U.S. DEPARTMENT OF
ENERGY
Office of Science

Exploiting Artificial Intelligence for Advanced Earth System Predictability

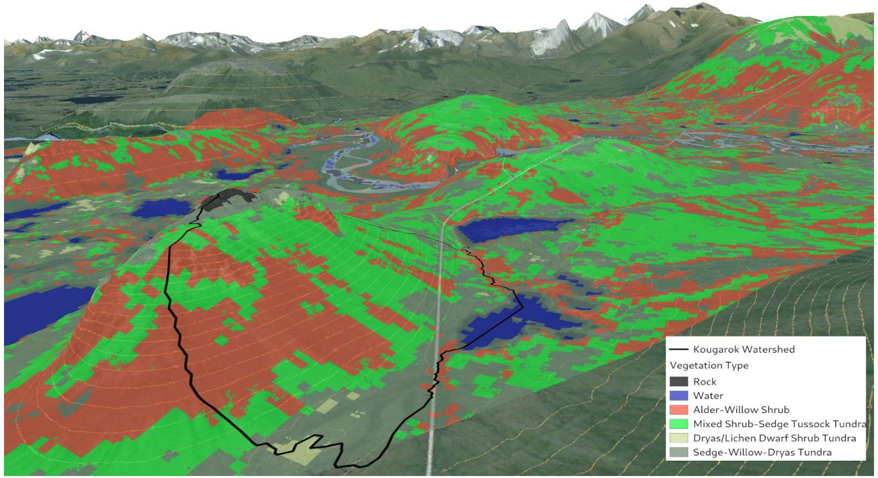
Forrest M. Hoffman
Oak Ridge National Laboratory

FASTMath Data and Optimization Symposium

February 22, 2021

Introduction

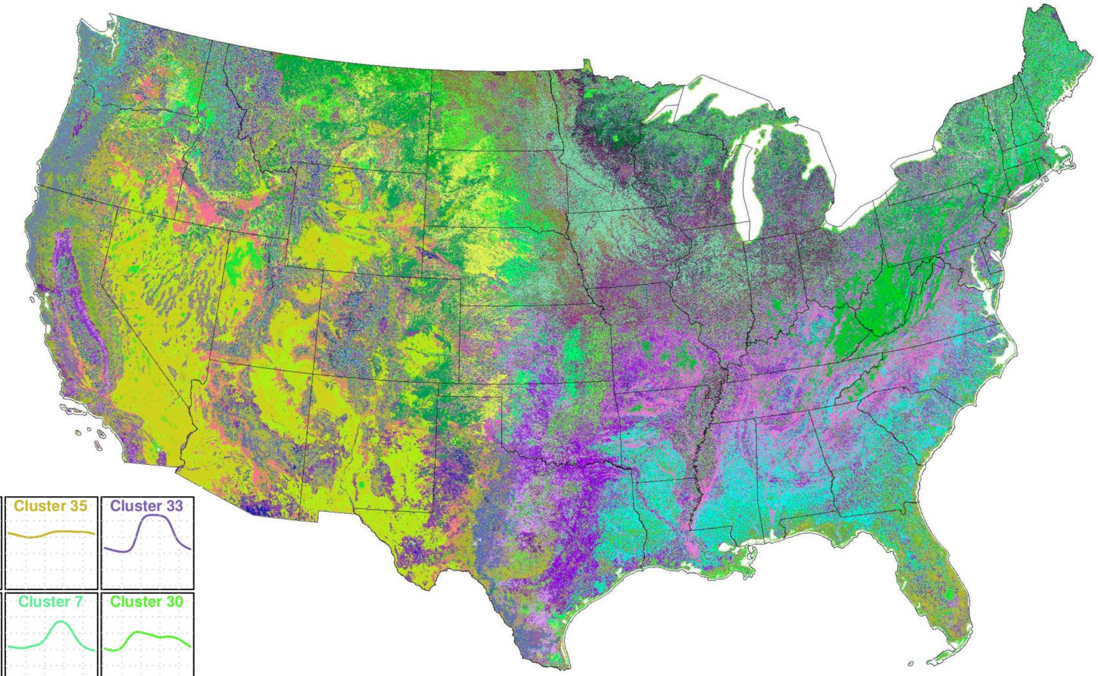
- Observations of the Earth system are increasing in spatial resolution and temporal frequency, and will grow exponentially over the next 5–10 years
- With Exascale computing, simulation output is growing even faster, outpacing our ability to evaluate and benchmark model results
- Explosive data growth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

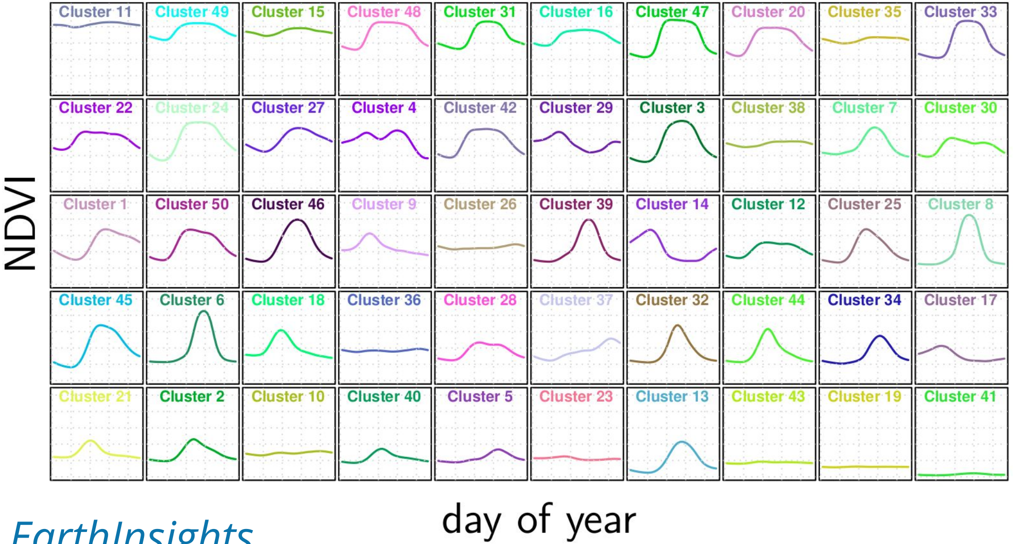


Langford et al. (2019)

50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI
Clustered from 2000 to present



