U.S. DEPARTMENT OF ENERGY

Office of Science

Exploiting Artificial Intelligence for Advancing Earth and Environmental System Science

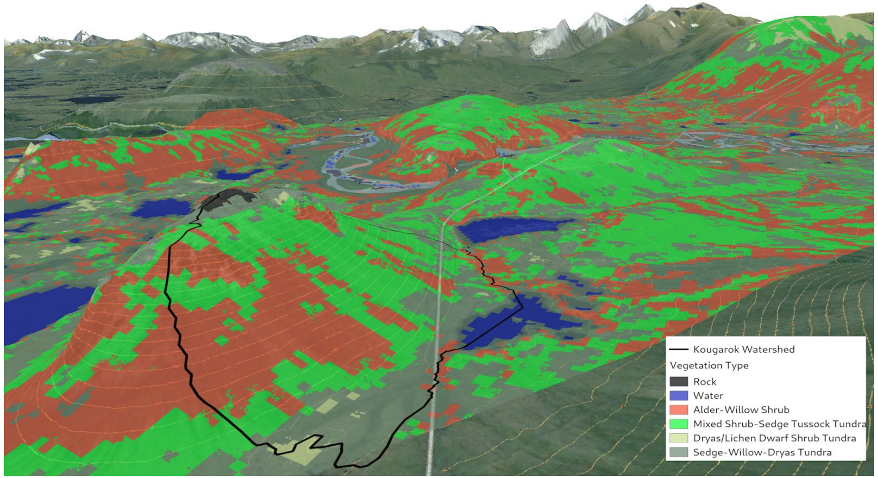
Forrest M. Hoffman[1], Jitendra Kumar[1], Zachary L. Langford[1], V. Shashank Konduri[2], Nathan Collier[1], Min Xu[1], and William W. Hargrove[3]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]USDA Forest Service, Asheville, NC, USA

SECOND SYMPOSIUM ON
SCIENCE-GUIDED AI
(SGAI-AAAI-21)
Held as part of AAAI Fall Symposium Series (FSS) 2021
from November 4-6, 2021

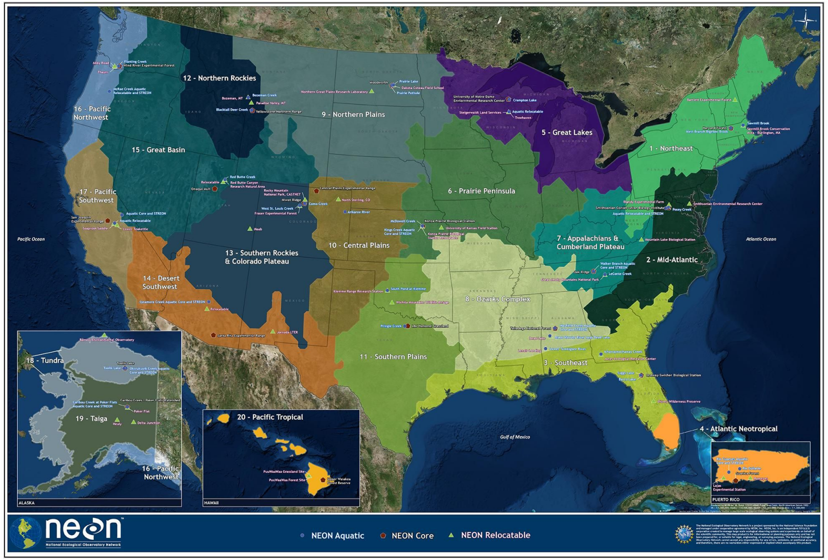
Introduction

- 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 evaluate and benchmark model results
- Explosive data growth and the promise of discovery through data-driven modeling necessitate new methods for feature extraction, change detection, data assimilation, simulation, and analysis



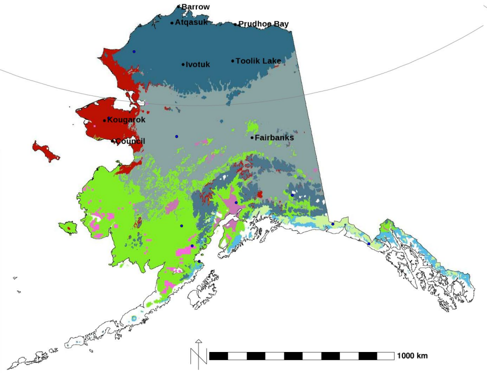
Langford et al. (2019)

Sampling Network Design

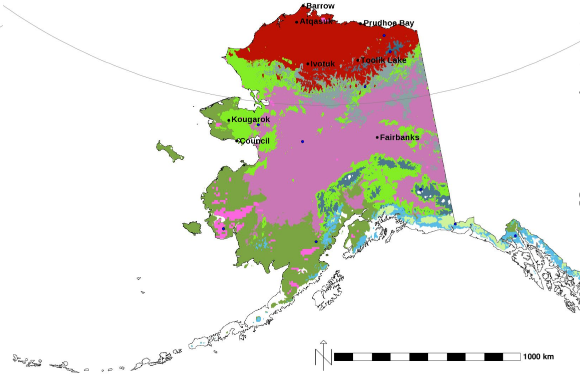


NSF's NEON Sampling Domains

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



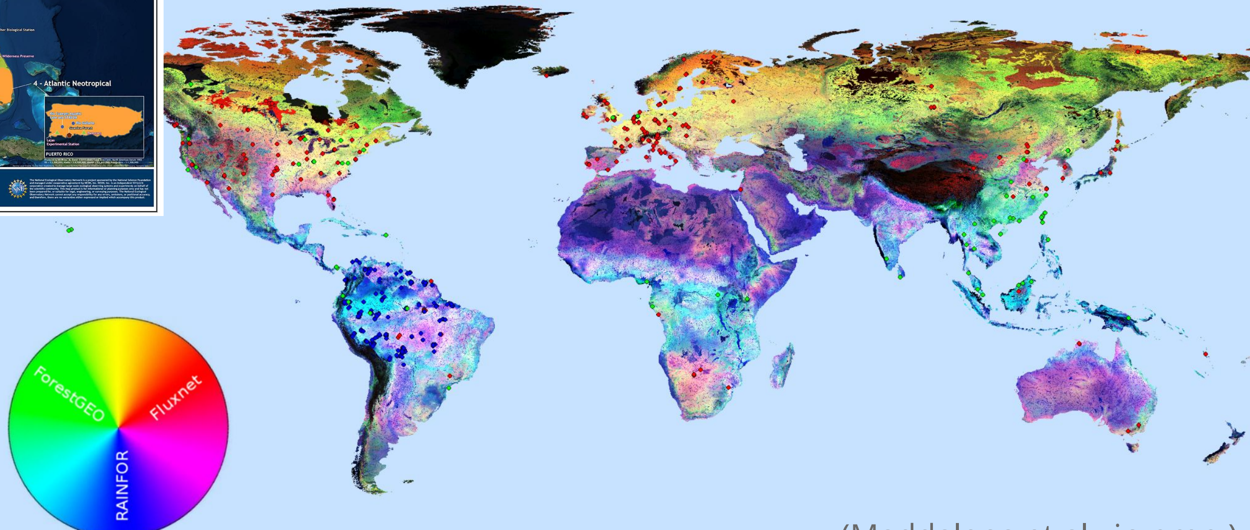
2000–2009



2090–2000

(Hoffman et al., 2013)

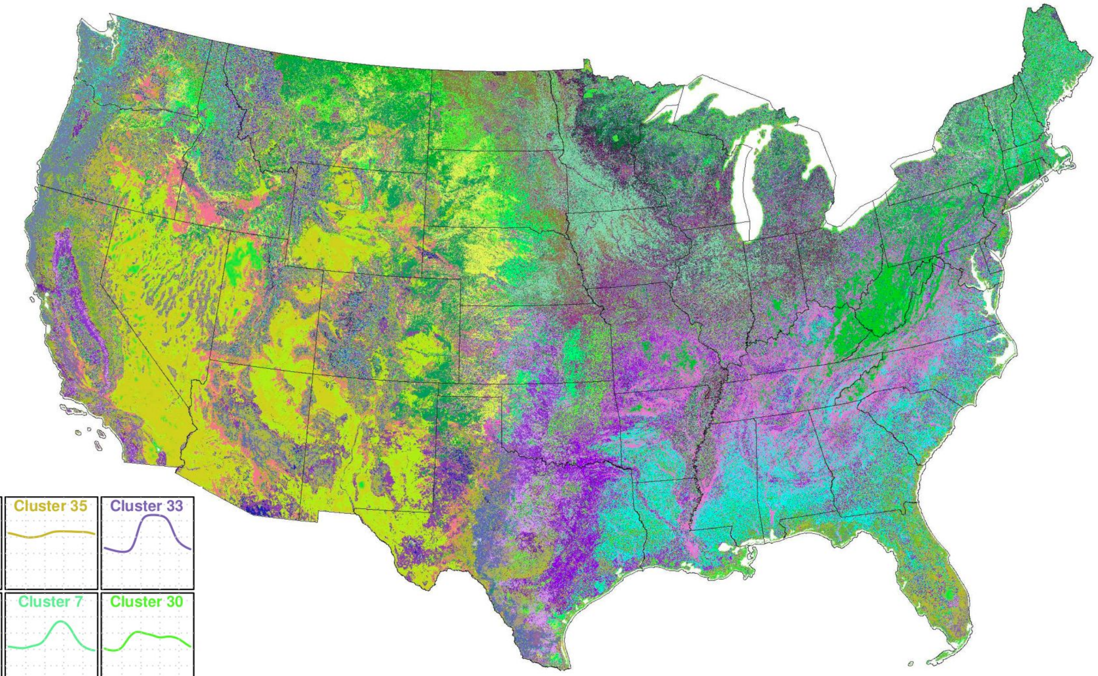
Triple-Network Global Representativeness

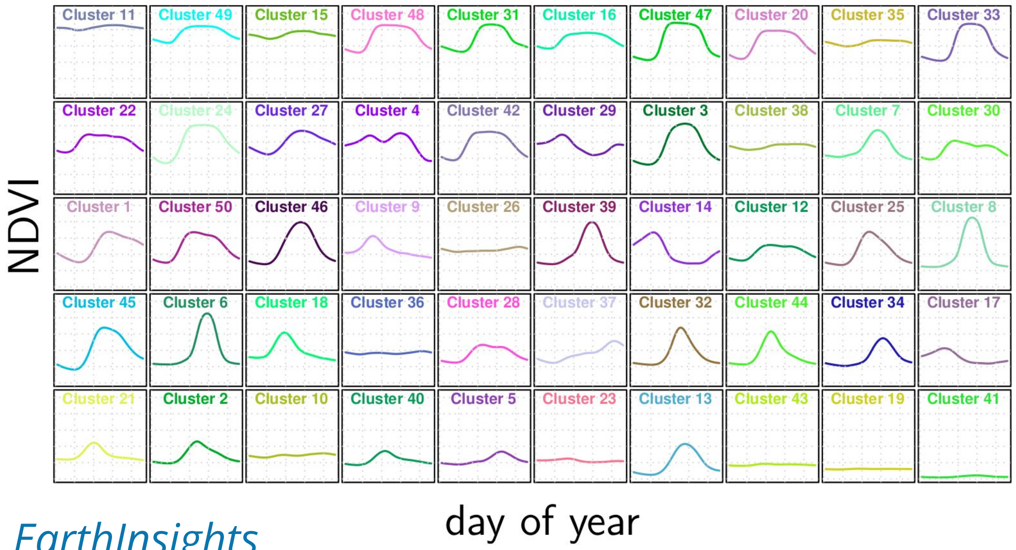


(Maddalena et al., in prep.)

