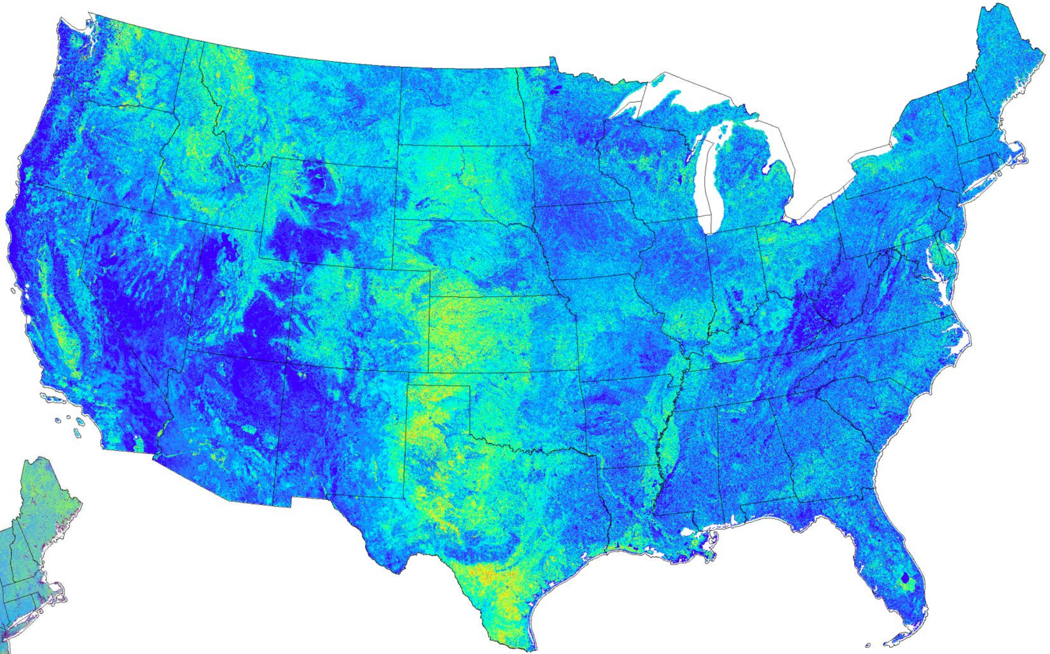


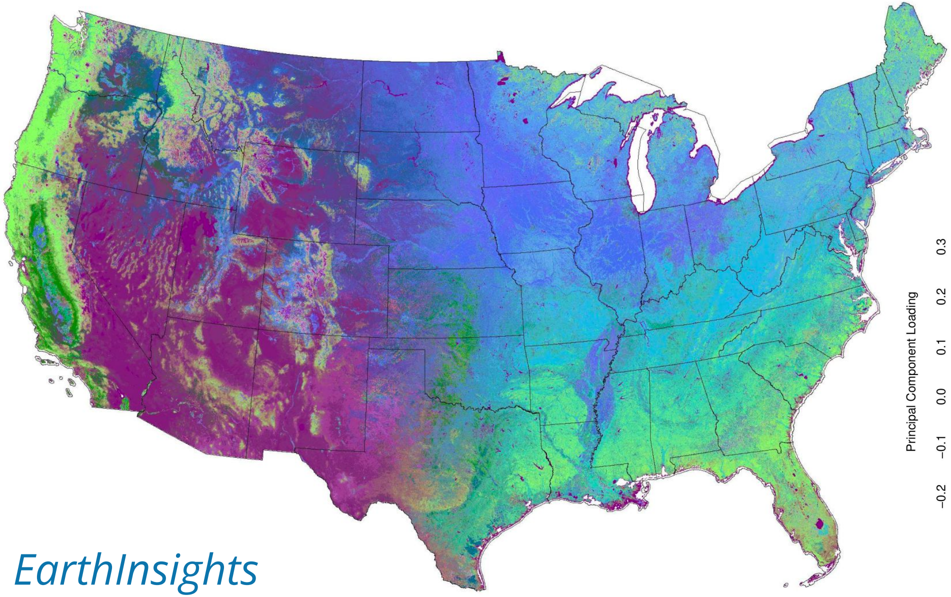
50 Phenoregion Prototypes (Random Colors)

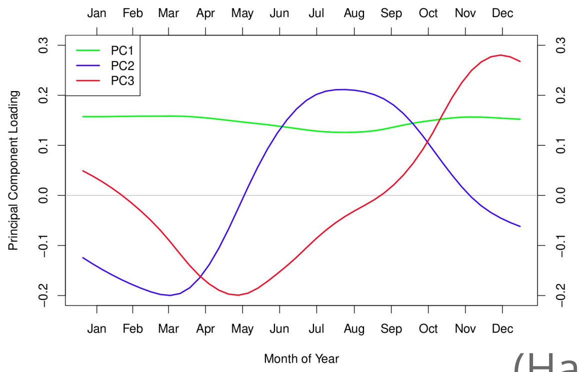
EarthInsights

(Hargrove et al., in prep.)

50 Phenoregions Persistence and 50 Phenoregions Max Mode (Similarity Colors)







Principal Components Analysis

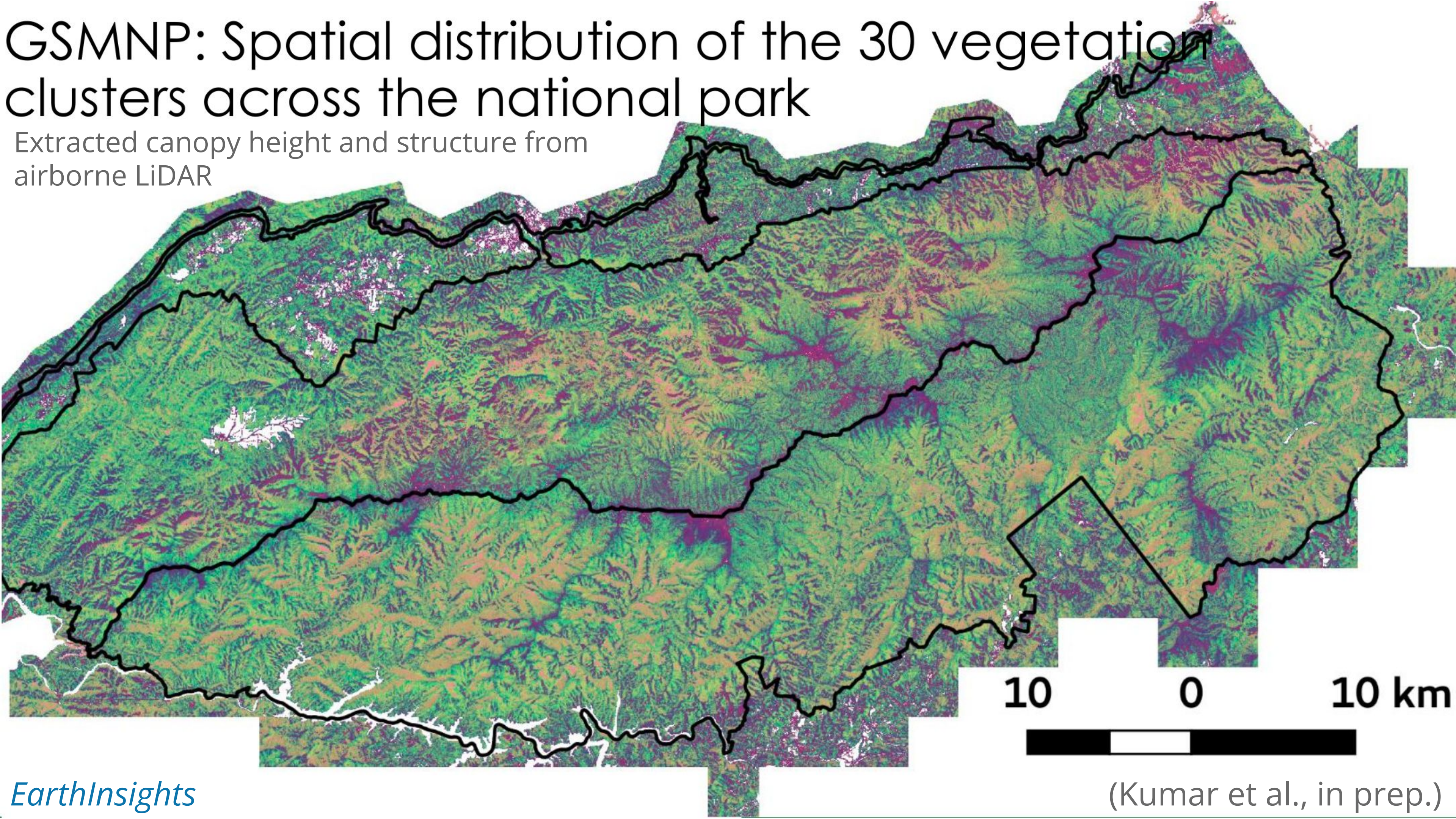
PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

EarthInsights

(Hargrove et al., in prep.)