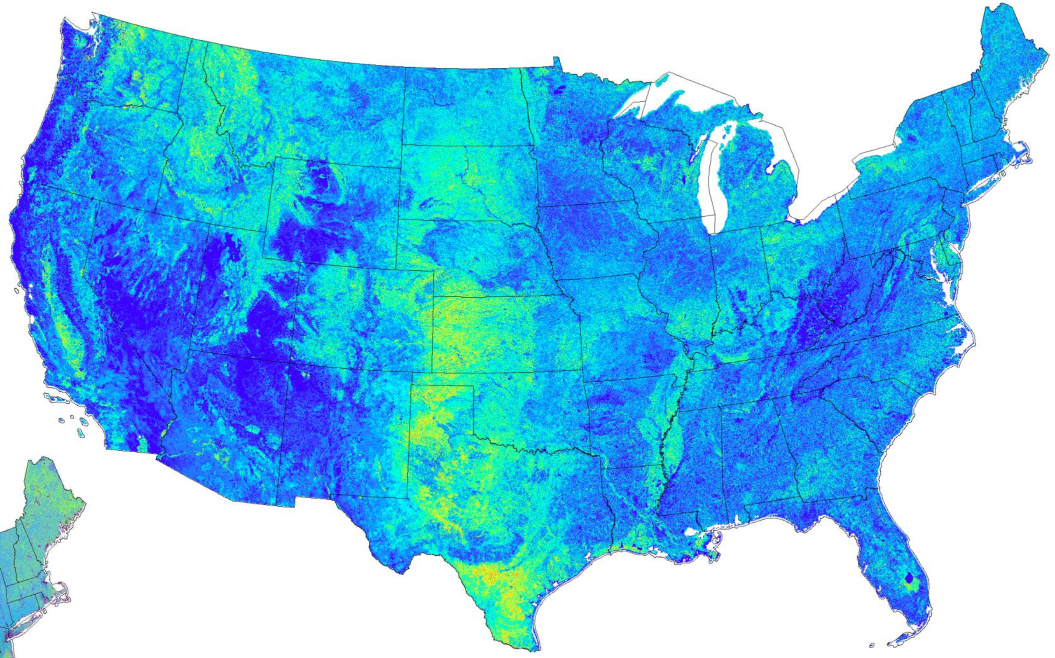


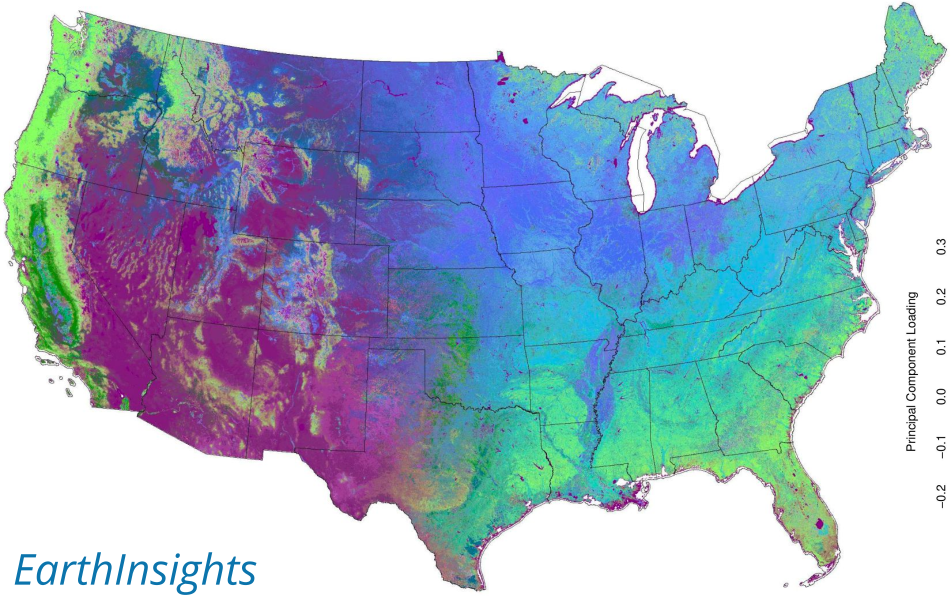
50 Phenoregion Prototypes (Random Colors)

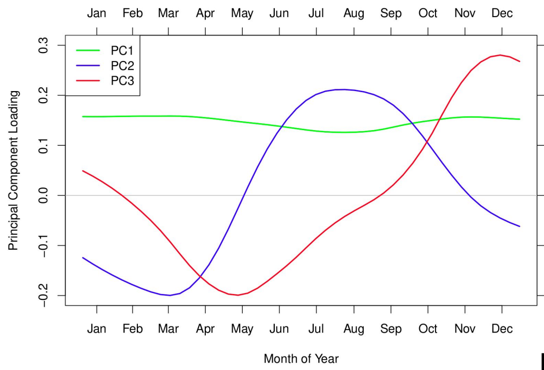
EarthInsights

Hargrove et al., in prep.

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







Principal Components Analysis

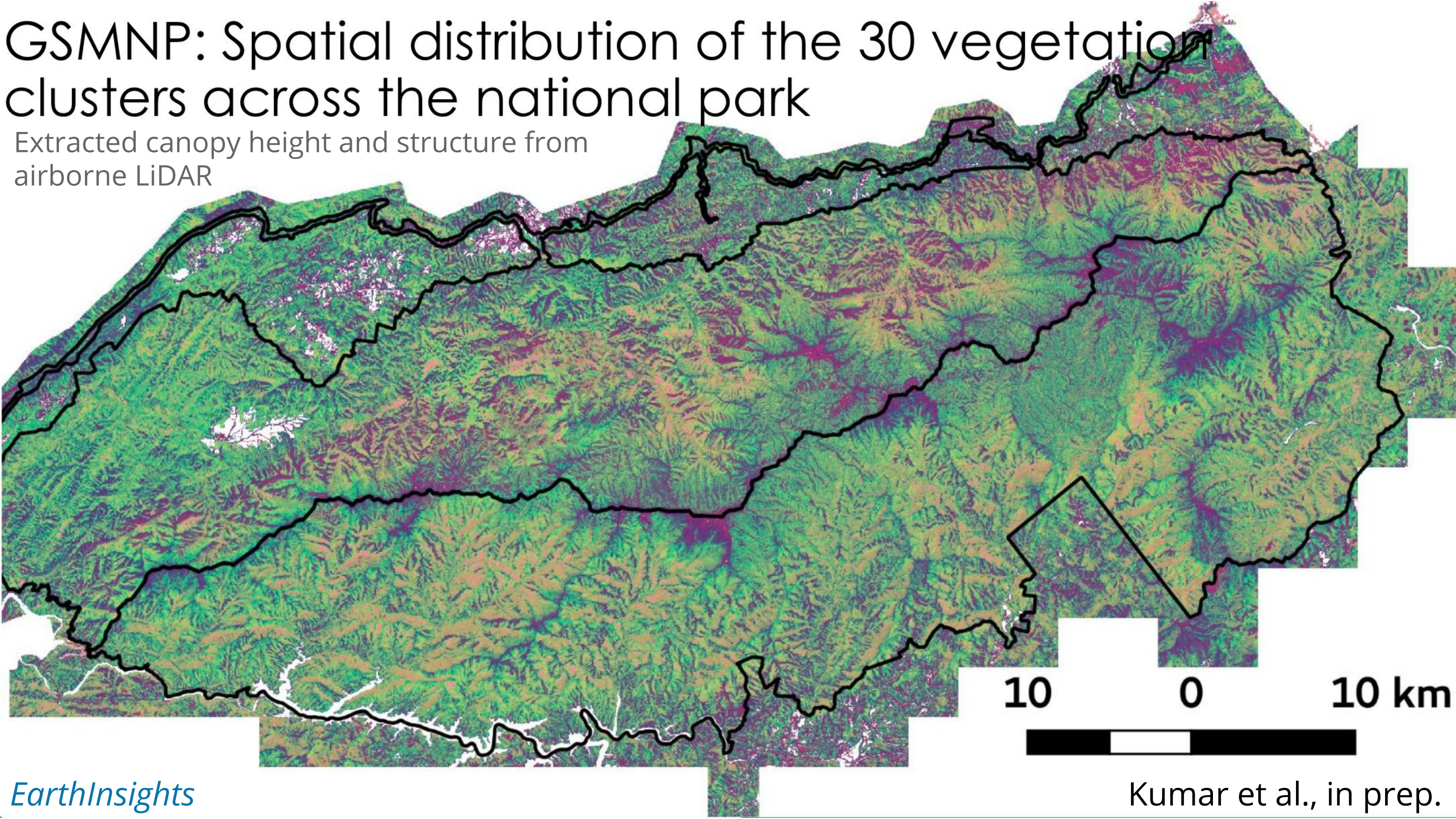
PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