50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI
Clustered from 2000 to present



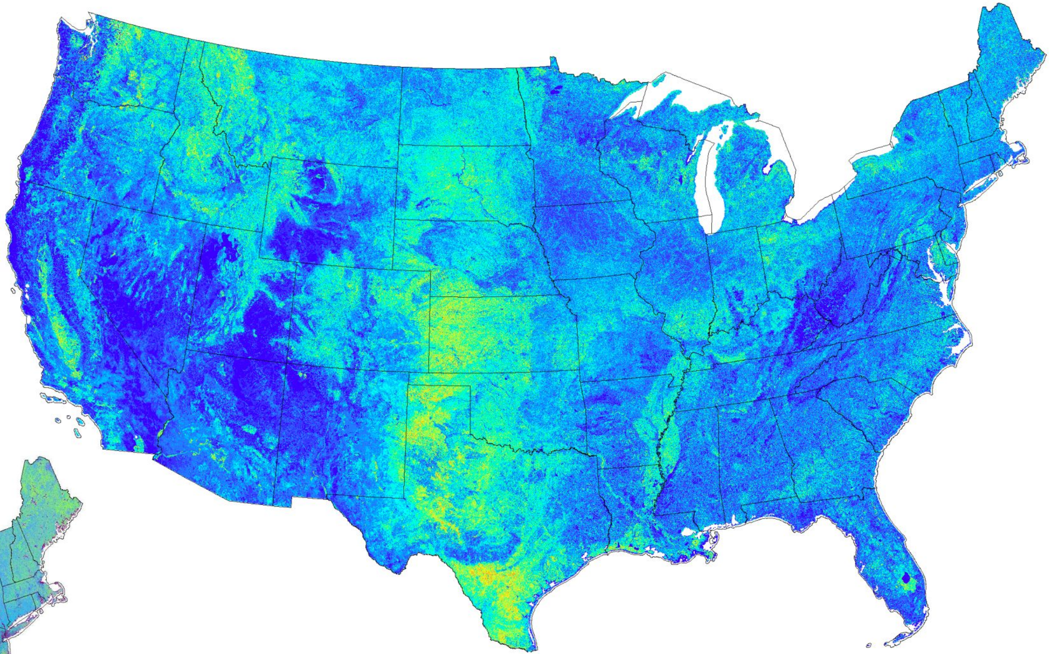


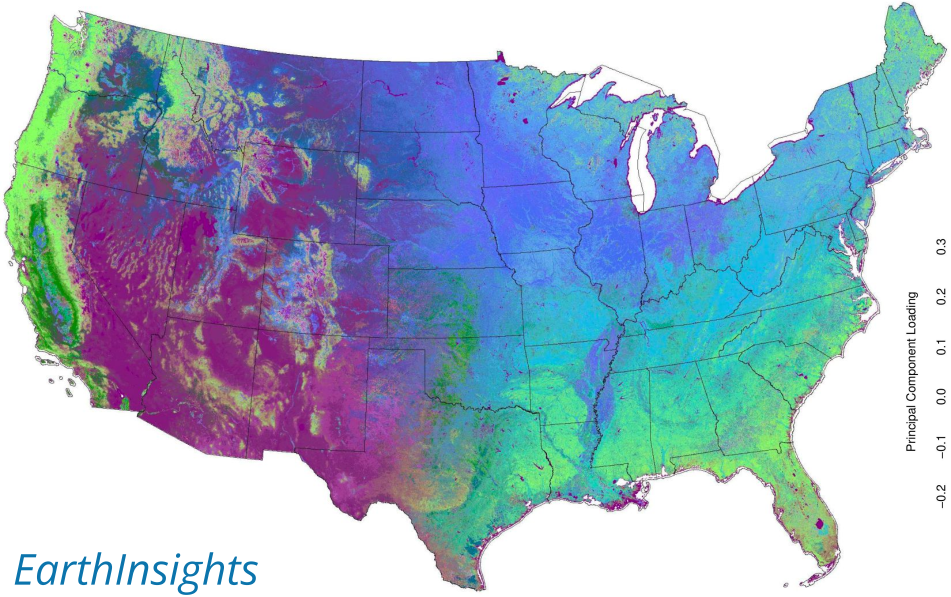
50 Phenoregion Prototypes (Random Colors)

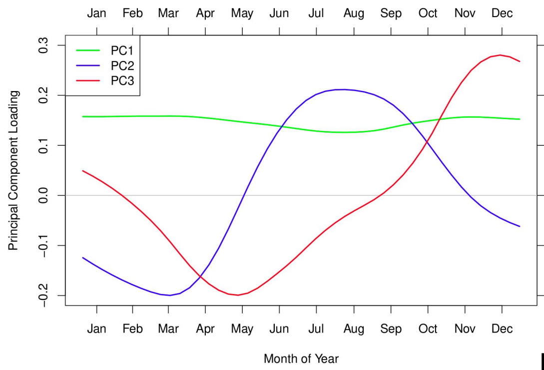
EarthInsights

Hargrove et al., in prep.

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







Principal Components Analysis

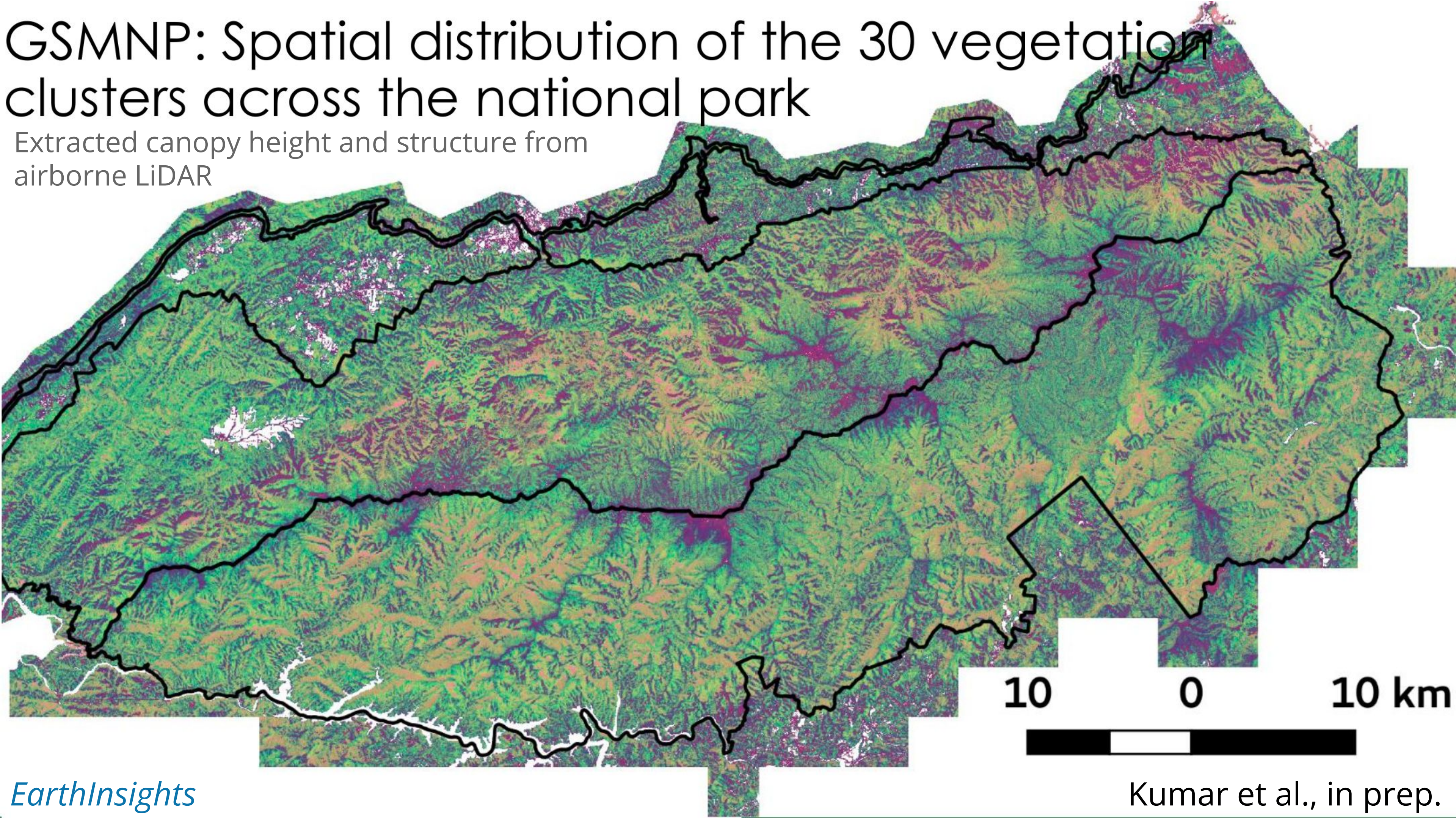
PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

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

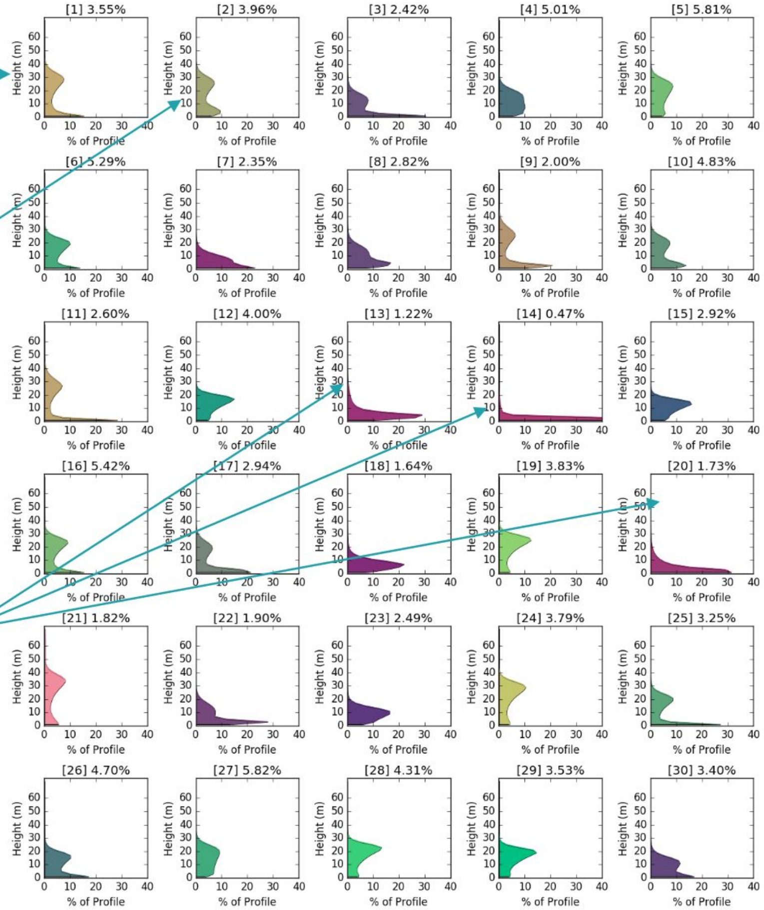
Kumar et al., in prep.

GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

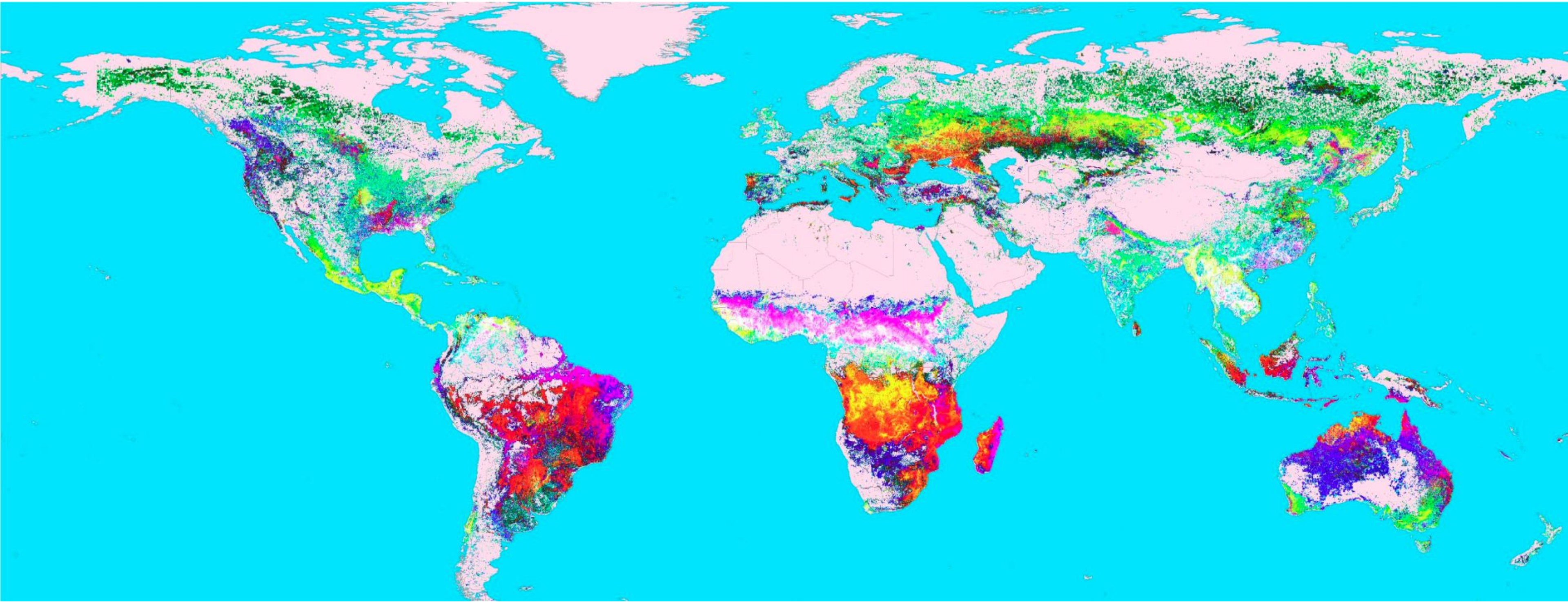
forests with slightly lower mean height with dense understory vegetation

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



Kumar et al., in prep.

Global Fire Regimes

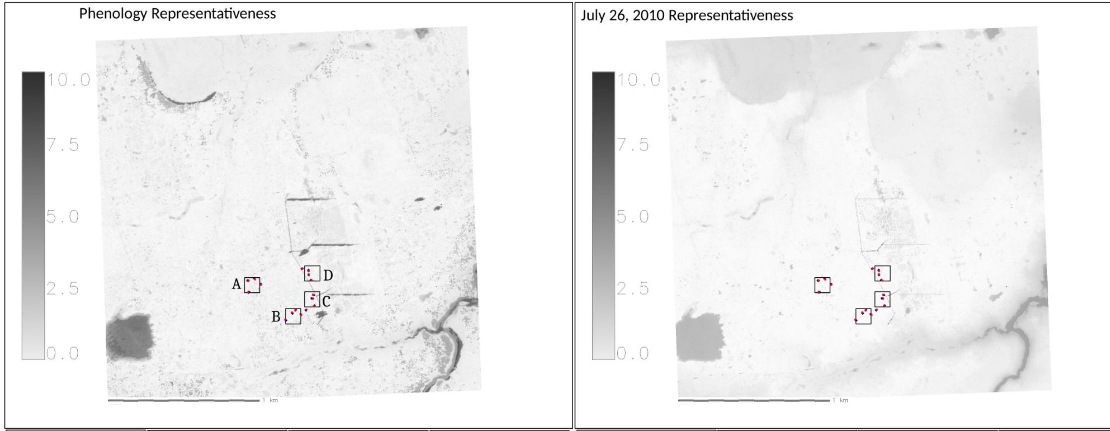


Regions that exhibit similar fire seasonality globally

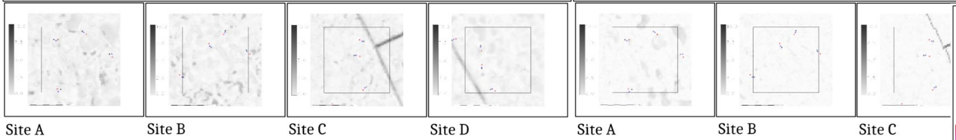
From MODIS "Hotspots" from 2002–2018

Norman et al., in prep.

Vegetation Distribution at Barrow Environmental Observatory

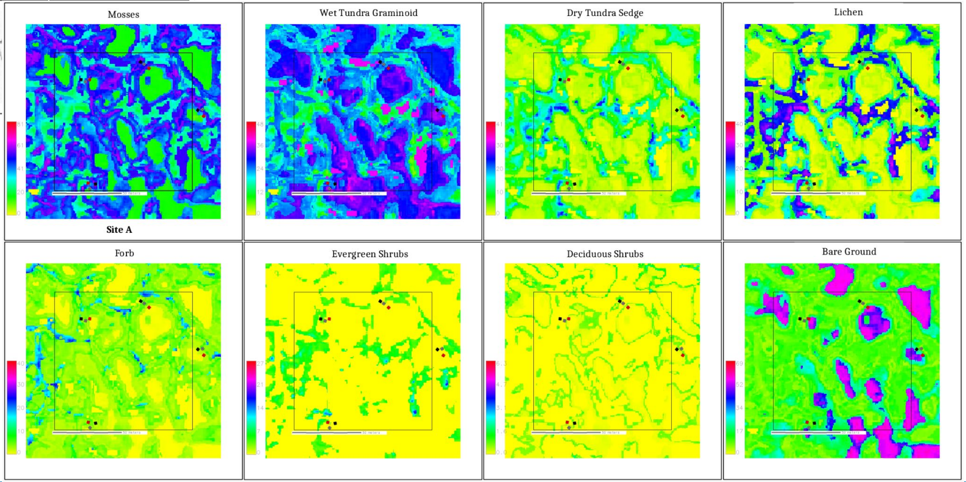


(Langford et al., 2016)



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

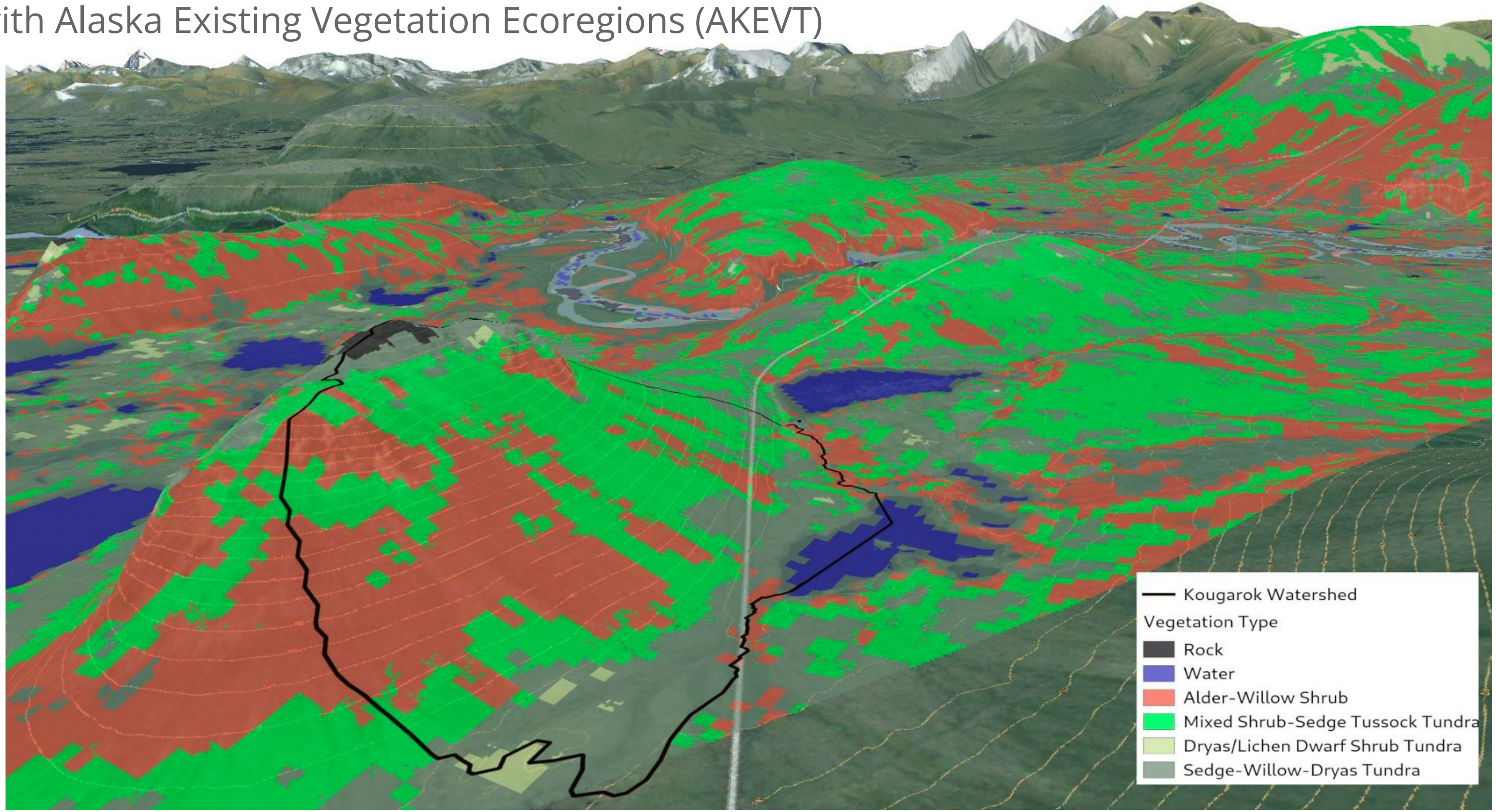


In situ data from field measurement activities inform the development of wide-scale maps of vegetation distribution through inference using remote sensing data as surrogate variables, and relationships with environmental controls can be extracted

EarthInsights

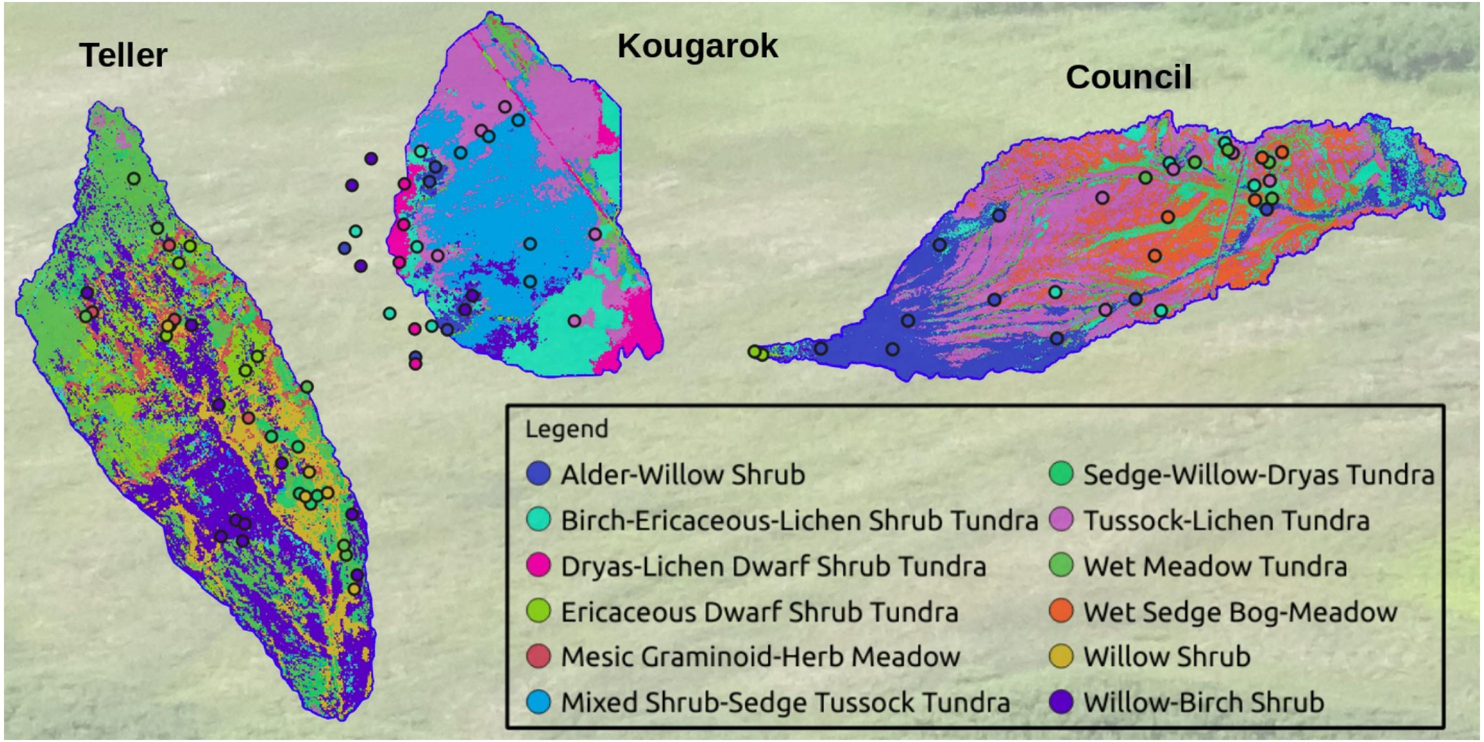
Arctic Vegetation Mapping from Multi-Sensor Fusion