GSMNP: Spatial distribution of the 30 vegetation clusters across the national park

Extracted canopy height and structure from airborne LiDAR



EarthInsights

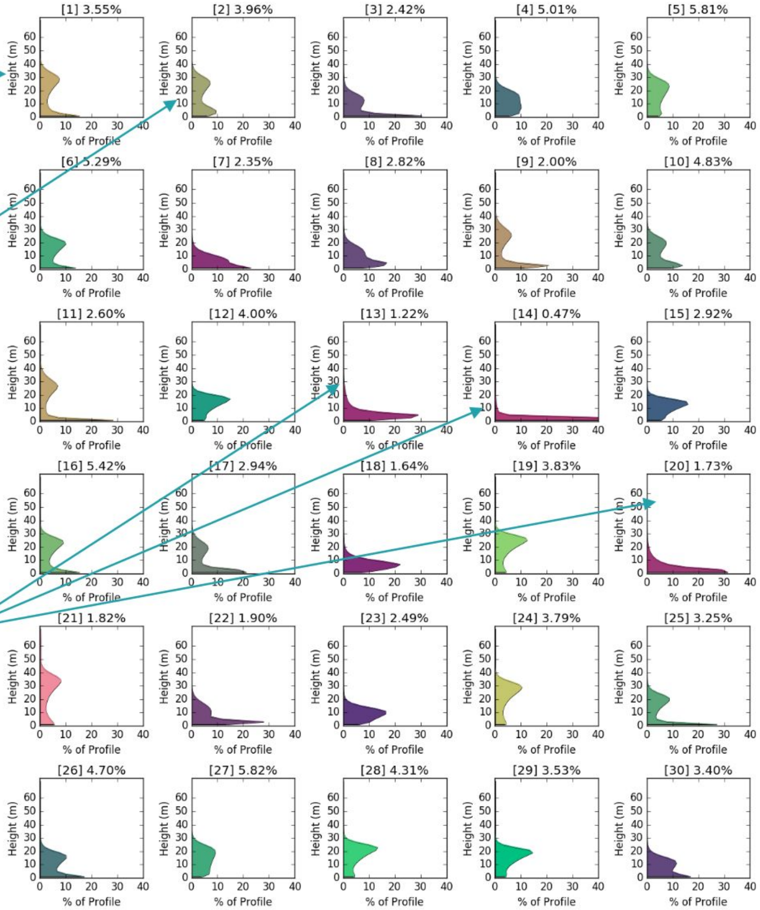
(Kumar et al., in prep.)

GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

forests with slightly lower mean height with dense understory vegetation

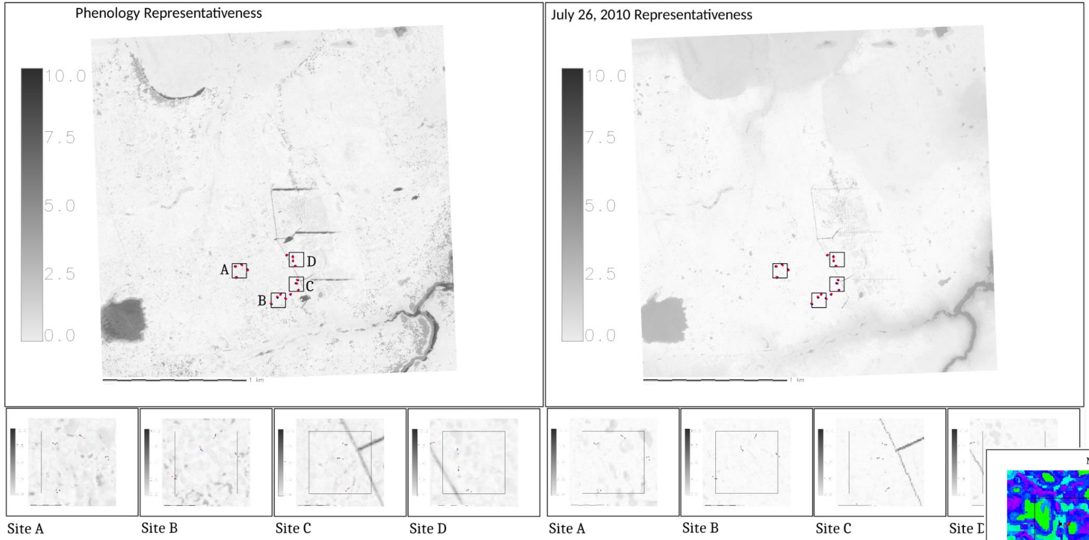
low height grasslands and heath balds that are small in area but distinct landscape type



EarthInsights

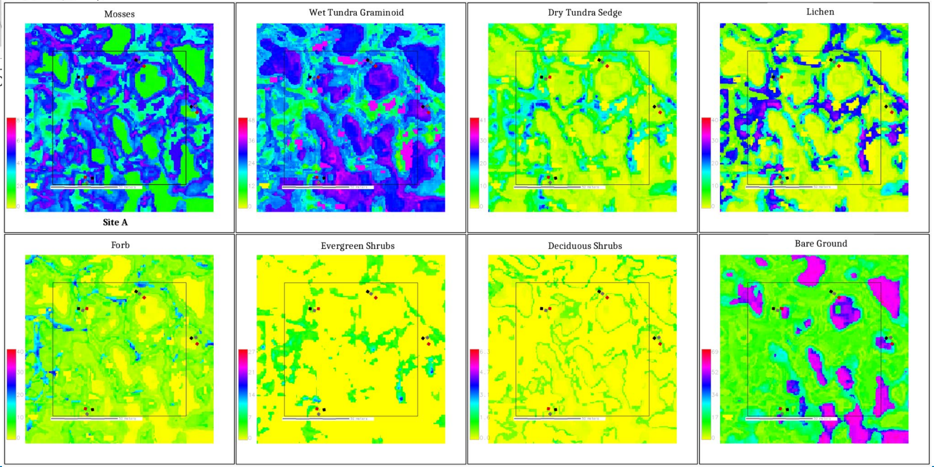
(Kumar et al., in prep.)

Vegetation Distribution at Barrow Environmental Observatory



Representativeness map for vegetation sampling points in sites A, B, C, and D with phenology (left) and without (right) from WorldView2 multispectral imagery for the year 2010 and LiDAR data

Example plant functional type (PFT) distributions scaled up from vegetation sampling locations