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

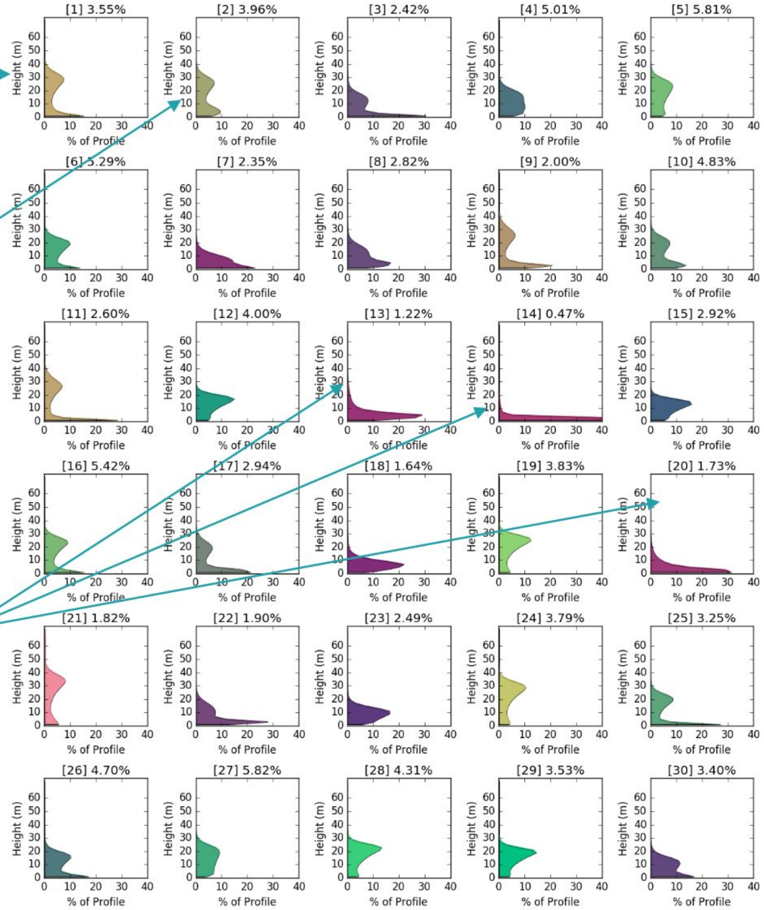
Kumar et al., in prep.

GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

forests with slightly lower mean height with dense understory vegetation

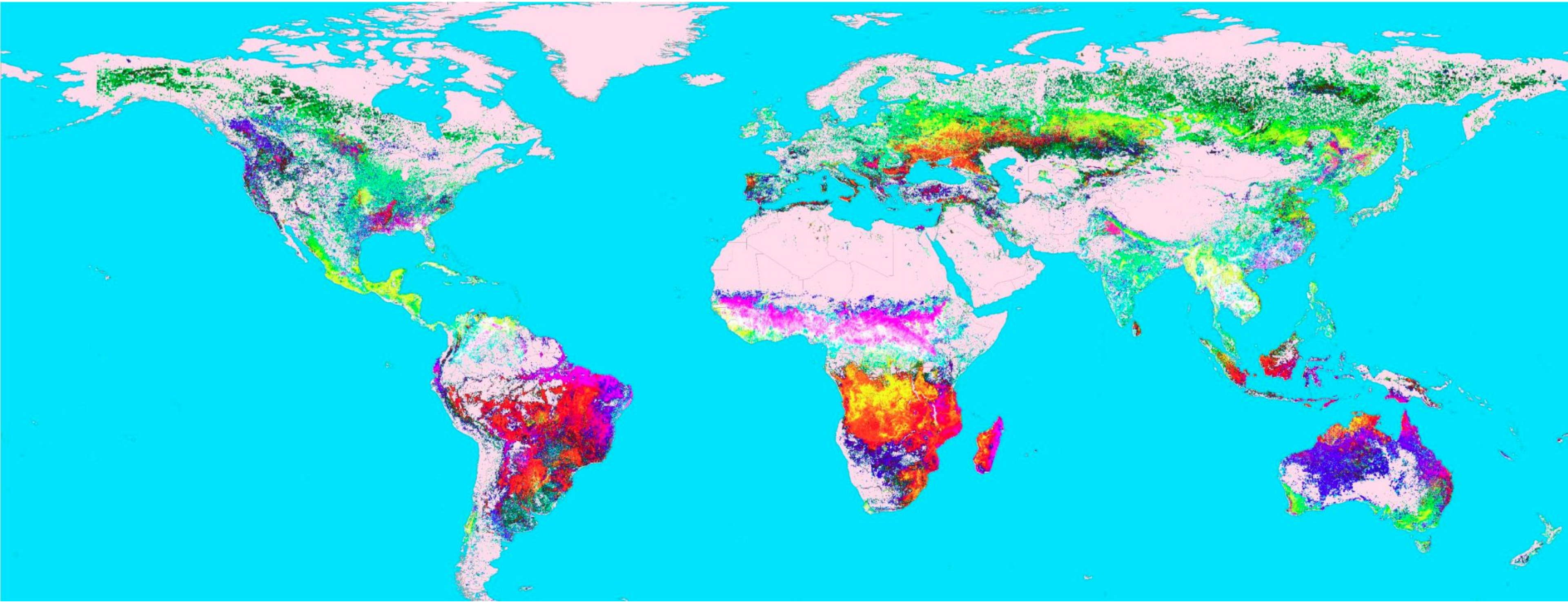
low height grasslands and heath balds that are small in area but distinct landscape type



EarthInsights

Kumar et al., in prep.

Global Fire Regimes



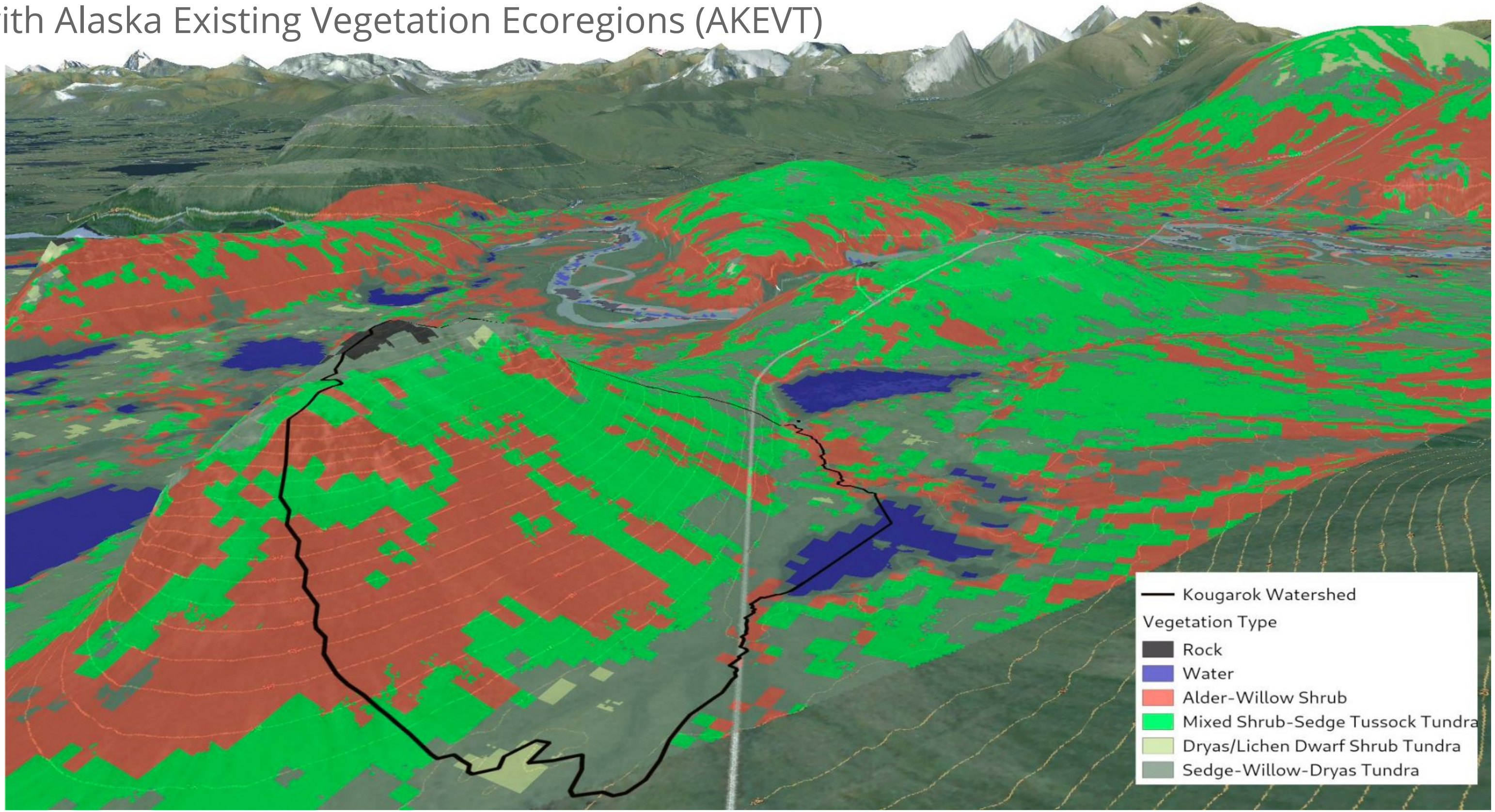
Regions that exhibit similar fire seasonality globally