Using Hyperion Multispectral and IfSAR-derived Digital Elevation Model
Trained with Alaska Existing Vegetation Ecoregions (AKEVT)



Langford et al., *Remote Sensing*, 2019

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



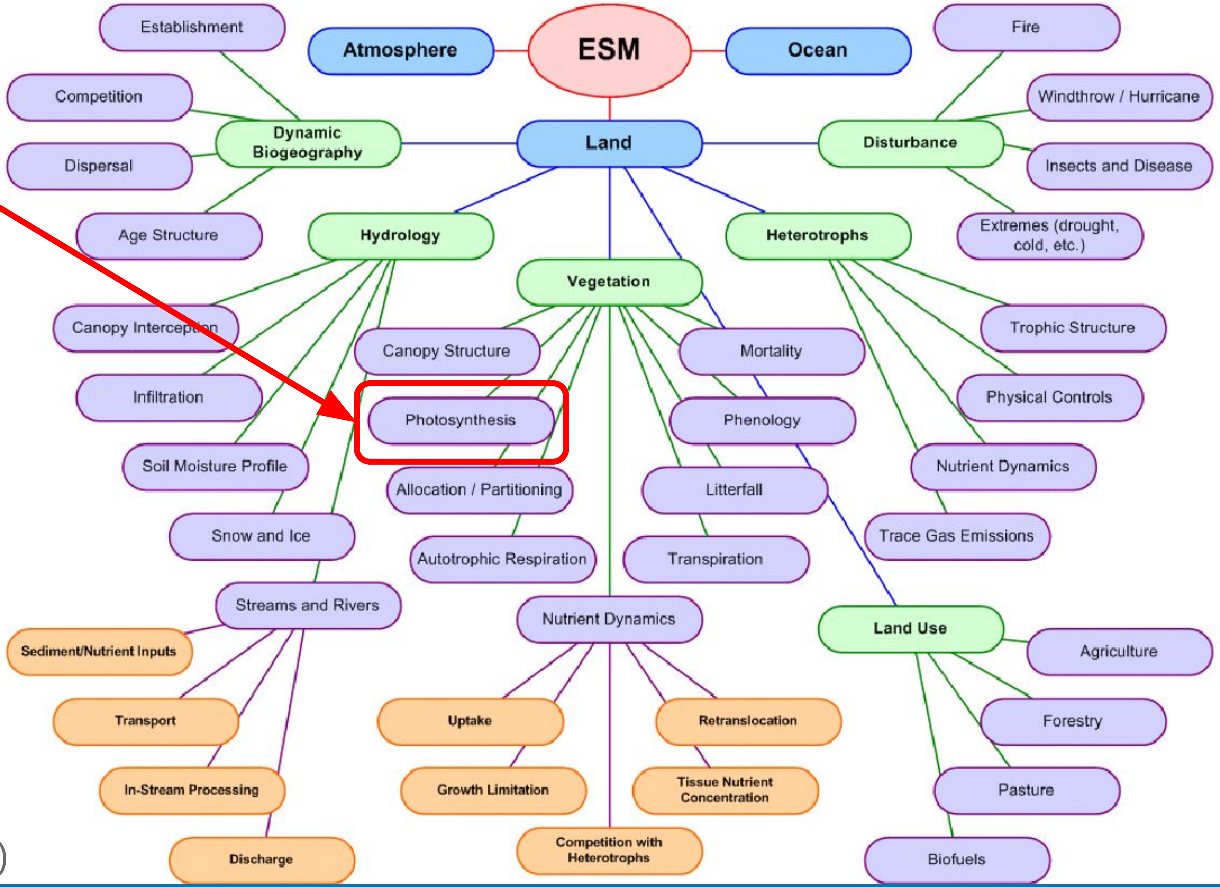
At the watershed scale, vegetation community distribution follows topographic and water controls.
At a fine scale, nutrients limit the distribution of vegetation types.

(Konduri et al., in prep.)

Hybrid ML/Process-based Modeling for Terrestrial Modeling

In the hierarchy of land model processes, we start with the **photosynthesis** parameterization because

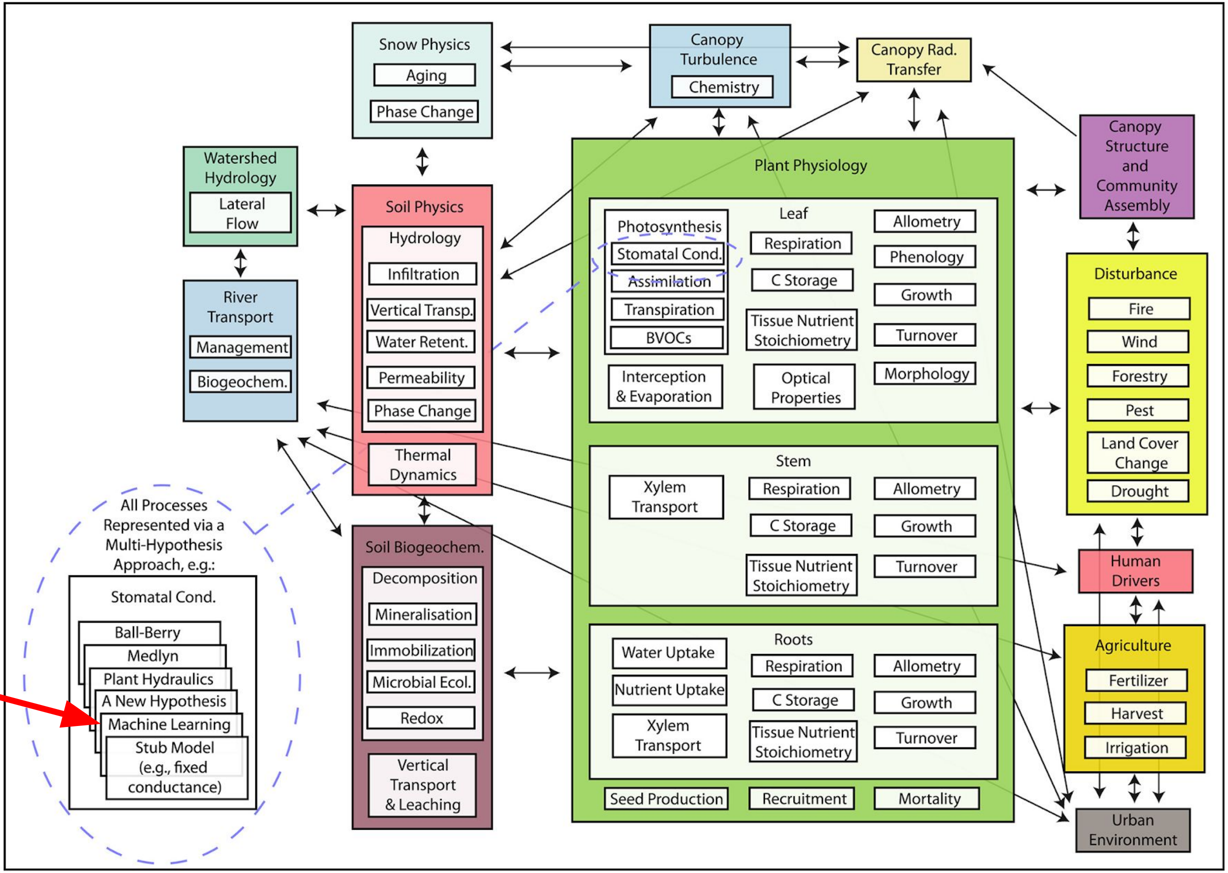
- Multiple hypotheses
- Many leaf-level measurements
- Most computationally intensive part of the land model



(Figure from P. E. Thornton)

Hybrid ML/Process-based Modeling for Terrestrial Modeling

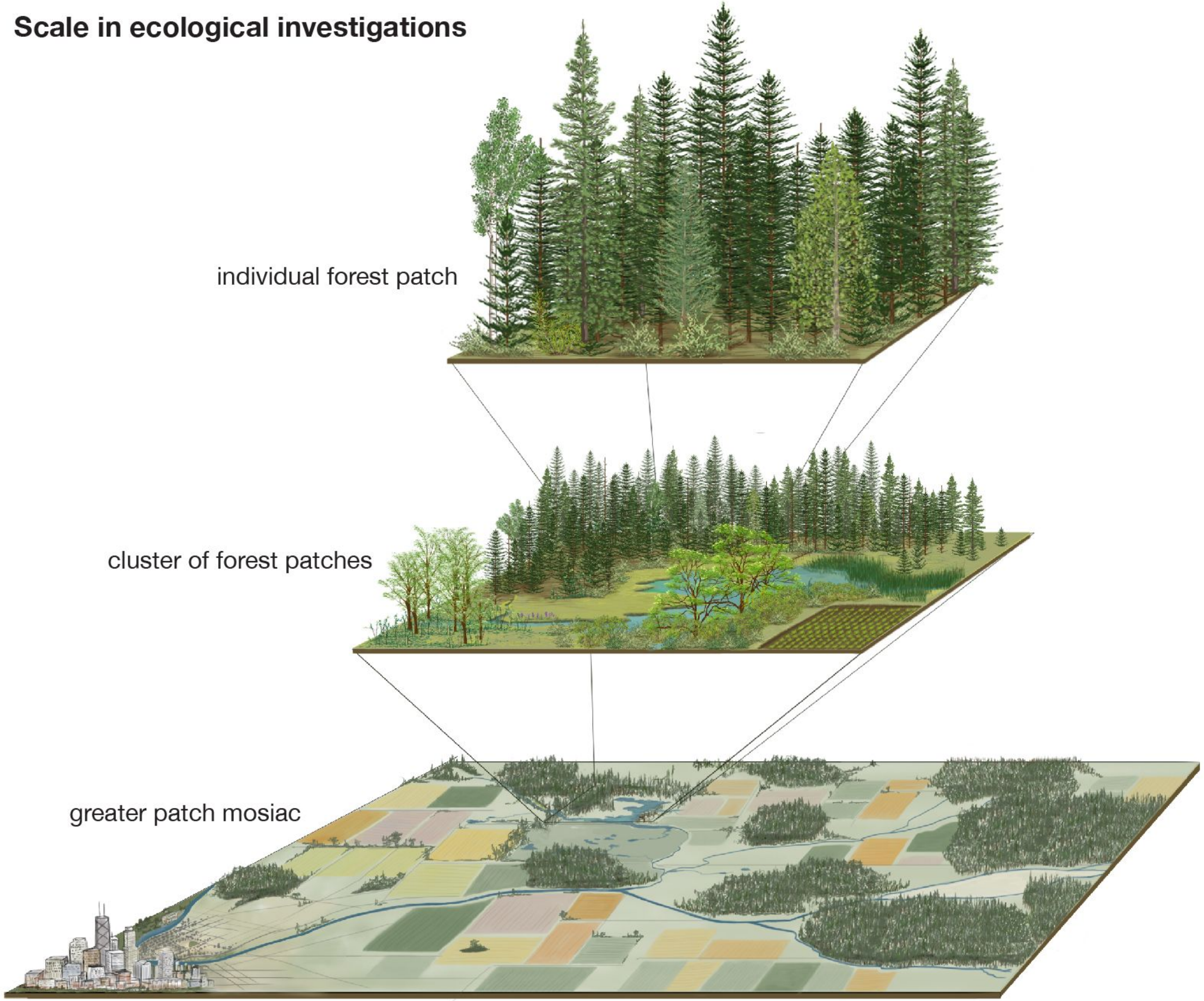
Individual processes can be represented by a multi-hypothesis approach, and ML provides an opportunity for a data-derived hypothesis that can be further explored or used to calibrate other hypotheses, when sufficient data are available.