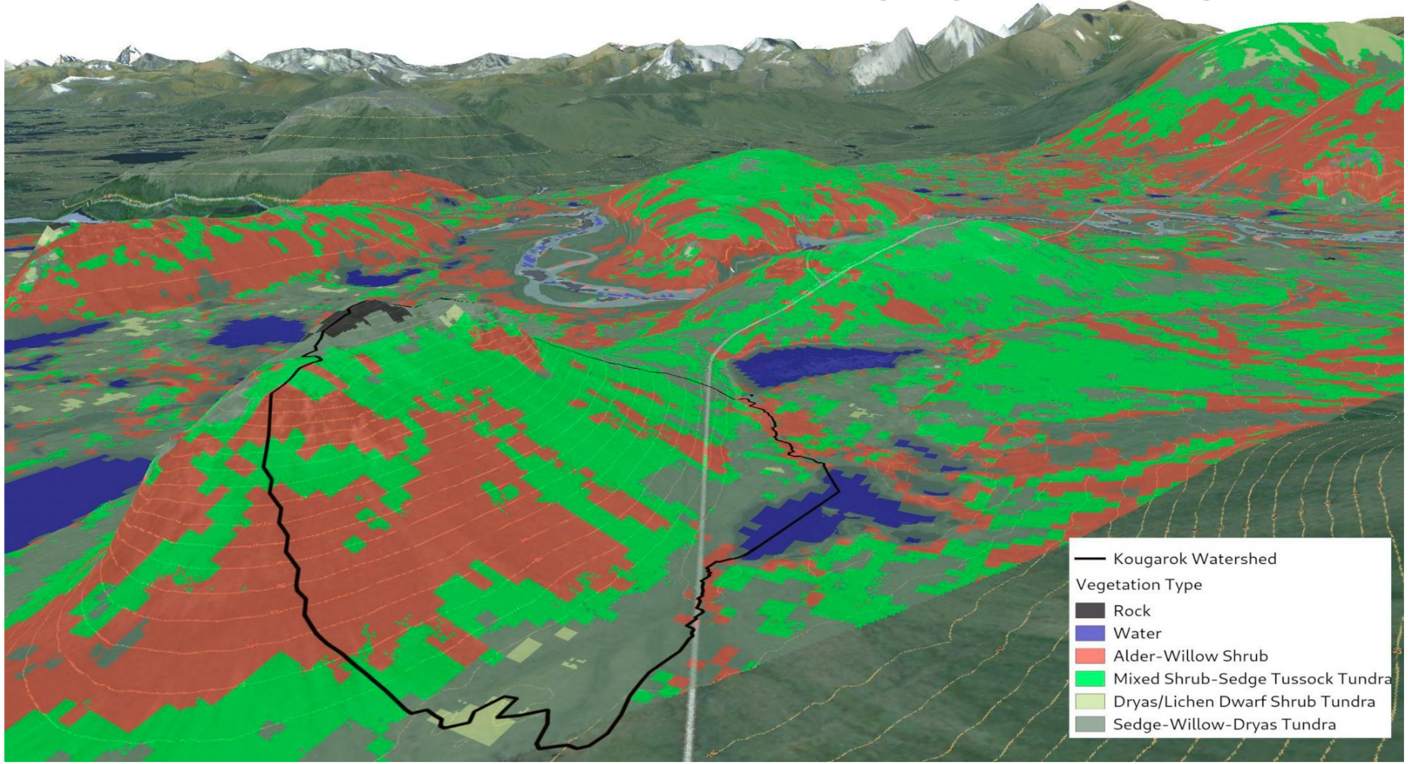


In situ data from field measurement activities inform the development of wide-scale maps of vegetation distribution through inference using remote sensing data as surrogate variables, and relationships with environmental controls can be extracted

Langford, Z. L., et al. (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 from Multi-Sensor Fusion

Used Hyperion Multispectral and IfSAR-derived Digital Elevation Model, applied cluster analysis, and trained a convolutional neural network (CNN) with Alaska Existing Vegetation Ecoregions (AKEVT)



Langford, Z. L., et al. (2019), Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks, *Remote Sens.*, 11(1):69, doi:[10.3390/rs11010069](https://doi.org/10.3390/rs11010069).

Satellite Data Analytics Enables Within-Season Crop Identification

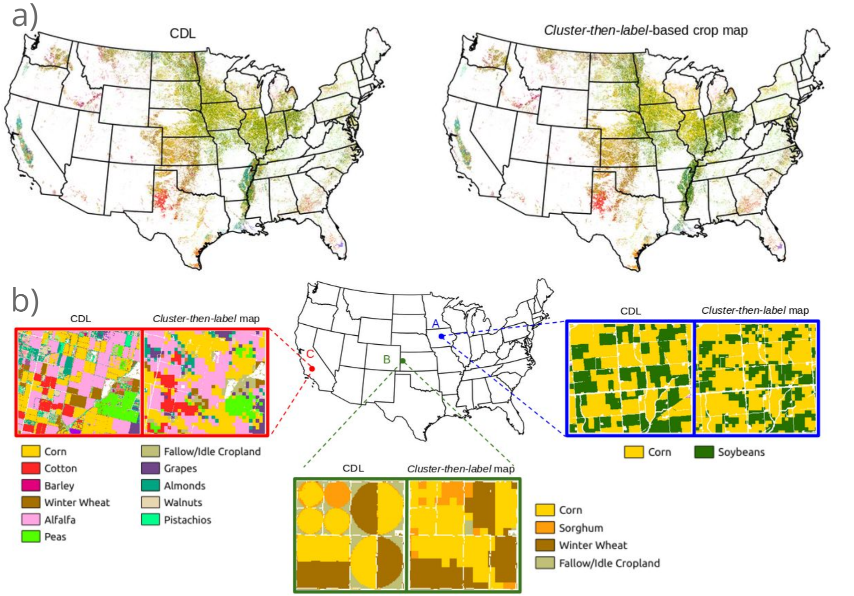
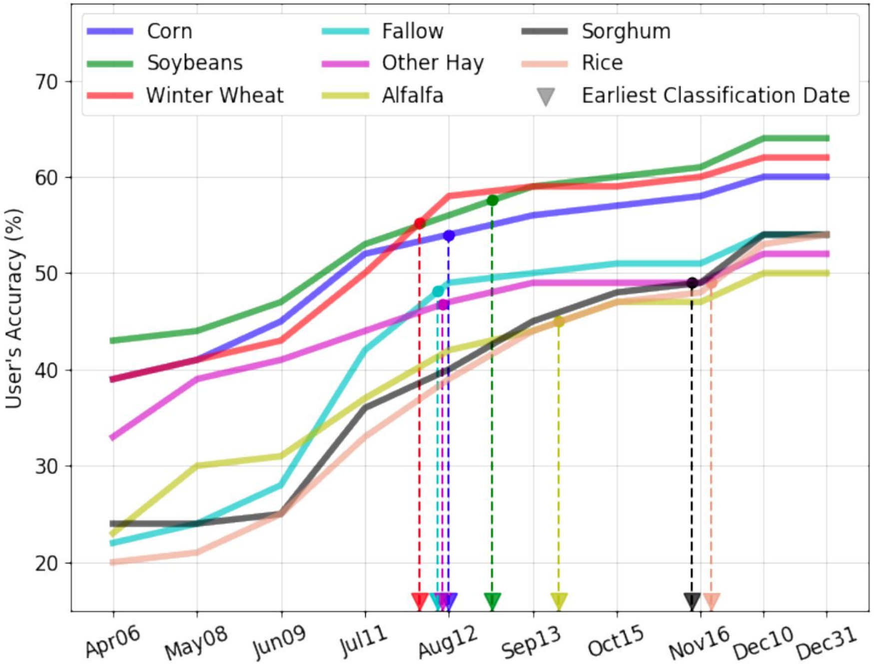


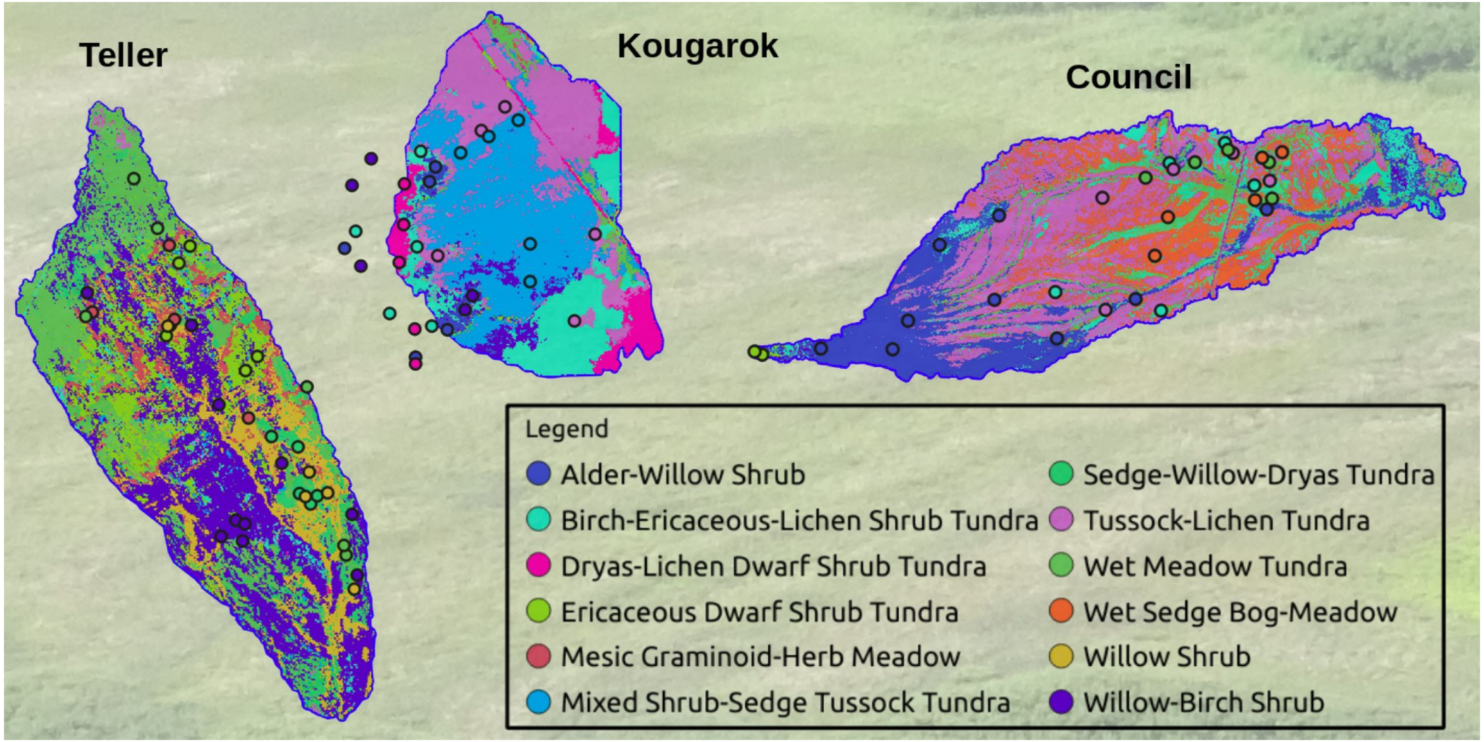
Figure: a) Comparison of cluster-then-label crop map with USDA Crop Data Layer (CDL) shows similar patterns at continental scale. b) Good spatial agreement is found at three selected regions, but cluster-then-label crop maps lack sharpness at field boundaries due to coarser resolution of MODIS data.