From MODIS “Hotspots” from 2002–2018

Norman et al., in prep.

Arctic Vegetation Mapping from Multi-Sensor Fusion

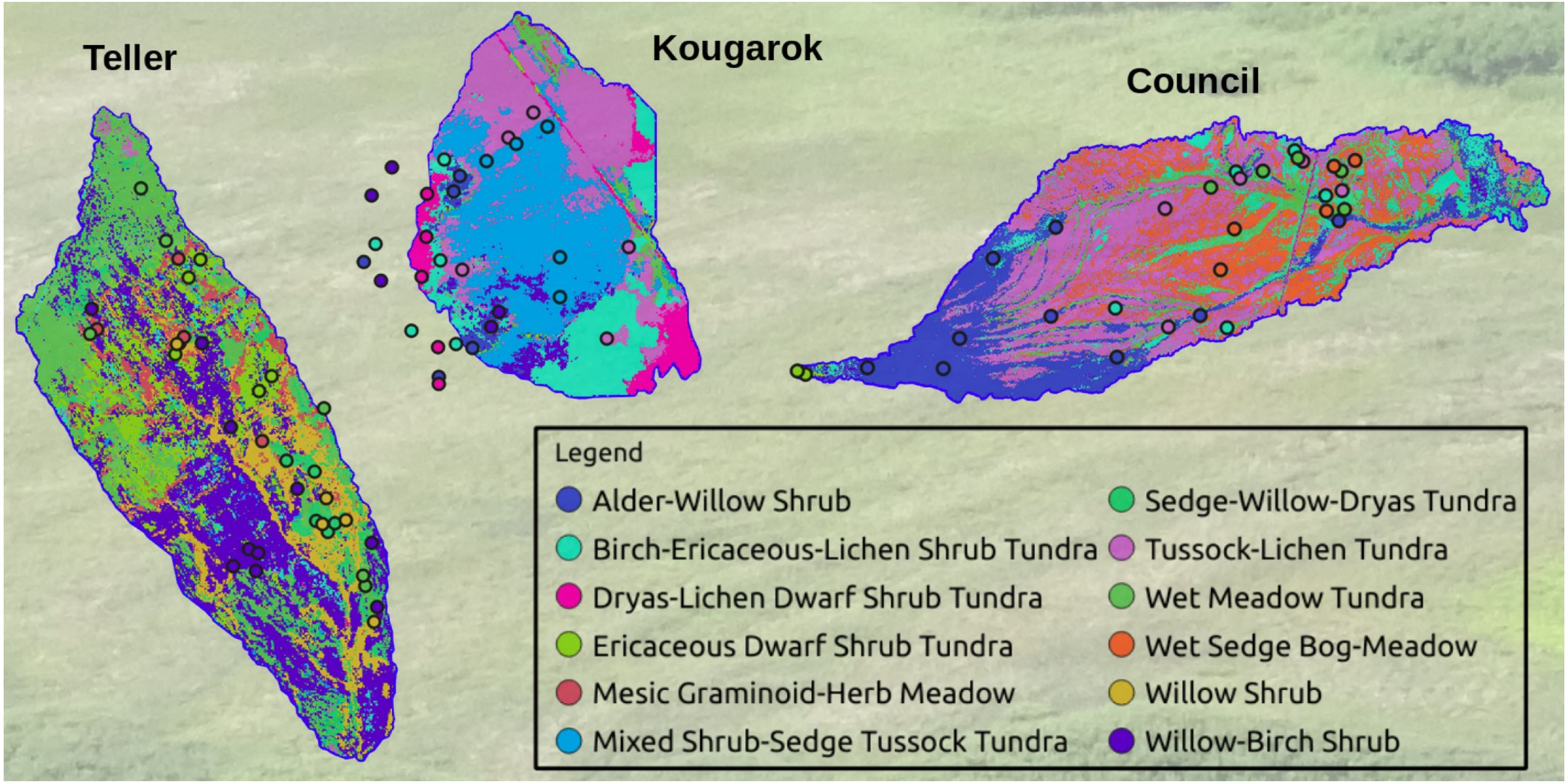
Using Hyperion Multispectral and IfSAR-derived Digital Elevation Model
Trained with Alaska Existing Vegetation Ecoregions (AKEVT)