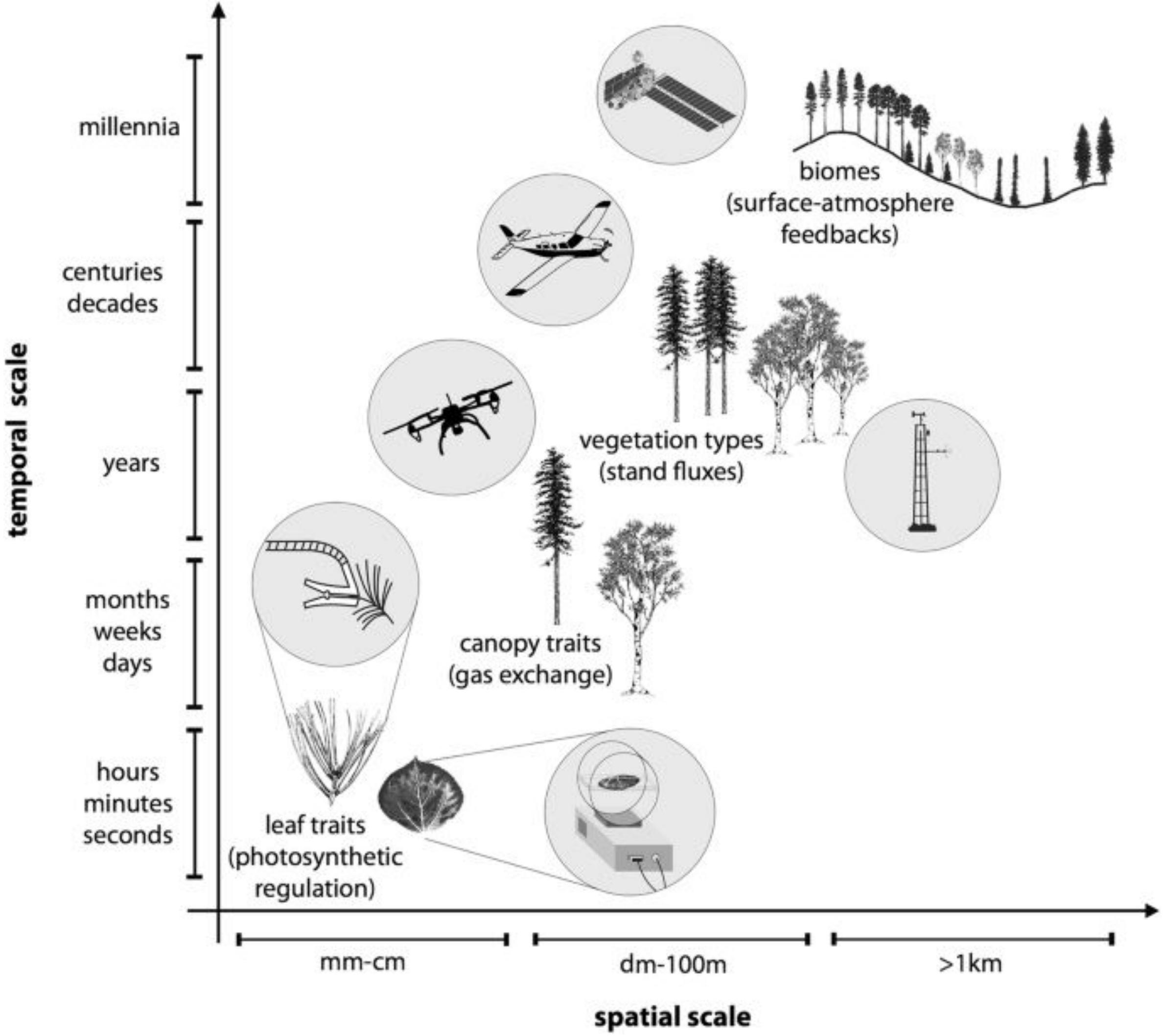


(a) Process Schematic of a Possible Full-Complexity Configuration of a Land Surface Model

(Fisher and Koven, 2020)

Scale in ecological investigations




temporal scale
millennia
centuries
decades
years
months
weeks
days
hours
minutes
seconds
leaf traits
(photosynthetic
regulation)
canopy traits
(gas exchange)
vegetation types
(stand fluxes)
biomes
(surface-atmosphere
feedbacks)
mm-cm
dm-100m
>1km
spatial scale

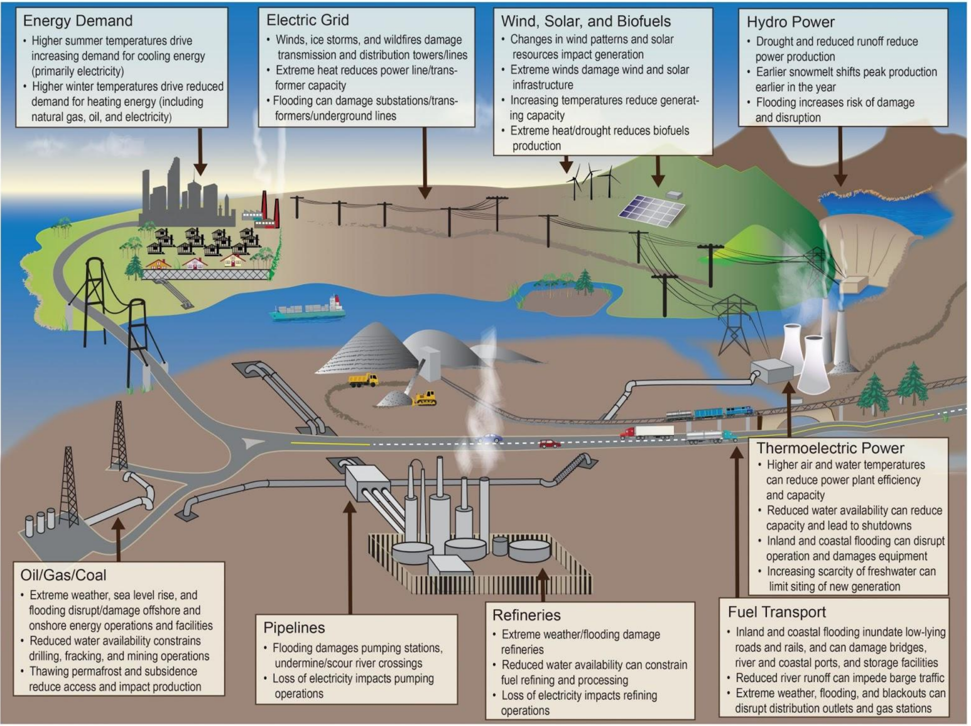
AI FOR SCIENCE TOWN HALL

Earth and Environmental Sciences

Forrest M. Hoffman (ORNL),
Rao Kotamarthi (ANL),
Haruko Wainwright (LBNL),
and the EES Writing Team

U.S. DEPARTMENT OF ENERGY | Office of Science

Grand Challenge #1



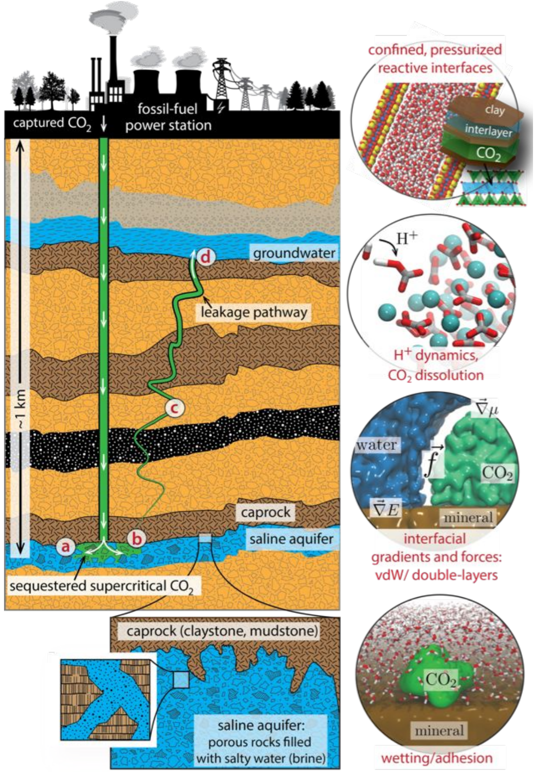
Project environmental risk and develop resiliency in a changing environment

- Increasing frequency of weather extremes and changing environment pose risks to energy infrastructure and the built environment
- Sparse observations and inadequate model fidelity limit the ability to identify vulnerability, mitigate risks, and respond to disasters

Grand Challenge #1

- New tools are needed to accelerate projection of weather extremes and impacts on energy infrastructure
- Building resiliency to address evolving risks will benefit from integration of smart sensing systems, built-for-purpose models, ensemble forecasts to quantify uncertainty, and dynamic decision support systems for critical infrastructure

Grand Challenge #2



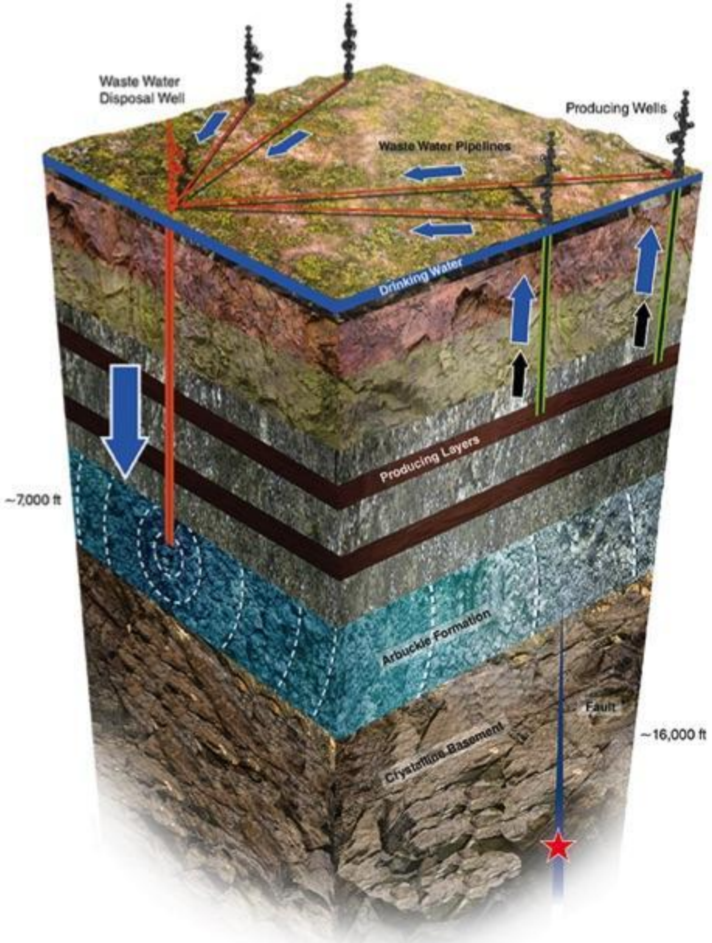
Characterize and modify subsurface conditions for responsible energy production, CO_2 storage, and contaminant remediation