Earliest date for crop type classification



Konduri, V. S., J. Kumar, W. W. Hargrove, F. M. Hoffman, and A. R. Ganguly (2020), Mapping Crops Within the Growing Season Across the United States, *Remote Sens. Environ.*, 251, 112048, doi:10.1016/j.rse.2020.112048.

Watershed-Scale Plant Communities Determined from DNN and AVIRIS-NG

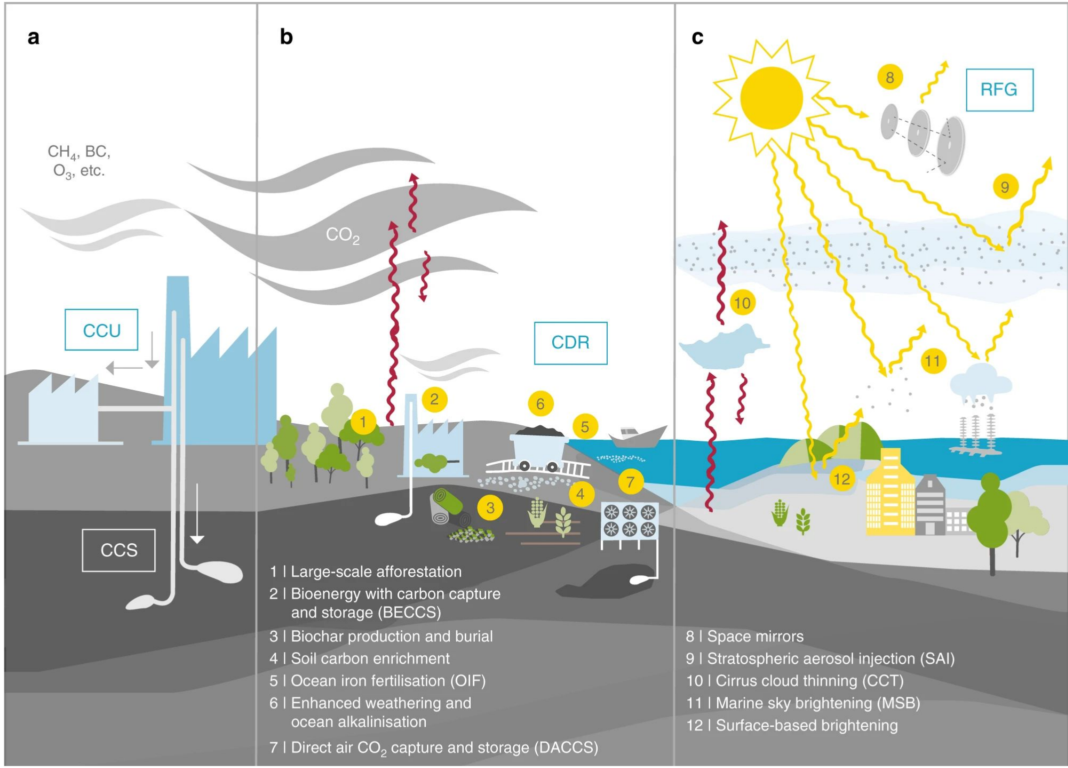


At the watershed scale, vegetation community distribution follows topographic and water controls. At a fine scale, nutrients limit the distribution of vegetation types.

(Konduri et al., in prep.)

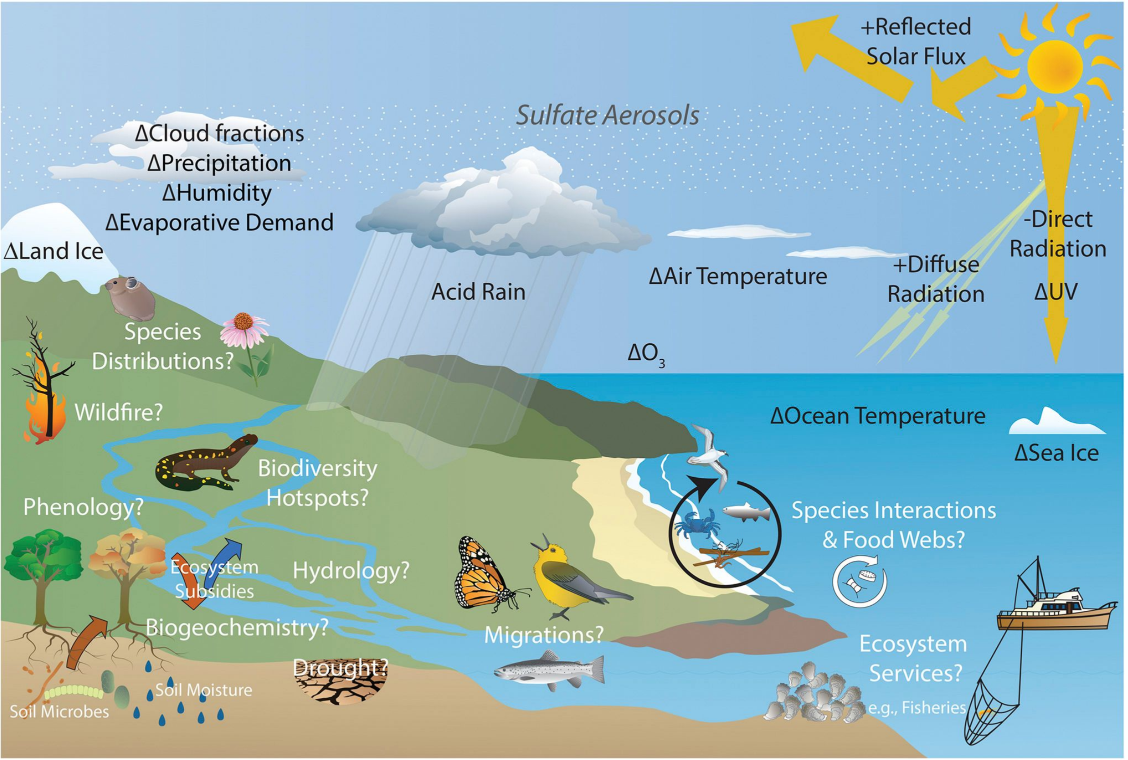
Climate Change Mitigation through Climate Intervention

- The increasing severity of extreme events and wildfire is threatening utilities, built infrastructure, and economic & national security
- Loss of life and property is motivating consideration of *climate intervention* or *geoengineering*
- In addition to *carbon dioxide removal (CDR)* through *direct air capture (DAC)* and other means, interest is growing in reducing or stabilizing Earth's surface temperature
- *Solar radiation management (SRM)* is an approach to partially reduce warming, and *stratospheric aerosol intervention (SAI)* by injecting sulfur into the lower stratosphere is considered the most feasible scheme



A wide variety of natural solutions and geoengineering techniques are proposed for mitigating the effects of climate change. Adopted from Lawrence et al. (2018).