EarthInsights

Langford et al., *Remote Sensing*, 2019

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



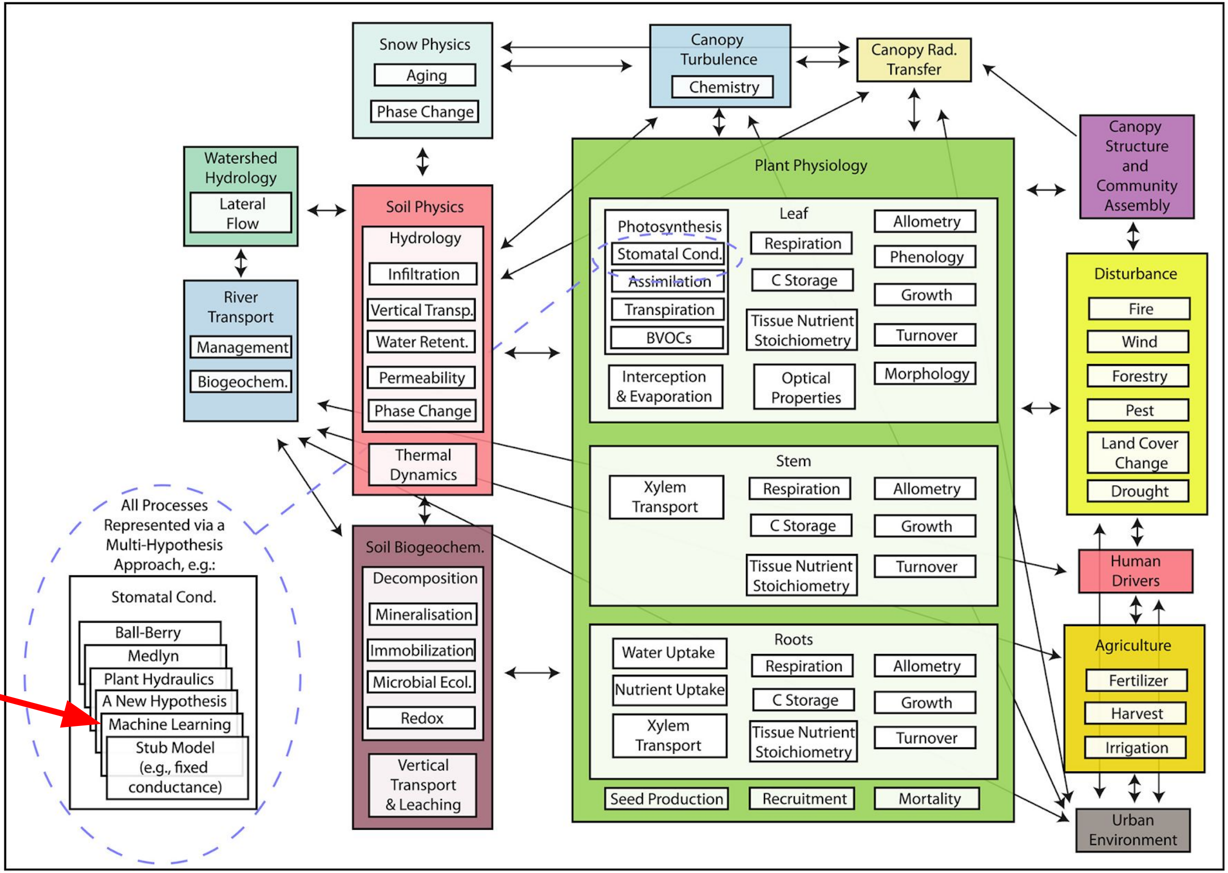
EarthInsights

Konduri et al., in prep.

Hybrid ML/Process-based Modeling for Terrestrial Modeling

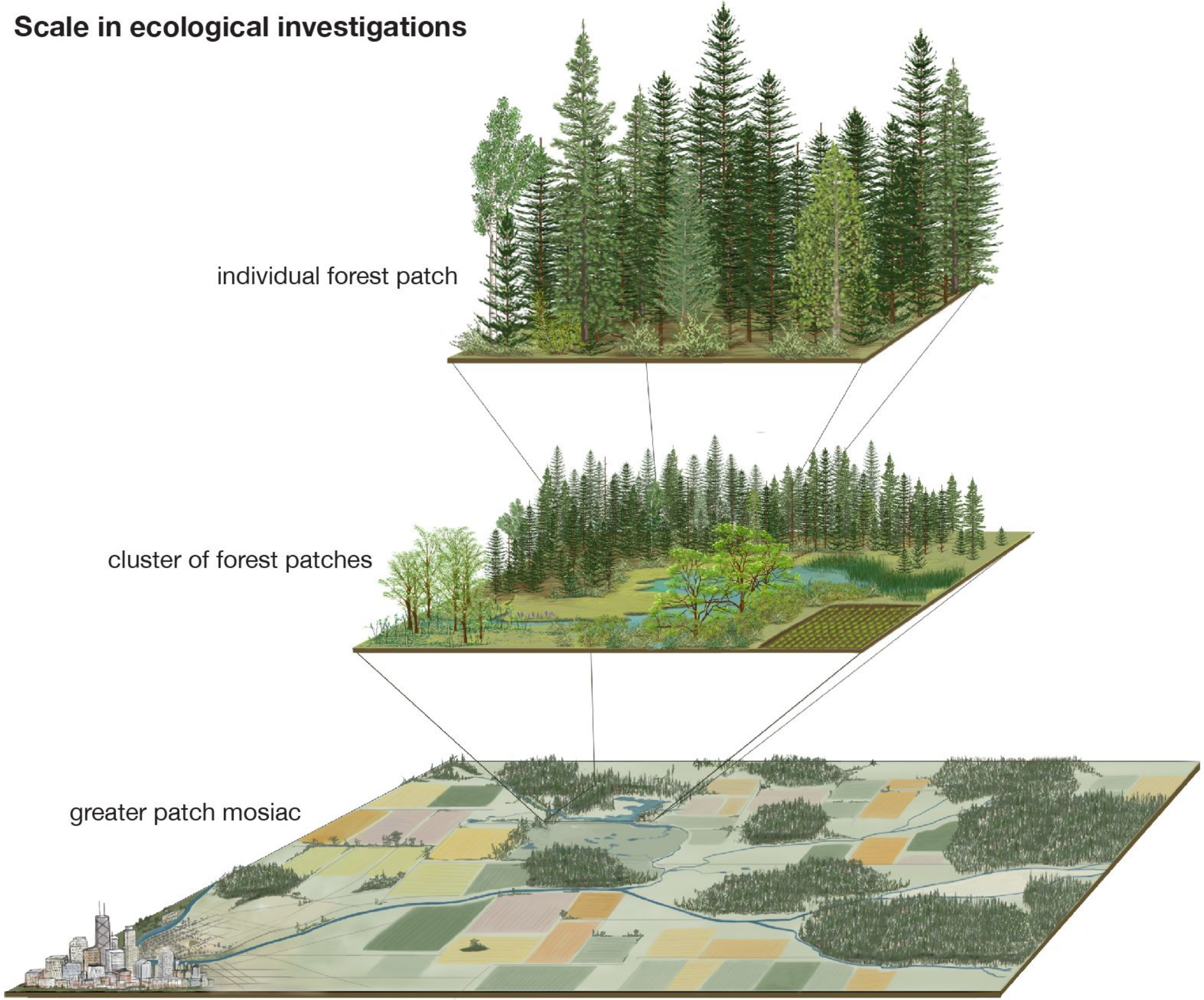
Individual processes can be represented by a multi-hypothesis approach, and ML provides an opportunity for a data-derived hypothesis that can be further explored or used to calibrate other hypotheses, when sufficient data are available.

(Fisher and Koven, 2020)



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

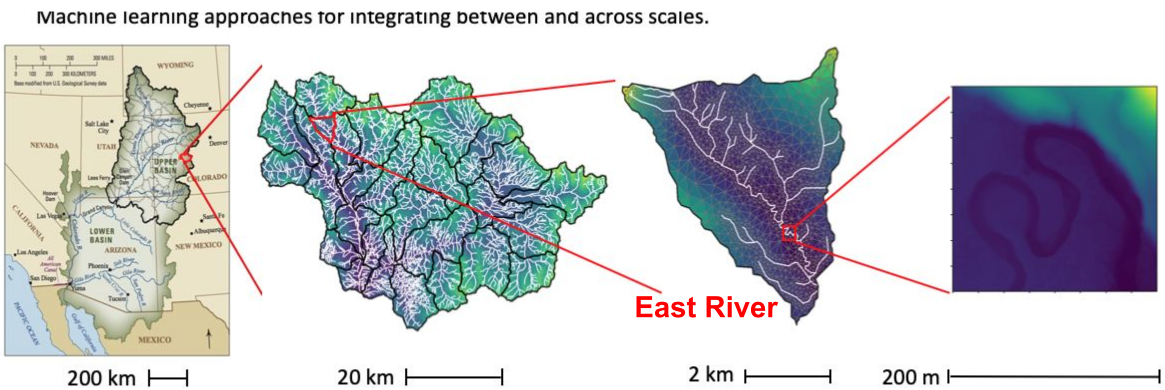
Scale in ecological investigations

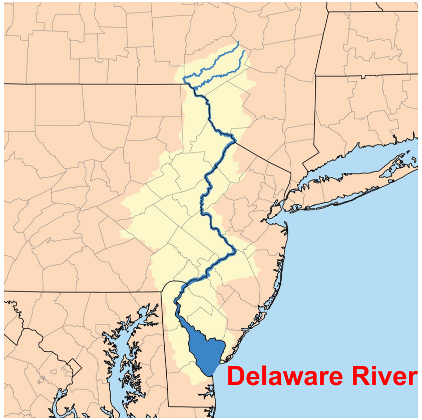


ExaSheds

A novel multiscale strategy fusing process-resolving simulations and machine learning