- National energy security and transition to renewable energy resources relies on utilization of subsurface reservoirs for energy production, carbon storage, and spent nuclear fuel storage
- Subsurface data are uncertain, disparate, diverse, sparse, and affected by scaling issues
- Subsurface process models are incomplete, uncertain, and frequently unreliable for prediction

U.S. DEPARTMENT OF ENERGY | Office of Science

Grand Challenge #2

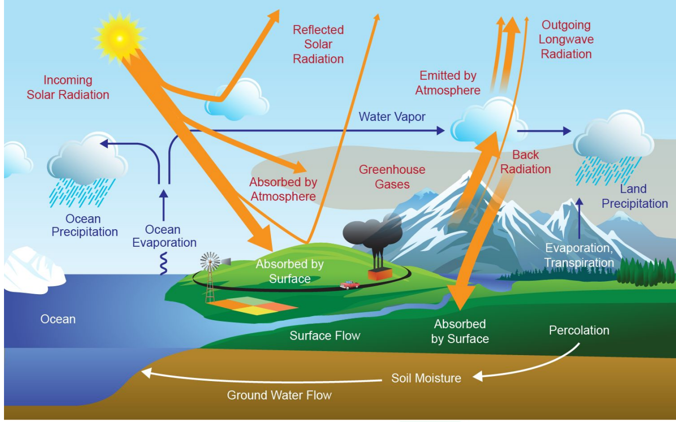
- We need to substantially increase hydrocarbon extraction efficiency, discover and exploit hidden geothermal resources, reduce induced seismicity and other impacts, improve geologic CO_2 storage, and predict long-term fate and transport of contaminants
- Mitigating risks requires improved subsurface characterization and assimilation of real-time data streams into predictive models of geological and ecological processes



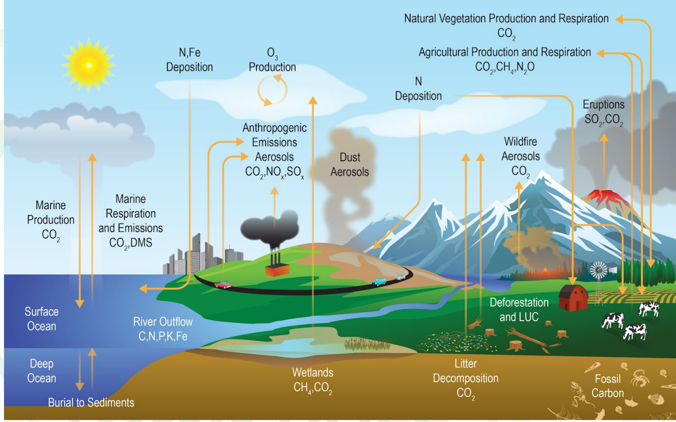
Grand Challenge #3

Develop a predictive understanding of the Earth system under a changing environment

- To advance the nation’s energy and infrastructure security, a foundational scientific understanding of complex and dynamic hydrological, biological, and geochemical processes and their interactions is required (across atmosphere, ocean, land, ice)
- Knowledge must be incorporated into Earth system models to project future climate conditions for various scenarios of population, socioeconomics, and energy production and use

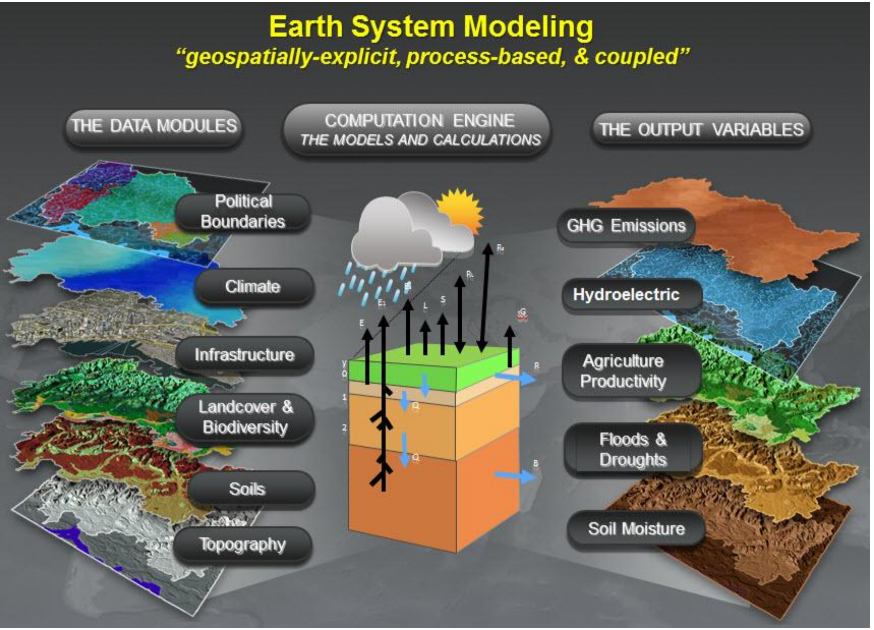


Energy & Water Cycles



Carbon & Biogeochemical Cycles

Grand Challenge #3

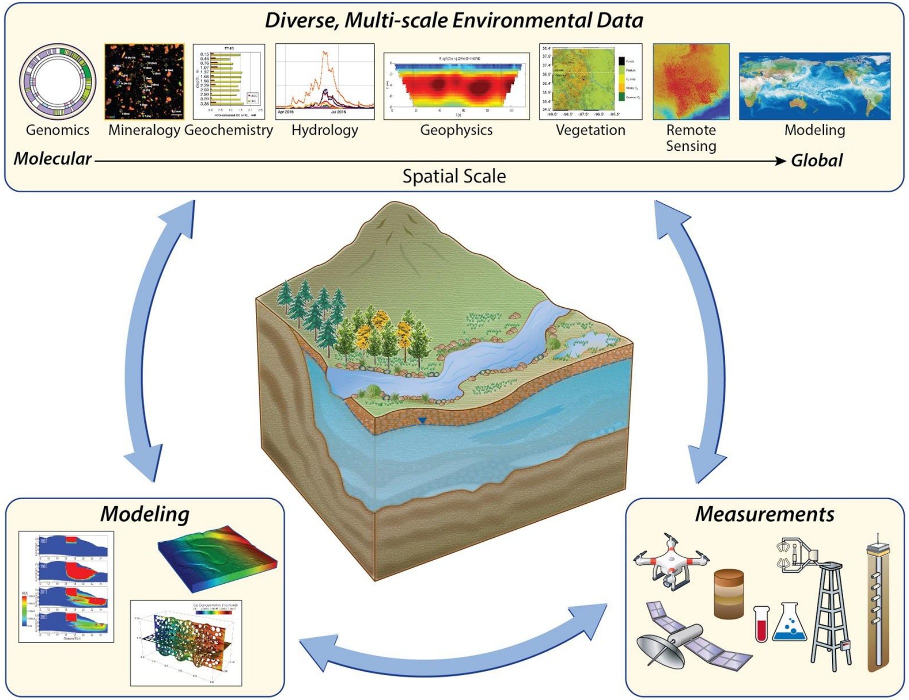


- Accurate predictions are needed to quantify changes in atmospheric and ocean circulation and weather extremes, to close the carbon cycle, and to understand responses and feedbacks of human, terrestrial, and marine ecosystems to environmental change
- Advances in genomics and bioscience data need to be leveraged to provide detailed understanding of plant–microbial interactions and their adaptations and feedbacks to the changing environment

Grand Challenge #4

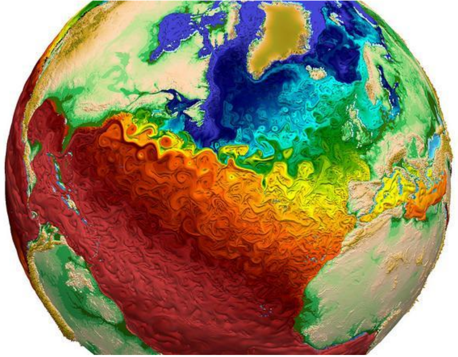
Ensure global water security under a changing environment

- Water resources are critical for energy production, human health, food security, and economic prosperity
- Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change



Grand Challenge #4

- Methods are needed to integrate disparate and diverse multi-scale data with models of watersheds, rivers, and water utility infrastructure
- Predictions of water quality and quantity require data-driven models and smart sensing systems
- Water resource management must account for changes in weather extremes, population, and economic growth

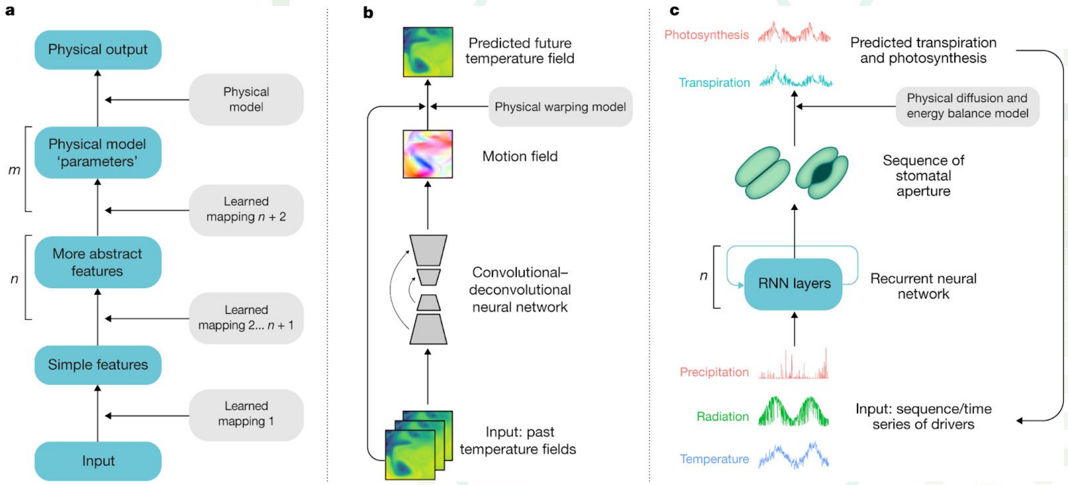


Accelerating Development

The near-term (5–10 years) priorities are to:

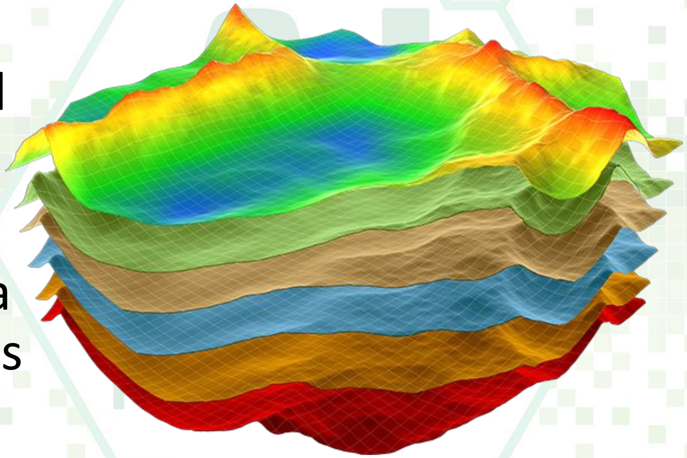
- Develop hybrid process-based/AI modeling frameworks for Exascale systems
- Develop strategies for mapping hybrid components on GPU/CPU based on computational density and communications patterns
- Develop physics / chemistry / biology-constrained ML
- Develop explainable AI and ML methods for hypothesis generation and testing

Hybrid Approaches to Earth Science Simulation (Reichstein et al., 2019)



Expected Outcomes

- Model testbeds and surrogate models are expected to yield insights into process understanding across all Grand Challenges
- Data-driven and physics-constrained hybrid models are expected to stimulate new discovery and bridge space and time scales
- Integrated models of Earth system processes and energy/built infrastructure will enhance national energy and water security through simulation
- AI methods will enable effective use of large data streams for energy production, predictive process understanding, and environmental resiliency



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 a workshop in Fall 2021.