Potential Ecological Impacts of Climate Intervention



Although some effects of SRM with SAI on climate are known from certain SAI scenarios, the effects of SAI on ecological systems are largely unknown. Adopted from Zarnetske et al. (2021).

- While climate research has focused on predicted climate effects of SRM, few studies have investigated impacts that SRM would have on ecological systems
- Impacts and risks posed by SRM would vary by implementation scenario, anthropogenic climate effects, geographic region, and by ecosystem, community, population, and organism
- A transdisciplinary approach is essential, and **new modeling paradigms are required**, to represent complex interactions across Earth system components, scales, and ecological systems

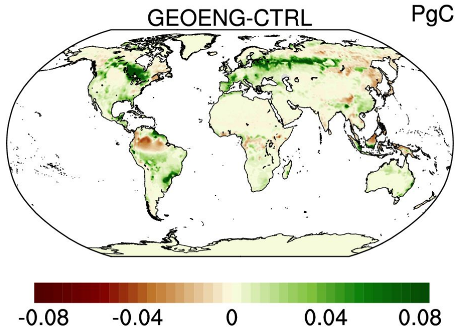
Geoengineering Increases the Global Land Carbon Sink

Objective: To examine stratospheric aerosol intervention (SAI) impacts on plant productivity and terrestrial biogeochemistry.

Approach: Analyze and compare simulation results from the Stratospheric Aerosol Geoengineering Large Ensemble (GLENS) project from 2010 to 2097 under RCP8.5 with and without SAI.

Results/Impacts: In this scenario, SAI causes terrestrial ecosystems to store an additional 79 Pg C globally as a result of lower ecosystem respiration and diminished disturbance effects by the end of the 21st century, yielding as much as a 4% reduction in atmospheric CO_2 mole fraction that progressively reduces the SAI effort required to stabilize surface temperature.

Yang, C.-E., **F. M. Hoffman**, D. M. Ricciuto, S. Tilmes, L. Xia, D. G. MacMartin, B. Kravitz, J. H. Richter, M. Mills, and J. S. Fu (2020), Assessing Terrestrial Biogeochemical Feedbacks in a Strategically Geoengineered Climate, *Environ. Res. Lett.*, doi:10.1088/1748-9326/abacf7.



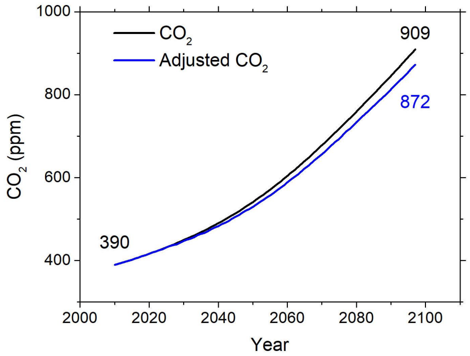
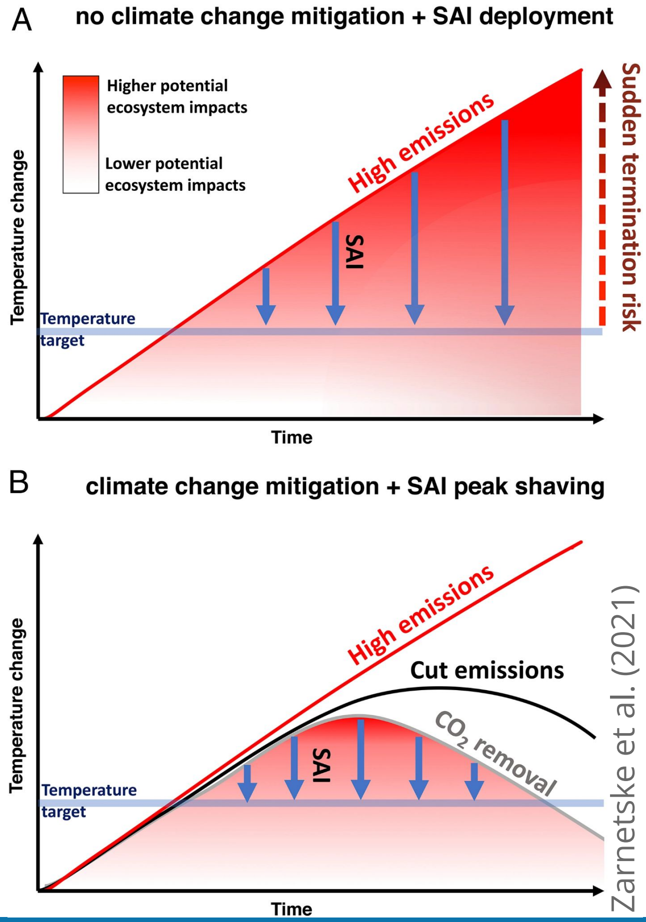


Figure: The larger sink under SAI increased land C storage by 79 Pg C by 2097, which would reduce the projected atmospheric CO_2 level.

Exploring Feedbacks of SAI

- To fill research gaps in understanding Earth system feedbacks of SAI on ecosystems, we are conducting a series of increasingly complex geoengineering simulations with **DOE's Energy Exascale Earth System Model (E3SM)**
- **Simulations will mimic effects of CDR, SAI, and CDR plus SAI**
- Start with SSP5-3.4-OS mid-range overshoot CO_2 trajectory from CMIP6, which prescribes a drawdown of CO_2
- Global surface temperatures will rise by >2.5°C around 2040, **above the 2°C threshold that may induce irreversible impacts**
- Next, introduce SAI to simultaneously cool the surface until drawdown is sufficient to assure < 2°C warming, called **temperature "peak shaving"**
- To quantify feedbacks from reducing, not increasing, atmospheric CO_2, **but may not capture all the as yet unobserved processes**



Zarnetske et al. (2021)

Leveraging Advances in Machine Learning for Earth Sciences

Existing machine learning techniques can improve understanding of biospheric processes and representation in Earth system models

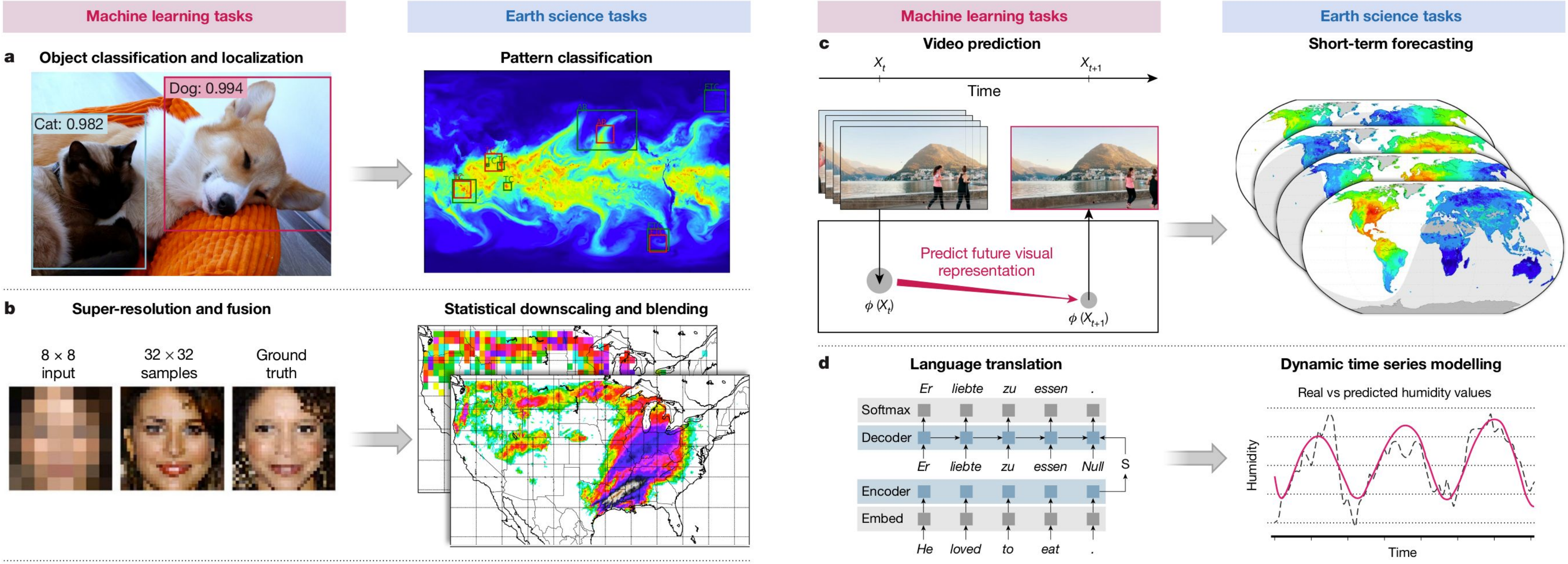
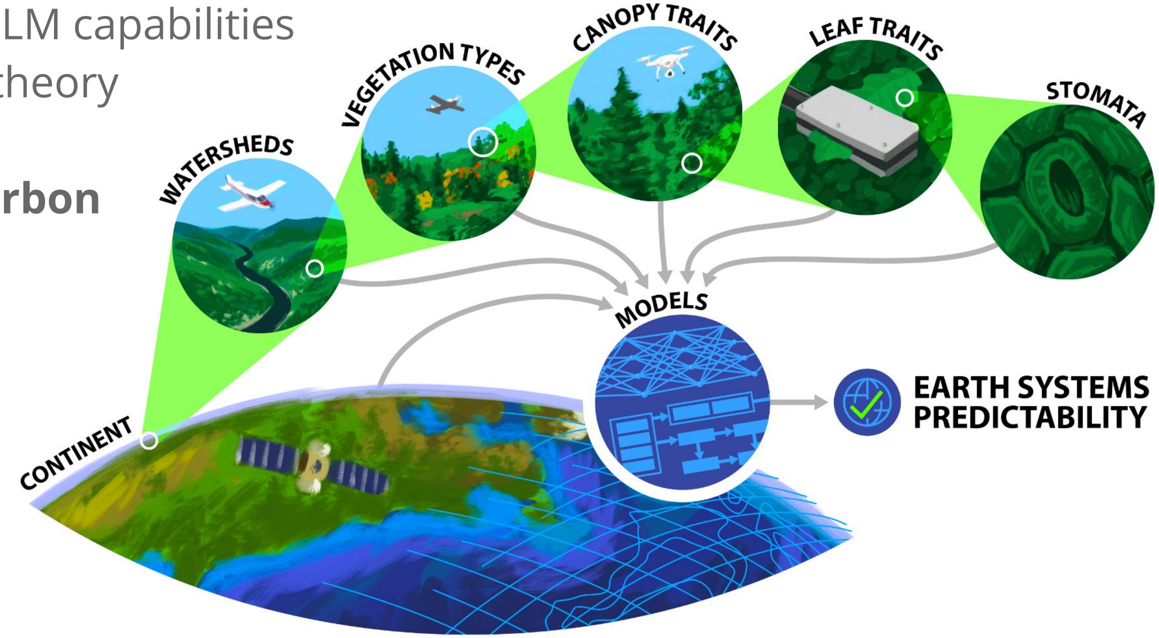