- Tightly integrated role for machine learning
  - Synthesize spatially distributed model inputs from diverse data streams; use inverse modeling; apply surrogate models across scales; co-design distributed sensor networks through feedback from modeling





- Process-explicit integrated surface/subsurface flow and reactive transport codes
  - Represent biogeochemical processes and their hydrologic controls at their native scales; leverage exascale computing; exploit high throughput model-data integration

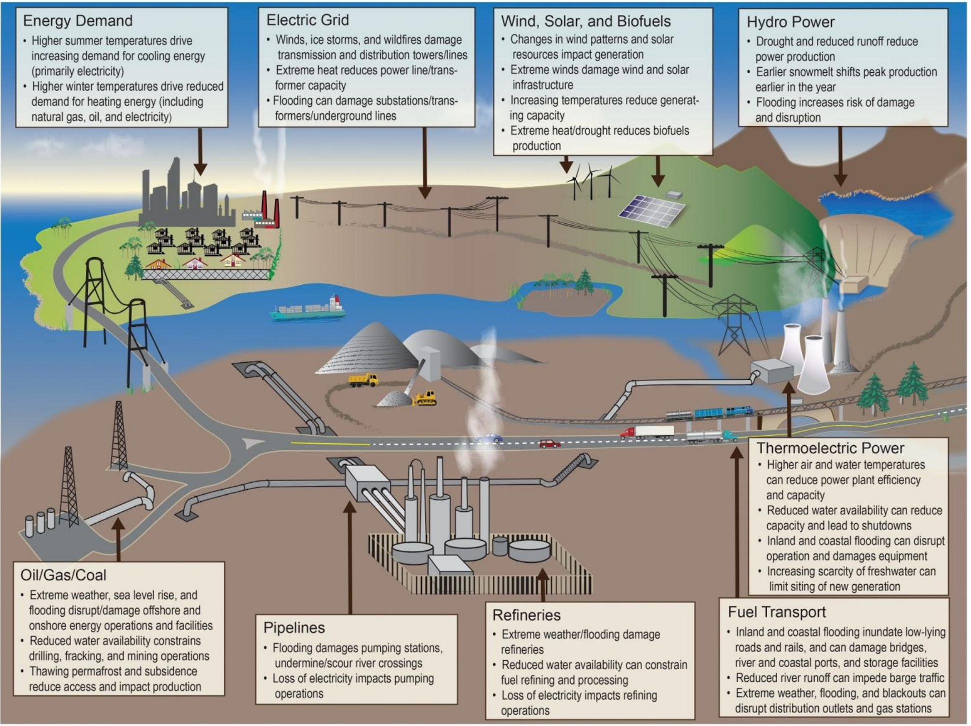
AI FOR SCIENCE TOWN HALL

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

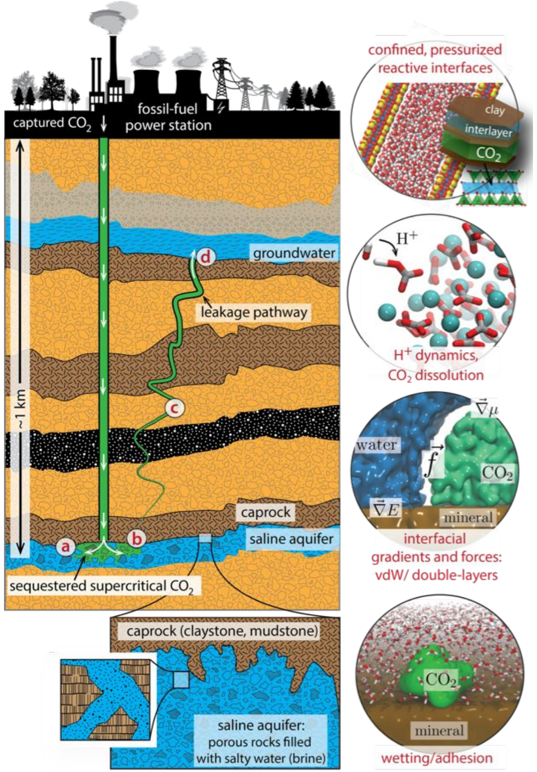
- Increasing frequency of weather extremes and changing environment pose risks to energy infrastructure and the built environment
- Sparse observations and inadequate model fidelity limit the ability to identify vulnerability, mitigate risks, and respond to disasters

U.S. DEPARTMENT OF ENERGY | Office of Science

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

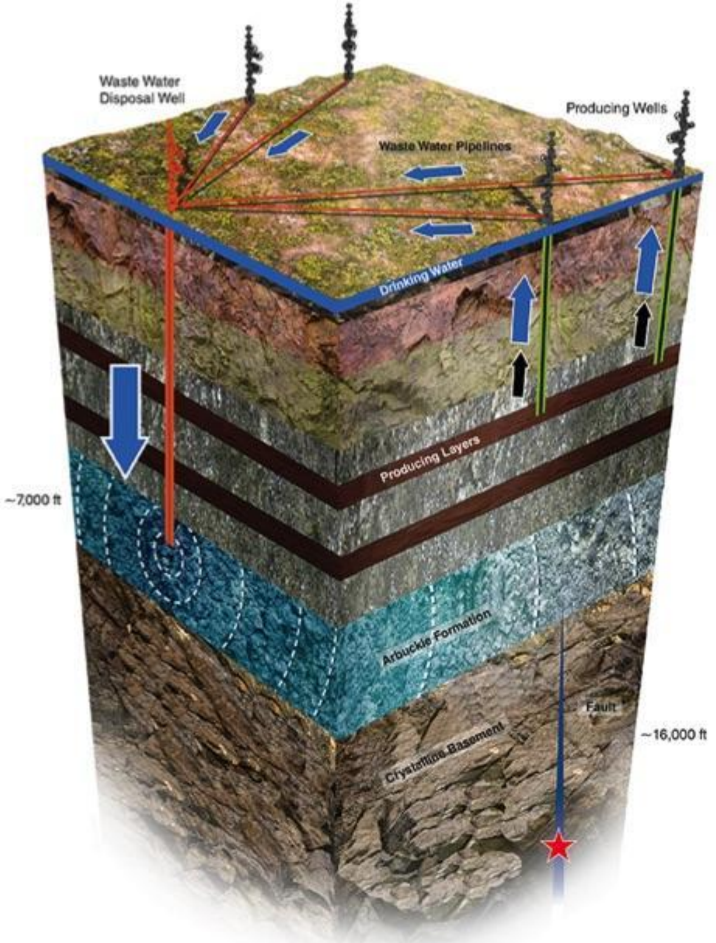


Characterize and modify subsurface conditions for responsible energy production, CO_2 storage, and contaminant remediation

- 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

Grand Challenge #2

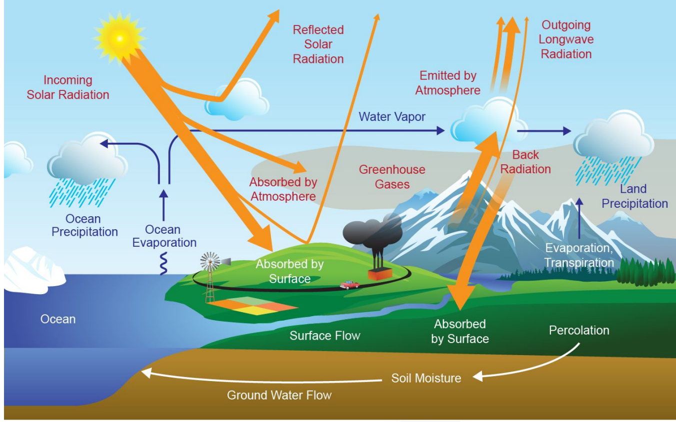
- We need to substantially increase hydrocarbon extraction efficiency, discover and exploit hidden geothermal resources, reduce induced seismicity and other impacts, improve geologic CO_2 storage, and predict long-term fate and transport of contaminants
- Mitigating risks requires improved subsurface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