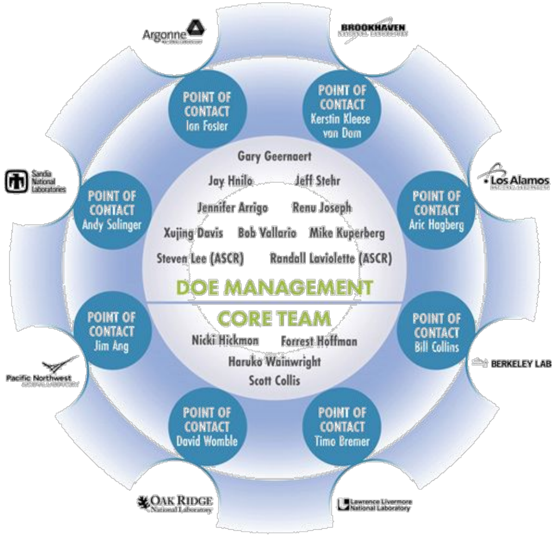
AI4ESP Workshop: Oct 25–Dec 3, 2021

Earth System Predictability Sessions

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

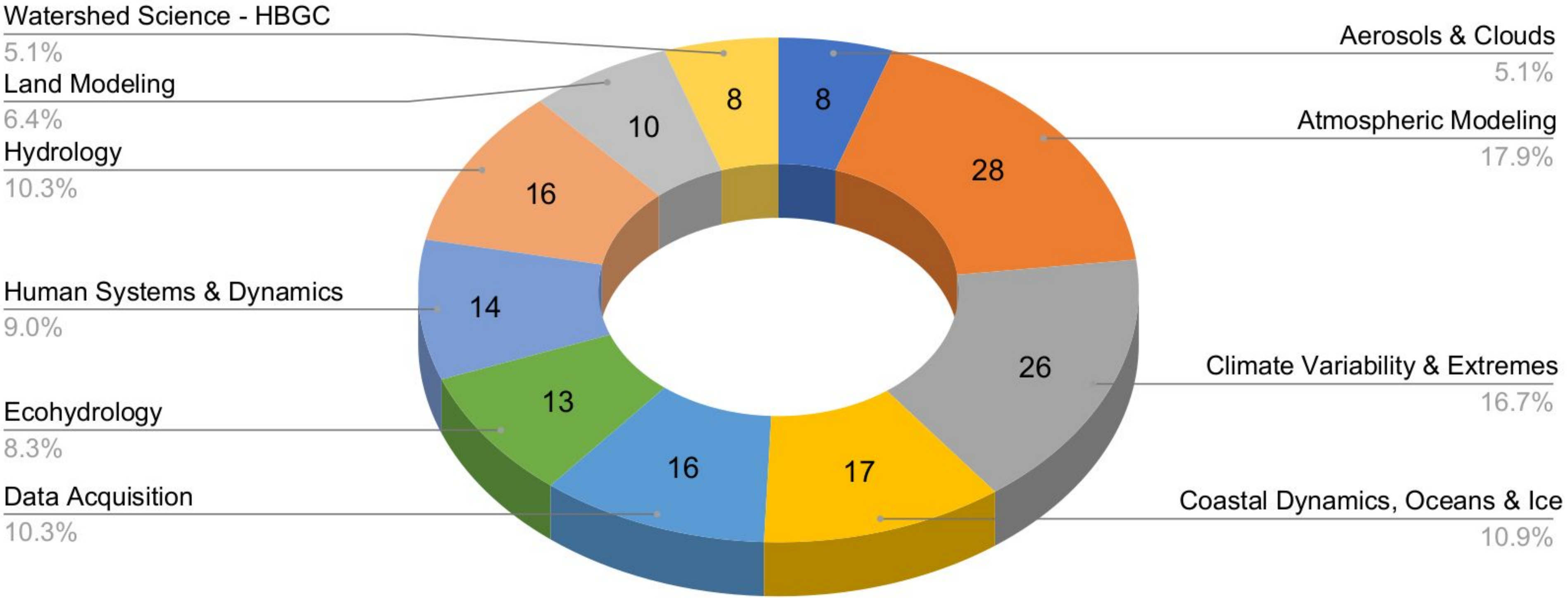
Cross-Cut Sessions

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



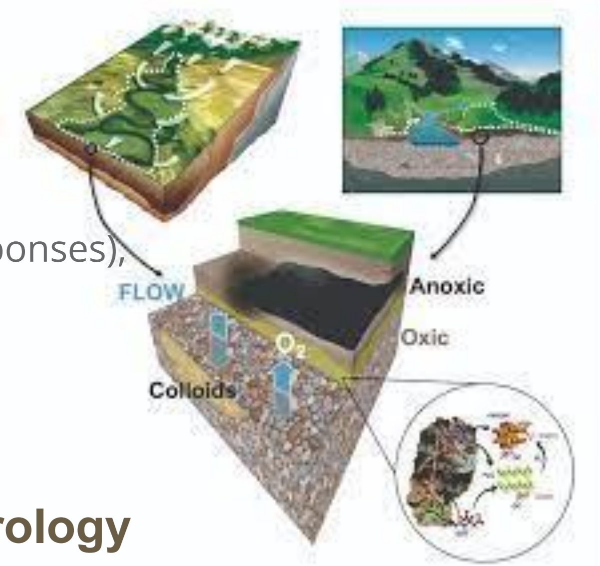
AI4ESP White Papers: Earth System Predictability Topics

Earth System Predictability Topics from 156 White Papers

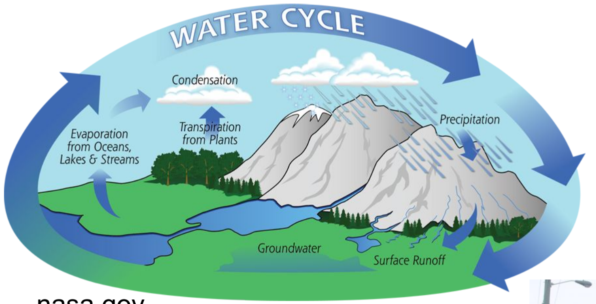


AI4ESP White Papers: Earth System Predictability Topics

- **Watershed science**
  - Hydro-Biogeochemistry, Soil biogeochemistry
  - Water quality
  - Lab-to-field, field-to-regional scale analysis
  - Experimental data, sensor networks (rapid responses), and experimental/network designs



ess.science.energy.gov



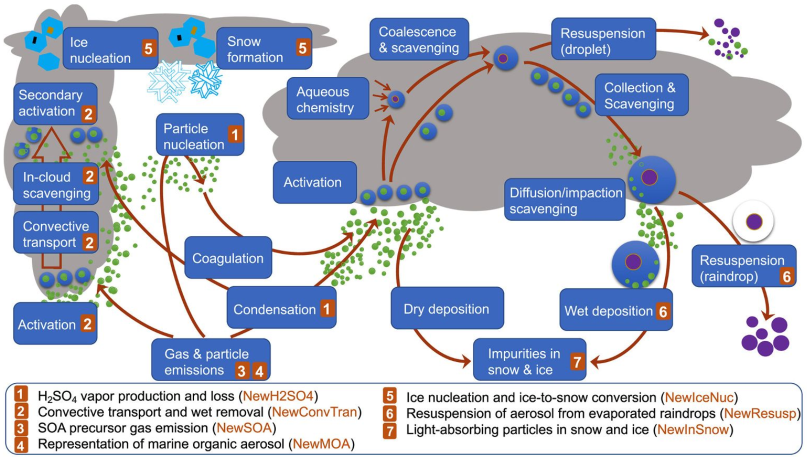
nasa.gov

- **Hydrology**
  - Water resources
  - Precipitation-induced hazards (floods etc)
  - Weather/hydrological monitoring
  - Groundwater to surface water models
  - Mountain hydrology
  - Regional to continental scale

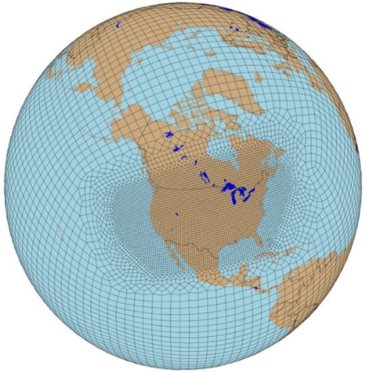
climate.gov

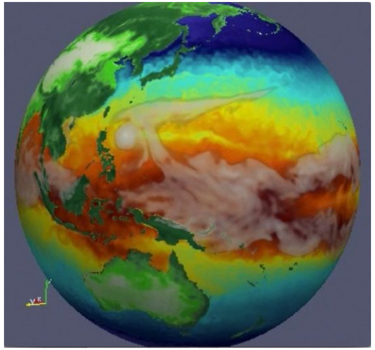
AI4ESP White Papers: Earth System Predictability Topics

- **Atmospheric Modeling**
  - Convection and turbulence
  - Surface Fluxes
  - Radiation
  - Model Tuning
  - General concepts that can generalized to other ESMs components



e3sm.org





e3sm.org

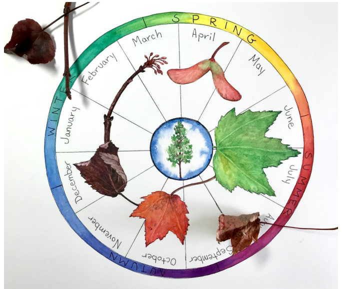
- **Aerosols and Clouds**
  - Cloud Classification
  - Aerosol cloud interactions

AI4ESP White Papers: Earth System Predictability Topics

- **Land Modeling**
  - Agriculture / Crops
  - Leaf Phenology
  - Streamflow / Water Availability
  - Wildfire
  - Satellite Data Assimilation

- **Ecohydrology**
  - Stomatal Conductance / Photosynthesis
  - Plant Hydraulics and Growth
  - Evapotranspiration
  - Soil Moisture
  - Soil Hydrology