Figure 2 in Reichstein et al. (2019)

Machine Learning for Understanding Biospheric Processes

- Widening adoption of deep neural networks and growth of climate data are fueling interest in AI/ML for use in weather and climate and Earth system models
- ML potential is high for improving predictability when (1) *sufficient data are available for process representations* and (2) *process representations are computationally expensive*
- Example methods for improving ELM capabilities by exploring ML and information theory approaches:
  - **Soil organic carbon & radiocarbon**
  - **Wildfire**
  - **Methane emissions**
  - **Ecohydrology**
- All of these applications involve unresolved, subgrid-scale processes that strongly influence results at the largest scales



AI-Constrained Ecohydrology for Improving Earth System Predictions

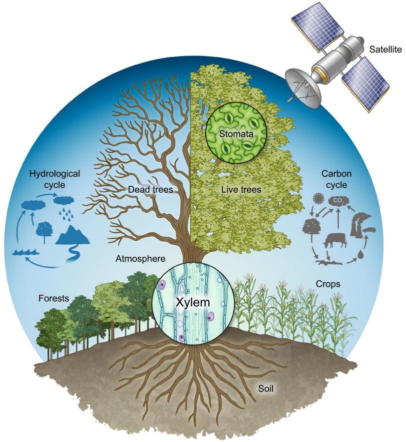
Collaboration among ORNL, LANL, Penn State, et al.

Contact: Forrest M. Hoffman

Project to prototype machine learning-based parameterizations for stomatal conductance and photosynthesis

- Photosynthesis is a computationally expensive part of land models and leaf-level flux and phenology data are available
- Use combinations of leaf-level and plant hydrodynamics data to build ML models of C_3, C_4, and CAM vegetation
- Investigate ML approaches for scaling to canopies and watersheds
- Prototype hybrid ML-/process-based components within the E3SM Land Model (ELM)
- Future efforts:
  - Conduct regional and global simulations to benchmark different combinations of process-based and ML modules
  - Explore approaches for building hybrid modeling interfaces within ELM

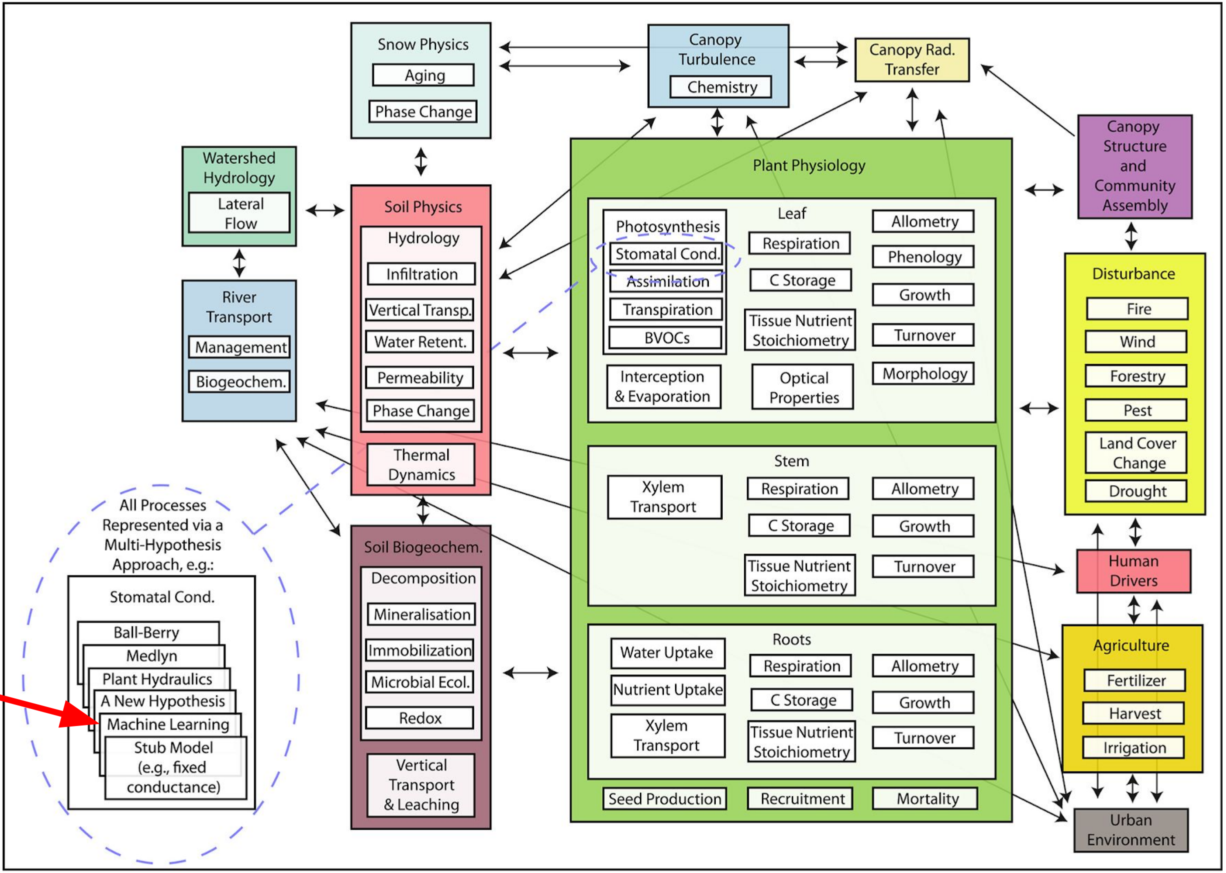
Nature



McDowell et al. (2019)

Hybrid ML/Process-based Modeling for Terrestrial Modeling

Individual processes can be represented by a multi-hypothesis approach, and ML provides an opportunity for a data-derived hypothesis that can be further explored or used to calibrate other hypotheses, when sufficient data are available.