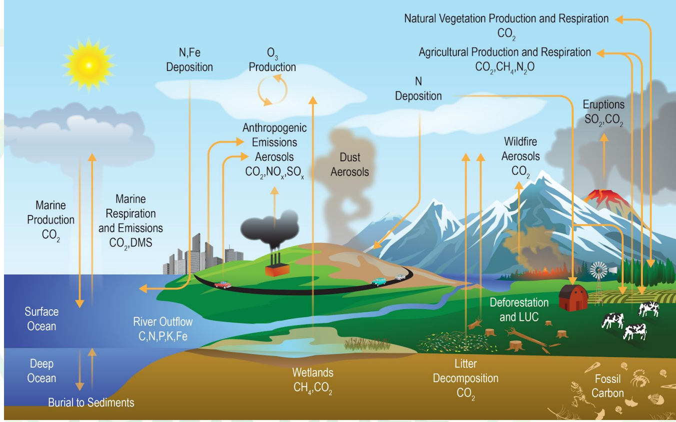
Grand Challenge #3

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



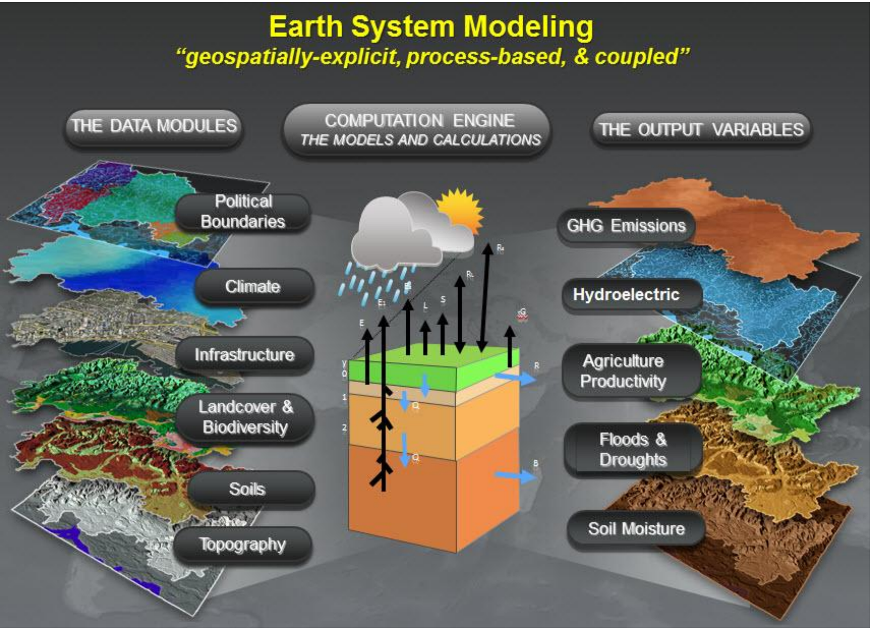
Energy & Water Cycles



Carbon & Biogeochemical Cycles

Washington DC Town Hall
October 22-23

Grand Challenge #3

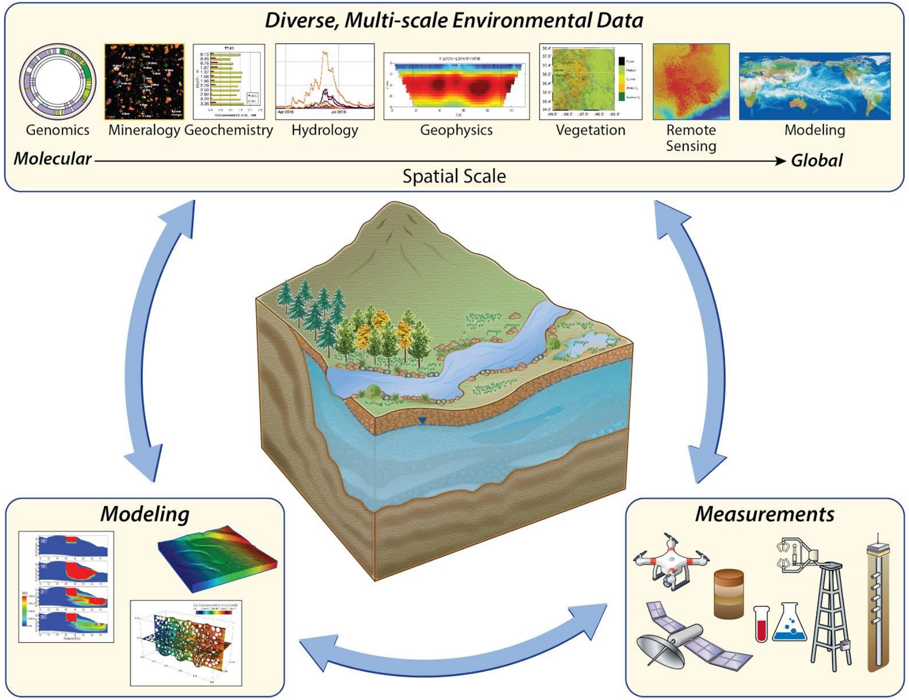


- Accurate predictions are needed to quantify changes in atmospheric and ocean circulation and weather extremes, to close the carbon cycle, and to understand responses and feedbacks of human, terrestrial, and marine ecosystems to environmental change
- Advances in genomics and bioscience data need to 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

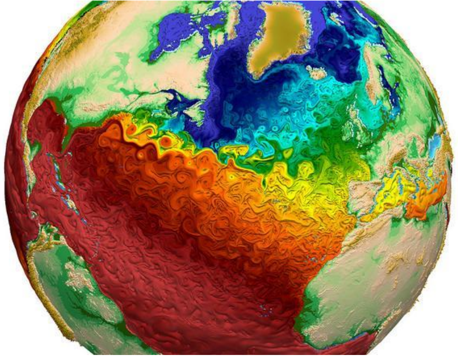
- Water resources are critical for energy production, human health, food security, and economic prosperity
- Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change



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

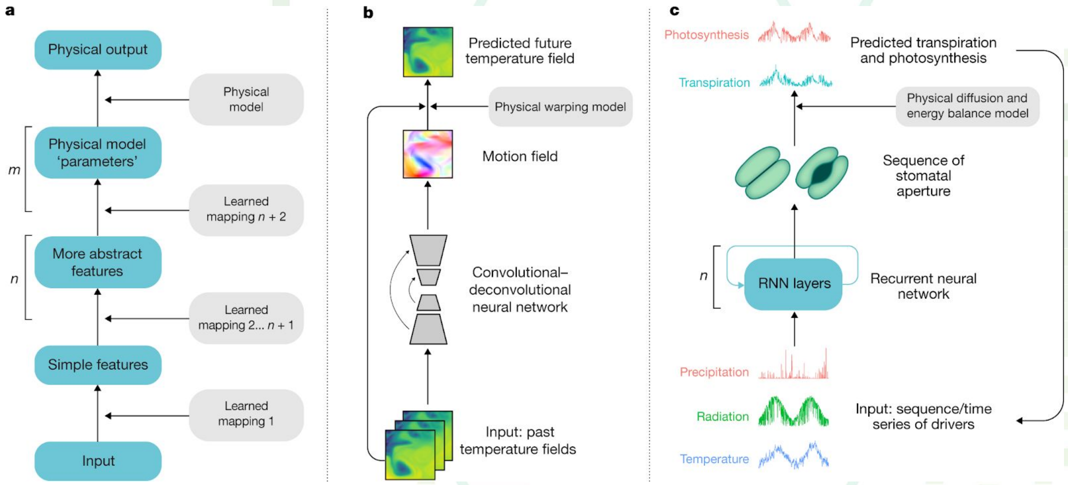
Accelerating Development



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)