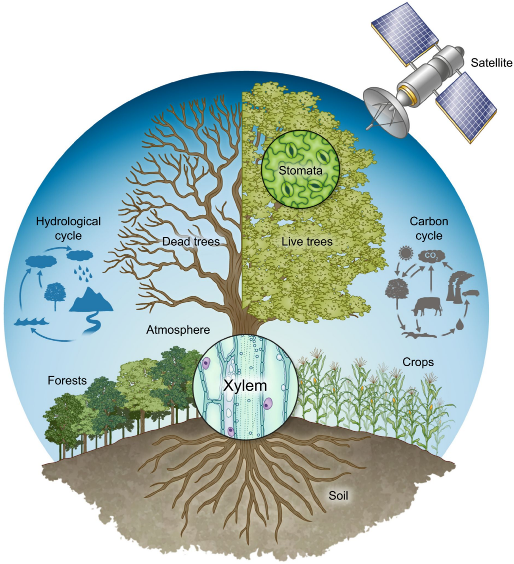

Getty Images



Adkins Arboretum


wallpaperbetter.com


ABC7 News



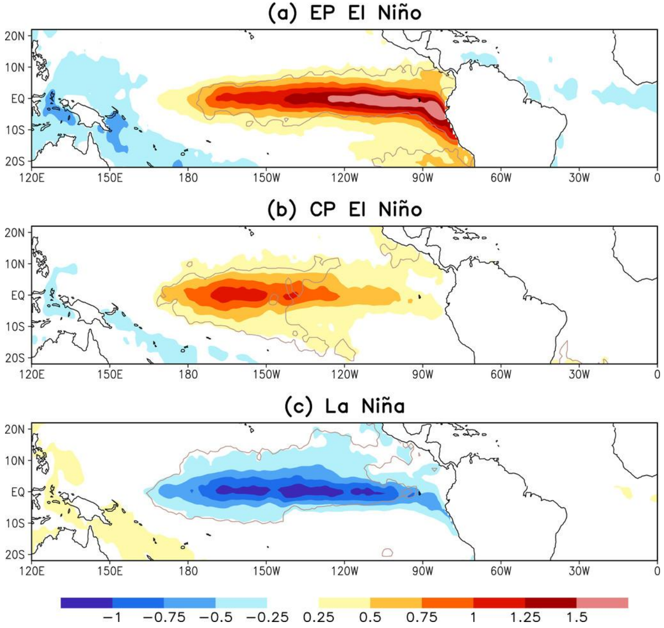
McDowell et al. (2019)


drought.gov

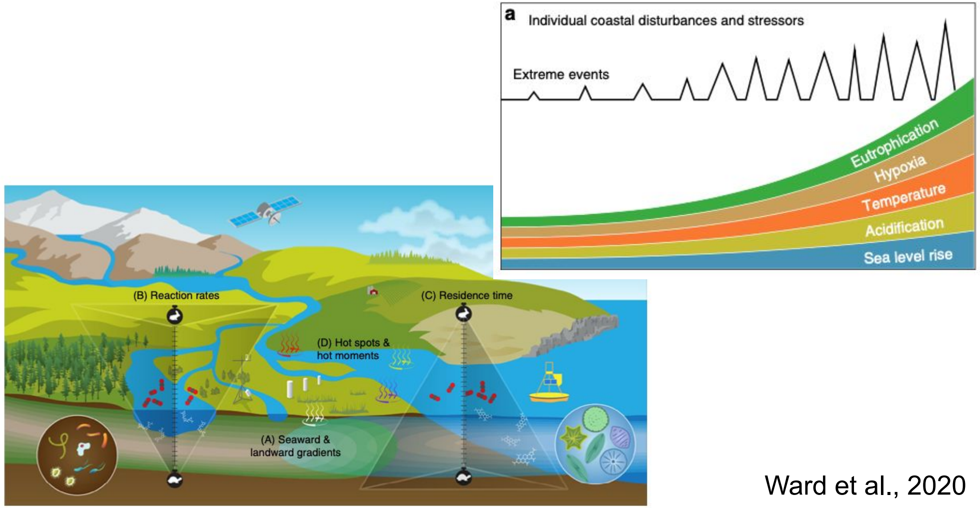

Nature

AI4ESP White Papers: Earth System Predictability Topics

- **Climate variability and Extremes**
  - TCs, ARs, Compound/Cascading events
  - Predictability
  - Circulation/climate variability (ENSO, NAO etc)
  - Telecommunication



Wang et al., 2014

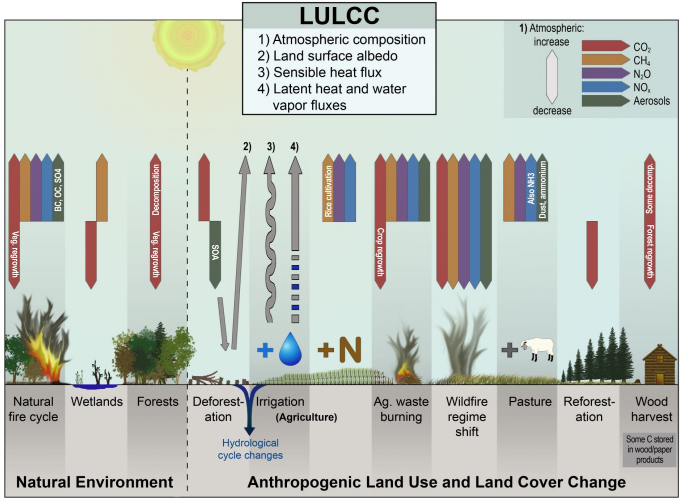


Ward et al., 2020

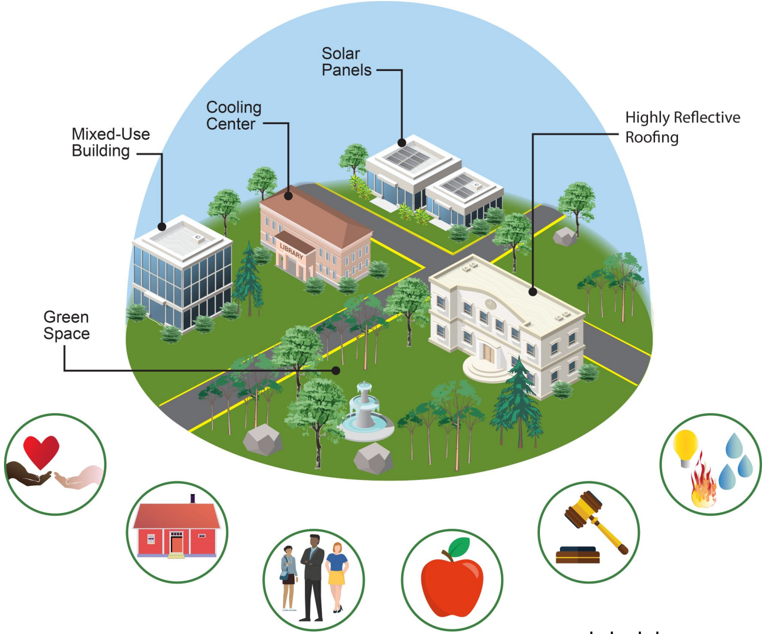
- **Coastal dynamics, Ocean/Ice**
  - Ocean/land/ice interface
  - Sea-level rise, storm surge
  - Coastal ecosystem/carbon cycling

AI4ESP White Papers: Earth System Predictability Topics

- **Human Systems and Dynamics**
  - Human activities/population
  - Energy-water-land nexus
  - Agriculture
  - Urban environment
  - Land use/cover changes



globalchange.gov



globalchange.gov

Workshop Agenda

WEEK 1		
North American Eastern Time	**Day 1** **Monday, October 25, 2021**	**Day 2** **Tuesday, October 26, 2021**
12:00	• Welcome - Nicki Hickmon • Deputy Secretary of Energy - David M. Turk • Introduction to AI4ESP Initiative - Nicki Hickmon • Earth & Environmental Systems Sciences Division (EESSD) - Gary Geernaert • Advanced Scientific Computing Research (ASCR) - Barb Helland	Plenary Talk - Amy McGovern
12:15		Plenary Talk - Pierre Gentine
12:30		Break
12:45		Earth Science Topic Session **Land Modeling (Invited Only)** Session Chair: Beth Drewniak
13:15	**AI4ESP Workshop Structure, Charge & State-of-the-Science** AI4ESP Core Group: Nicki Hickmon, Haruko Wainwright, Forrest Hoffman, Scott Collis	
14:00	Break	
14:15	Panel Discussion Panel Chair: Rick Stevens Panel: Grace E. Kim, Prabhat Ram, Kirk Borne	
14:45		Break
15:00	Earth System Predictability Session **Atmospheric Modeling (Invited Only)** Session Chair: Ruby Leung	Cross-cut Session **Data Acquisition to Distribution (Invited Only)** Session Chair: Giri Prakash
17:00	Adjourn	Adjourn

WEEK 2		
North American Eastern Time	**Day 3** **Monday, November 1, 2021**	**Day 4** **Tuesday, November 2, 2021**
12:00	Reports from Previous Sessions (15 min) • Atmospheric Modeling • Land Modeling • Data Acquisition	Plenary Talk - Chaopeng Shen
12:15		Plenary Talk - Rob Ross
12:30	Break	Break
12:45	Earth Science Topic Session **Human Systems & Dynamics (Invited Only)** Session Chair: Christa Brelsford	Earth Science Topic Session **Watershed Science (Invited Only)** Session Chair: Mavrik Zavarin
14:45	Break	Break
15:00	Earth Science Topic Session **Hydrology (Invited Only)** Session Chair: Charuleka Varadharajan	Cross-cut Session **Neural Networks (Invited Only)** Session Chair: Nathan Hodas
17:00	Adjourn	Adjourn

WEEK 3		
North American Eastern Time	**Day 5** **Monday, November 8, 2021**	**Day 6** **Tuesday, November 9, 2021**
12:00	Reports from Previous Sessions (15 min) • Human Systems & Dynamics • Hydrology • Watershed Science • Neural Networks	Plenary Talk - Tapio Schneider
12:15		Plenary Talk - Alison Appling
12:30	Break	Break
12:45	Earth Science Session **Ecohydrology (Invited Only)** Session Chair: Forrest Hoffman	Earth Science Session **Aerosols & Clouds (Invited Only)** Session Chair: Po-Lun Ma
14:45	Break	Break
15:00	Cross-cut Session **Surrogate Models & Emulators (Invited Only)** Session Chair: Nathan Urban	Cross-cut Session **Knowledge-Informed Machine Learning (Invited Only)** Session Chair: Frank Alexander
17:00	Adjourn	Adjourn

WEEK 4		
North American Eastern Time	**Day 7** **Monday, November 29, 2021**	**Day 8** **Tuesday, November 30, 2021**
12:00	Reports from Previous Sessions • Ecohydrology • Surrogate Models and Emulators • Aerosols & Clouds • Knowledge-Informed Machine Learning	Plenary Talk - Laure Zanna
12:15		Plenary Talk - Katie Dagon
12:30	Break	Break
12:45	Earth Science Session **Coastal Dynamics, Oceans & Ice (Invited Only)** Session Chair: Matt Hoffman	Earth Science Topic Session **Climate Variability & Extremes (Invited Only)** Session Chair: Maria Molina
14:45	Break	Break
15:00	Cross-cut Session **Knowledge Discovery & Statistical Learning (Invited Only)** Session Chair: Xingyuan Chen	Cross-cut Session **Explainable/Interpretable/Trustworthy AI (Invited Only)** Session Chair: Line Pouchard
17:00	Adjourn	Adjourn

WEEK 5		
North American Eastern Time	**Day 9** **Thursday, December 2, 2021**	**Day 10** **Friday, December 3, 2021**
12:00	Reports from Previous Sessions • Coastal Dynamics, Oceans, & Ice • Knowledge Discovery & Statistical Learning • Climate Variability & Extremes • Explainable/Interpretable/Trustworthy AI	Reports From Previous Sessions • Hybrid Modeling • AI Architecture & CoDesign
12:15		Workshop session wrap-up and discussion motivation
12:30	Break	Break
12:45	Cross-cut Session **Hybrid Modeling (Invited Only)** Session Chair: Sivasankaran Rajamanickam	Panel/Open Discussion (Invited Only) **Common challenges & opportunities Resources, capabilities, and facilities** (DOE + Multi-agency)
14:45	Break	Break
15:00	Cross-cut Session **AI Architecture Co-Design (Invited Only)** Session Chair: Jim Ang	Panel/Open discussion (Invited Only) **Short-term, 5-year, 10-year goals Earth system predictability and applied math and computer science research priorities**
17:00	Adjourn	Adjourn

WEEK 6		
North American Eastern Time	**Day 11** **December 7 or 8, 2021**	
12:00	Authors Meeting (Invited Only)	

- Public sessions (highlighted in green) are open to anyone; requires registration at https://ai4esp.org/workshop/
- Invitation-only sessions (highlighted in pink) are open to invited active participants and selected listening participants; requires registration on the Google Form link in the invitation email