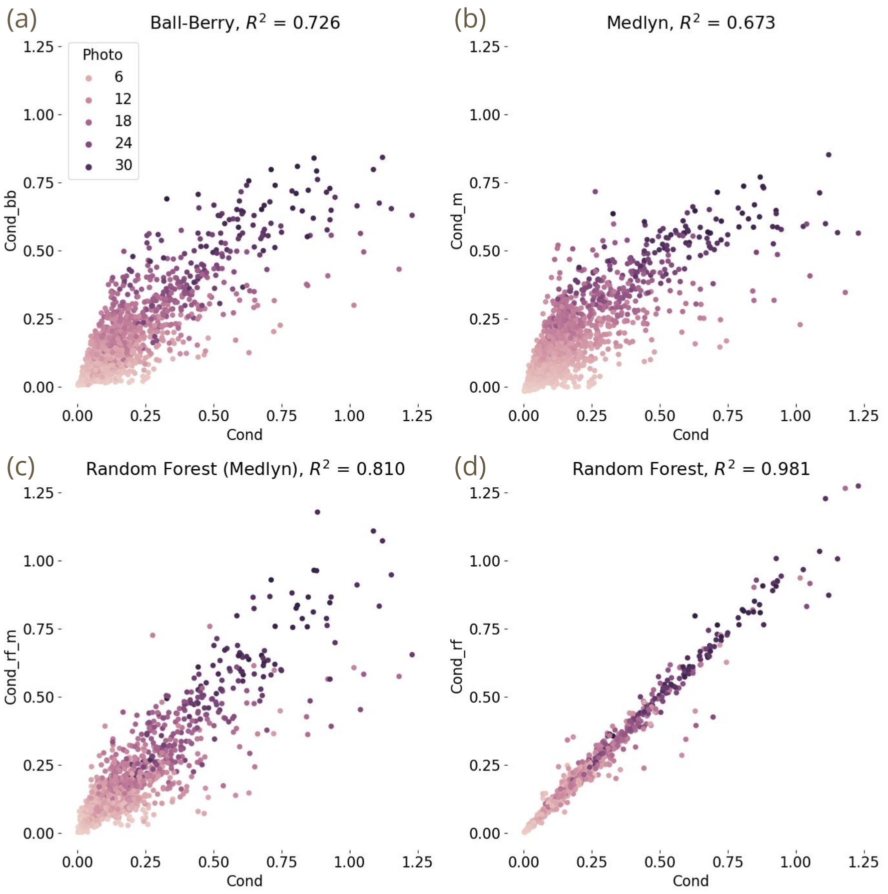
(a) Process Schematic of a Possible Full-Complexity Configuration of a Land Surface Model

(Fisher and Koven, 2020)

Hybrid Modeling of Photosynthesis and Ecohydrology

- Significant leaf-level data may be used to train ML parameterizations to **improve accuracy** and **computational performance**
- **Estimated stomatal conductance** vs. measured stomatal conductance for (a) Ball-Berry, (b) Medlyn, (c) Random forest (with Medlyn inputs), and (d) Random forest with all inputs from Lin et al. (2015)
- Inputs to the Medlyn parameterization are leaf-level CO_2, photosynthesis, and vapor pressure deficit
- Random forest trained on these three inputs (c) performs slightly better than Medlyn
- Random forest trained on more variables (d) achieves an R^2 of 0.98

(Massoud, Collier, et al. in prep)



https://ai4esp.org/ https://ai4esp.slack.com/

AI4ESP | Artificial Intelligence for Earth System Predictability

A multi-lab initiative working with the Earth and Environmental Systems Science Division (EESSD) of the Office of Biological and Environmental Research (BER) to develop a new paradigm for Earth system predictability focused on enabling artificial intelligence across field, lab, modeling, and analysis activities.

White papers were solicited for development and application of AI methods in areas relevant to EESSD research with an emphasis on quantifying and improving Earth system predictability, particularly related to the integrative water cycle and extreme events.

> ***How can DOE directly leverage artificial intelligence (AI) to engineer a substantial (paradigm-changing) improvement in Earth System Predictability?***

156 white papers were received and read to plan the organization of the **AI4ESP Workshop on Oct 25–Dec 3, 2021**

Earth System Predictability Sessions

- Atmospheric Modeling
- Land Modeling
- Human Systems & Dynamics
- Hydrology
- Watershed Science
- Ecohydrology
- Aerosols & Clouds
- Climate Variability & Extremes
- Coastal Dynamics, Oceans & Ice

Cross-Cut Sessions

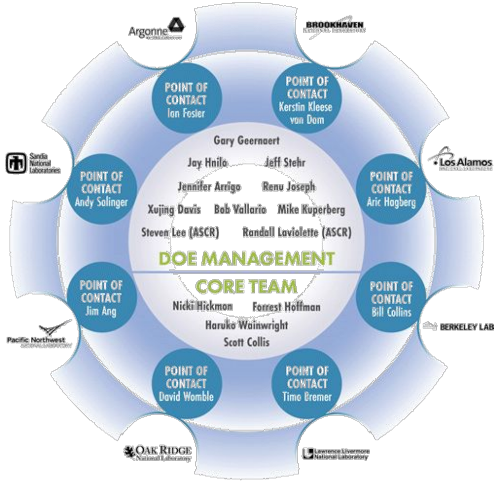
- Data Acquisition
- Neural Networks
- Surrogate models and emulators
- Knowledge-Informed Machine Learning
- Hybrid Modeling
- Explainable/Interpretable/Trustworthy AI
- Knowledge Discovery & Statistical Learning
- AI Architectures and Co-design

Workshop Report

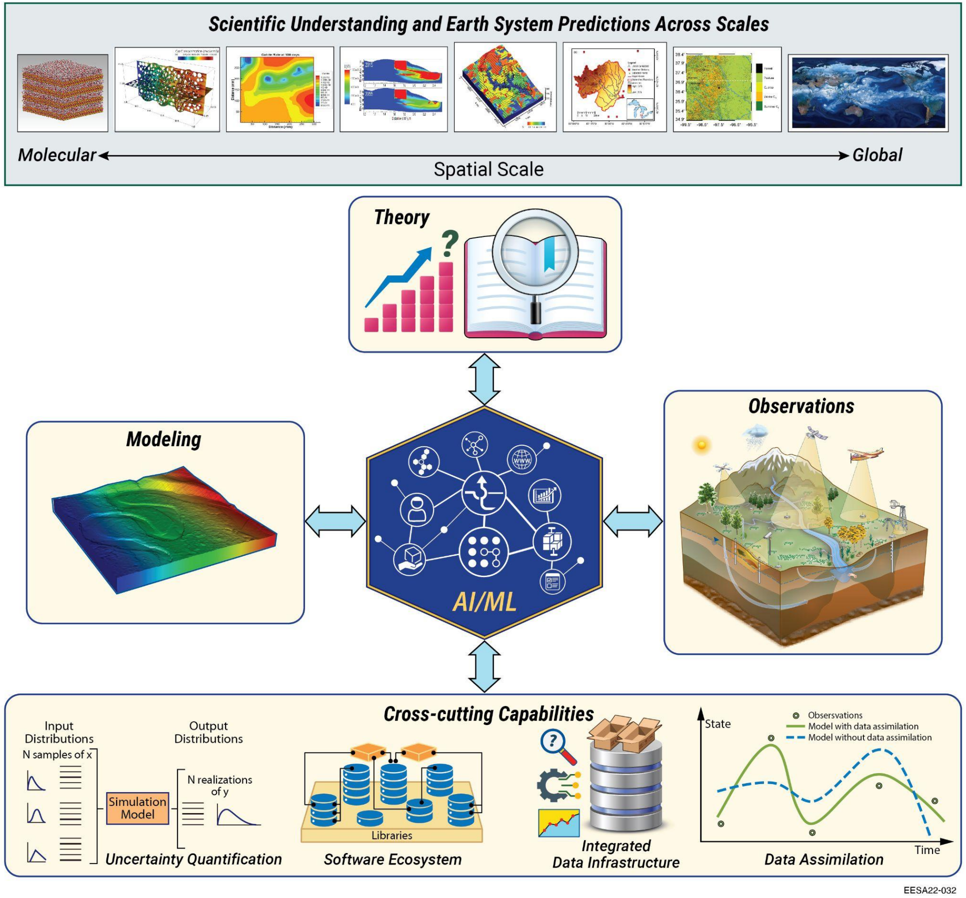
- Posted on ai4esp.org
- Executive Summary
- Long summary
- Earth science chapters
- Computational science chapters

AMS Special Collection

- Open submissions for new AI for the Earth Systems journal



AI4ESP Workshop Highlights



AI4ESP Workshop Highlights

Overview of priorities emerging from the AI4ESP workshop across 3 key themes.