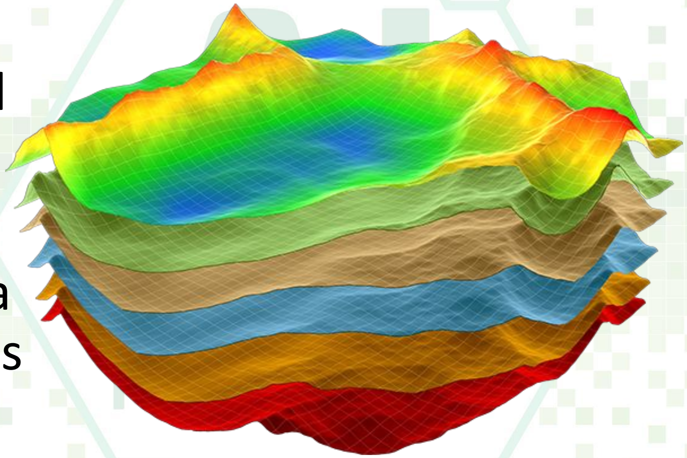


U.S. DEPARTMENT OF ENERGY | Office of Science

Washington DC Town Hall
October 22-23

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



U.S. DEPARTMENT OF ENERGY

Office of Science

AI for Earth System Predictability (AI4ESP): A Vision for a Machine Learning Framework Enabling End-to-End Earth System Predictability Research

Nicki L. Hickmon[1], Scott M. Collis[1], Forrest M. Hoffman[2], and Haruko Wainwright[3]

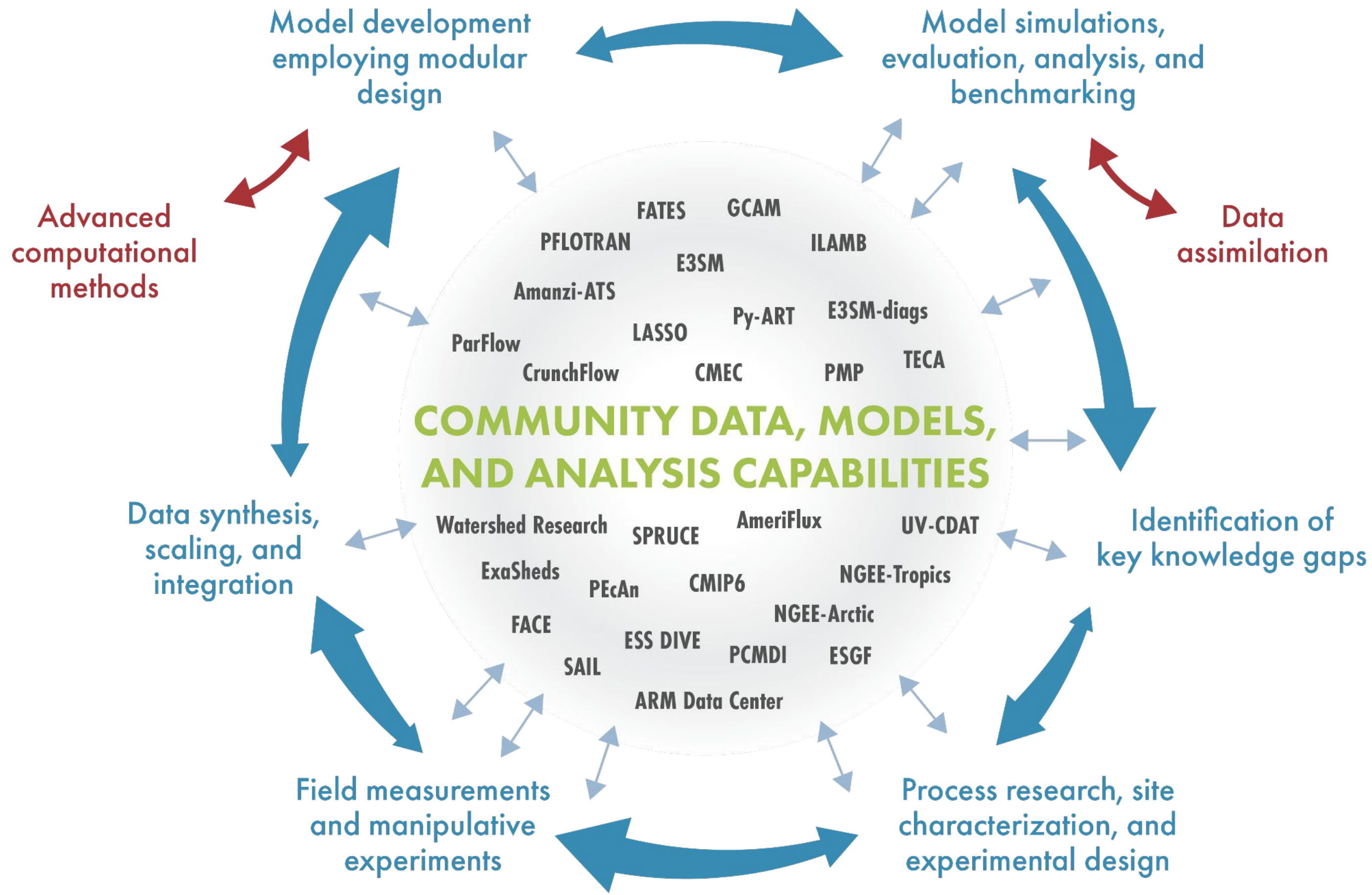
[1]Argonne National Laboratory
[2]Oak Ridge National Laboratory
[3]Lawrence Berkeley National Laboratory

Developing Vision Enabling AI from Obs to Earth System Models

- What does the framework look like that combines DOE's experiment/observation and simulation capabilities to quantify and reduce the uncertainty in high-resolution Earth systems models?

- Goal: Define the paradigm shift required to employ artificial intelligence and machine learning across field, lab, modeling, and analysis activities. Non-incremental advancement built for the future EESSD program needs (5-10yr).
  - Bridge the gap: state-of-the-art in AI/ML research & EESSD program needs
  - Harness Earth System Data including inter-agency resources
  - Harness DOE computing resources, i.e. Exascale
  - Domain-specific machine learning applications

DOE's Model-Data-Experiment Enterprise



Novel AI/ML Framework for Land–Atmosphere Interactions

- Domain-specific machine learning applications from field and lab activities to models and analysis
- AI/ML at every aspect in the wheel (examples; not exhaustive)
  - Simulation-guided experiment/sampling design
  - Dynamic/responsive AI-controlled measurement systems
  - Edge computing and 5G sensor networks
  - Pattern recognition and process discovery through large data
  - Hybrid process-/machine learning-based coupled Earth system modeling
  - Data-driven multiscale modeling and data–model integration and analytics

White Paper Call

AI methods in areas relevant to EESSD research emphasis on quantifying and improving Earth system predictability (particularly the integrative water cycle and associated water cycle extremes)

Focal Areas (helps organize the responses for designing workshops):

1. Data acquisition and assimilation enabled by ML, AI, and advanced methods
2. Predictive modeling using AI techniques and AI-derived model components; use of AI and other tools to design a prediction system comprising of a hierarchy of models
3. Insight gleaned from complex data (both observed and simulated) using AI, big data analytics, and other advanced methods, including explainable AI and physics- or knowledge-guided AI

DOE-BER-EESSD AI/ML Activity

- Core Team, Lab POCs, and DOE Management will develop a vision for workshops based on AI4ESP White Papers (submitted February 15) for a new paradigm for Earth system predictability focused on enabling artificial intelligence and machine learning across field, lab, modeling, and analysis activities

- How to engage:
  - **Sign up** for more information at http://bit.ly/MLAI4earth
  - **Join Slack Workspace** at https://join.slack.com/t/ai4esp/signup