These priorities will help address major challenges for Earth system predictability

Earth Science Priorities

- New observations
- AI-ready data products
- Data-driven and hybrid models
- Analytical approaches
- Uncertainty quantification, model parametrization & calibration

To Tackle Challenges

- Significant data gaps
- Scaling and heterogeneity
- Extreme events
- Representation of human activities
- Knowledge discovery
- Accurate high-resolution predictions with low bias, uncertainty
- Providing actionable, timely information for decision making

Computational Science Priorities

- Hybrid models
- Fundamental math and algorithms
- Interpretable, trustworthy AI
- AI-enabled data acquisition
- Data, software, hardware infrastructure

To Tackle Challenges

- Physically consistent predictions for data-driven models
- Computational costs of process models
- Sparse data, extreme values
- Identifying causality
- Interpretable, trustworthy predictions
- Data discovery, access, synthesis
- Model development and comparison

Programmatic and Cultural Priorities

- AI research centers
- Workforce development
- Codesign infrastructure
- Common standards, benchmarks
- Seed projects, integrate AI into programs
- AI ethics and policies

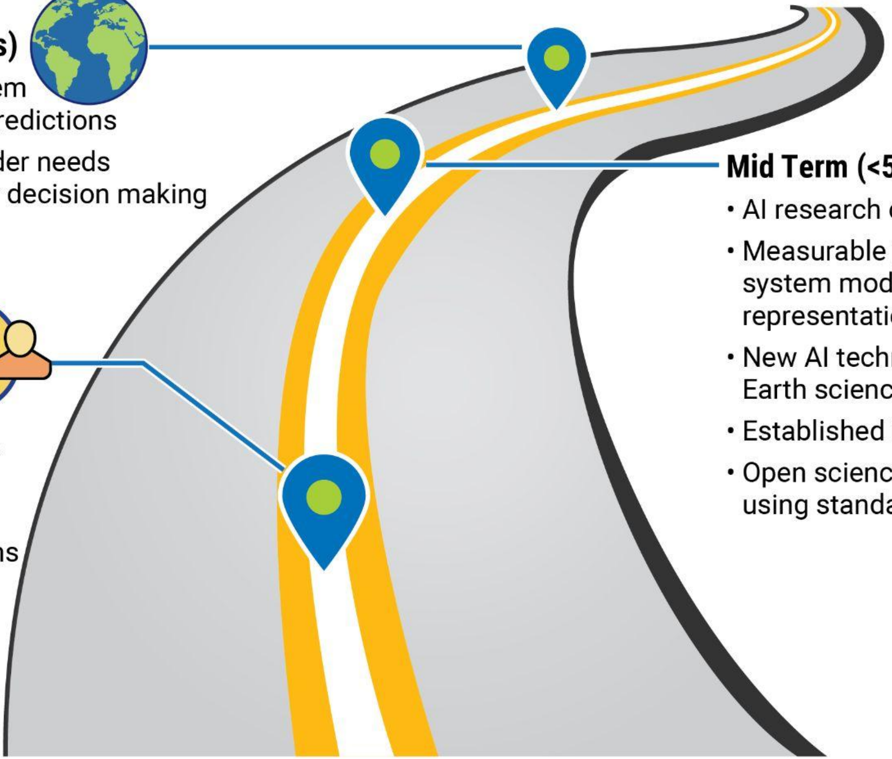
To Tackle Challenges

- Interdisciplinary scientific research
- Diverse organizational missions
- Personnel lack training in AI/ML
- Using data, communicating across research domains, organizations
- Data bias, model fairness, explainability of predictions

EESA22-031

AI4ESP Workshop Highlights

Idealized Roadmap for Success AI4ESP



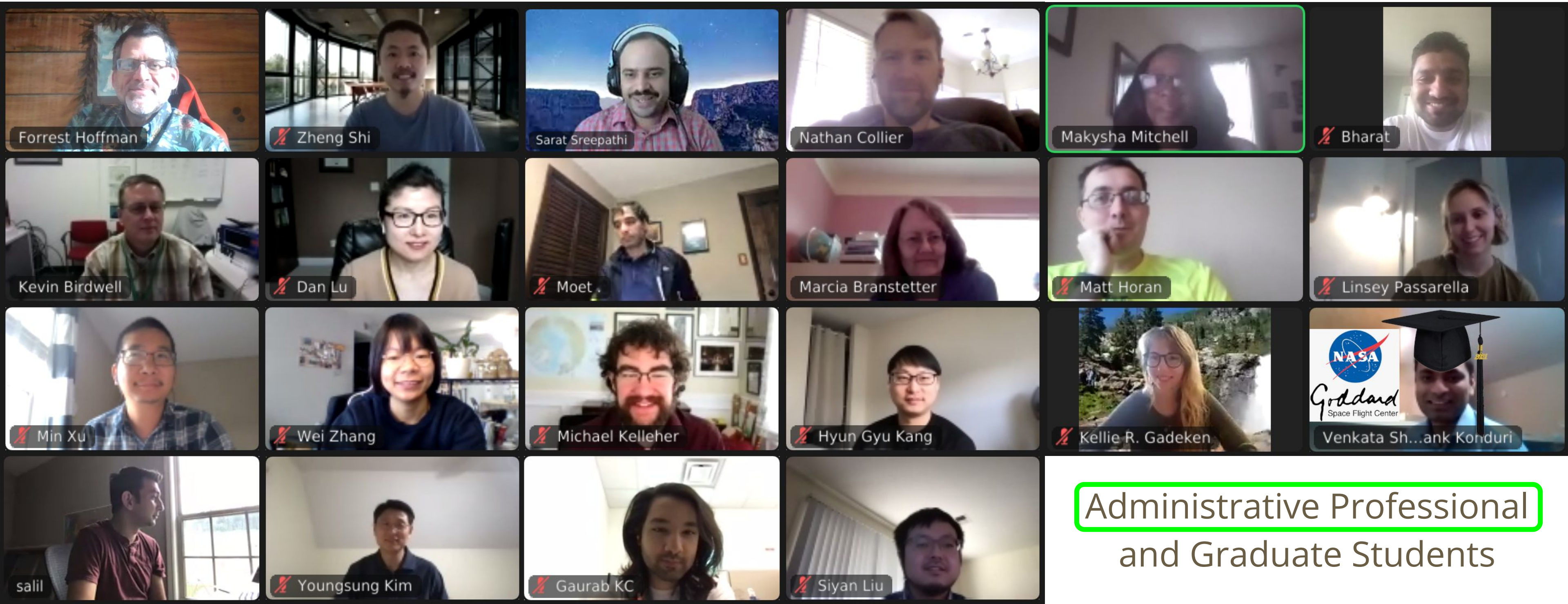
Long Term (<10 years)
- Improved Earth system understanding and predictions
- Supporting stakeholder needs at relevant scales for decision making

Mid Term (<5 years)
- AI research centers
- Measurable improvement in Earth system models with better representation of human activities
- New AI techniques tailored for Earth science applications
- Established interdisciplinary workforce
- Open science culture with data sharing using standards, co-developed models

Near Term (<2 years)
- Open benchmark datasets
- AI-enabled observations and data products based on gaps
- Seed efforts to demonstrate potential of AI in existing programs and modeling frameworks
- Cross-disciplinary collaborations to initiate activities

EESA22-033

Computational Earth Sciences Group Members



Administrative Professional

and Graduate Students

Staff and Postdoctoral Scholars

Located in the ORNL Climate Change Science Institute (CCSI) in Building 4500N, F Corridor

Frontier *at Oak Ridge National Laboratory is the #1 fastest supercomputer on the [TOP500] List and the first supercomputer to break the exaflop barrier (May 30, 2022).*

University of Tennessee, Knoxville

The Bredesen Center

The Bredesen Center for Interdisciplinary Research and Graduate Education unites resources and capabilities from the University of Tennessee and Oak Ridge National Laboratory to promote advanced research and to provide innovative solutions to global challenges in energy, engineering, and computation under the umbrella of the UT-Oak Ridge Innovation Institute (UT-ORII).

Seeking to create opportunities for exceptional students to engage in interdisciplinary research and education, the Bredesen Center offers a doctoral degree in the following areas:

- Energy Science and Engineering (ESE)
- Data Science and Engineering (DSE)

Questions?