Artificial Intelligence and Machine Learning for Advancing Predictive Process Understanding

Forrest M. Hoffman[1], Jitendra Kumar[1], Zachary L. Langford[1], V. Shashank Konduri[2], Auroop R. Ganguly[3], Zheng Shi[4], Elias Massoud[1], Nathan Collier[1], Min Xu[1], William W. Hargrove[5], Nicki L. Hickmon[6], Scott M. Collis[6], Charuleka Varadharajan[7], and Haruko Wainwright[7]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]National Ecological Observatory Network, Boulder, CO, USA
[3]Northeastern University, Boston, MA, USA
[4]University of Oklahoma, Norman, OK, USA
[5]US Department of Agriculture - Forest Service, Asheville, NC, USA
[6]Argonne National Laboratory, Lemont, IL, USA
[7]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

AI4ESS Virtual Roundtable Discussion

August 20, 2025

Introduction

U.S. DEPARTMENT of ENERGY | Office of Science

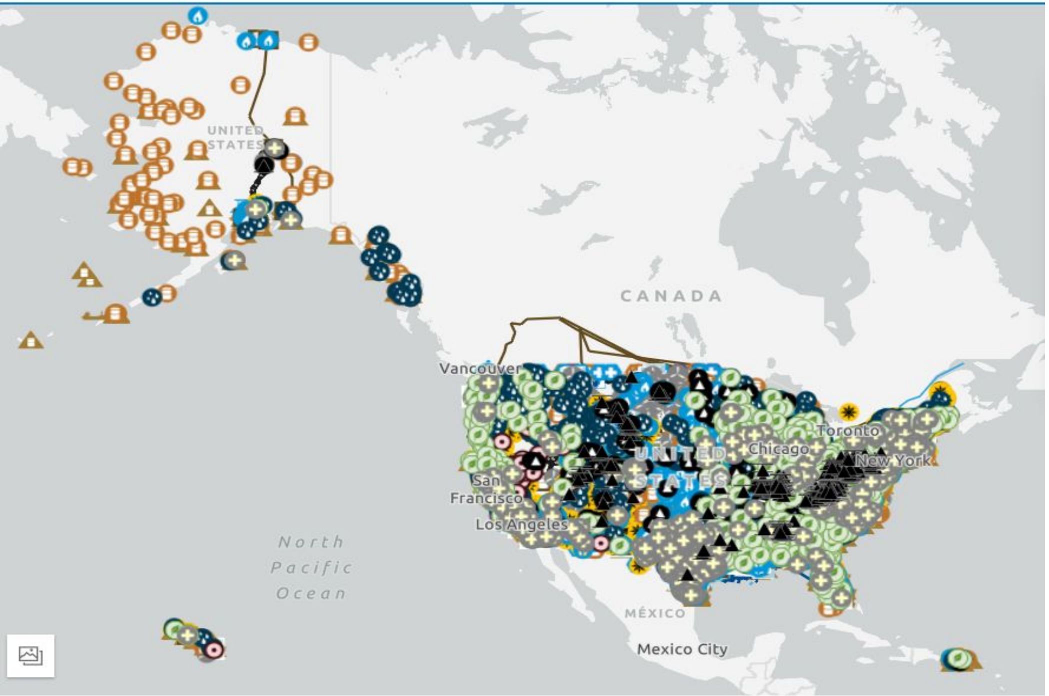
- Observations of the Earth system are increasing in spatial resolution and temporal frequency, and will grow exponentially over the next 5–10 years
- With Exascale computing, simulation output is growing even faster, outpacing our ability to analyze, interpret and evaluate model results
- Explosive 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 #2 fastest supercomputer on the [TOP500](TOP500) List and the first supercomputer to break the exaflop barrier (June 10, 2025).

US Energy Infrastructure and Resources

U.S. DEPARTMENT of ENERGY | Office of Science

All Energy Infrastructure and Resources

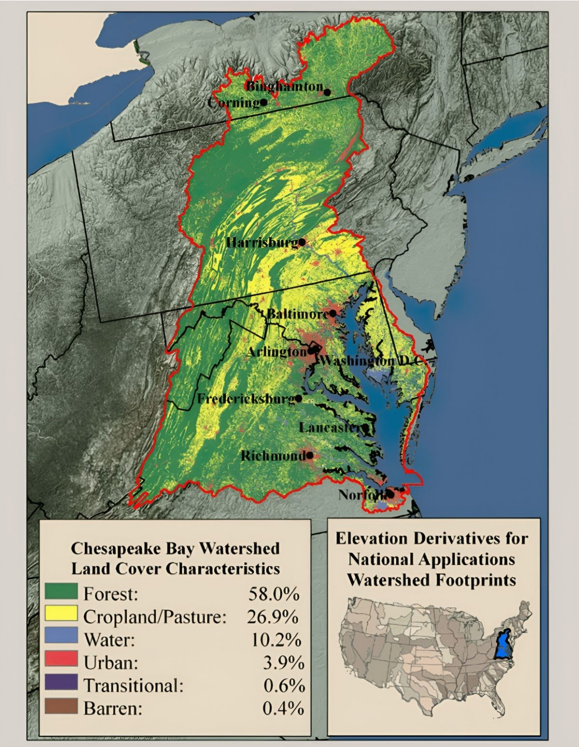


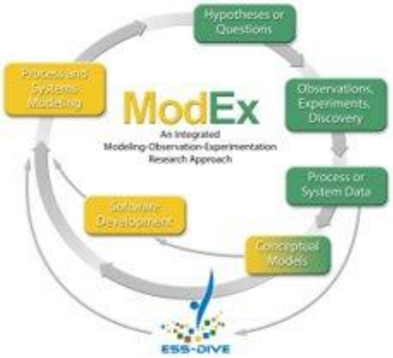
https://atlas.eia.gov/apps/5039a1a01ec34b6bbf0ab4fd57da5eb4/explore

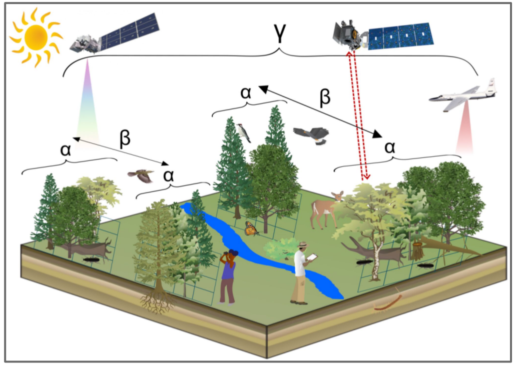
- US energy infrastructure is scattered across all states and territories
- Resources include
  - Coal mines and oil wells
  - Pipelines
  - Oil/gas refining & processing
  - Power plants & generation
    - Coal and oil
    - Nuclear
    - Dams
    - Geothermal
    - Wind and solar
  - Power transmission infrastructure
  - Market & trading hubs
  - More

AI/ML and ESS Observations Could Support Regional Testbeds

Local- to regional-scale **field observations** and **machine learning** to support **multi-scale modeling** for predictions of **natural hazard impacts** on **energy and water systems**

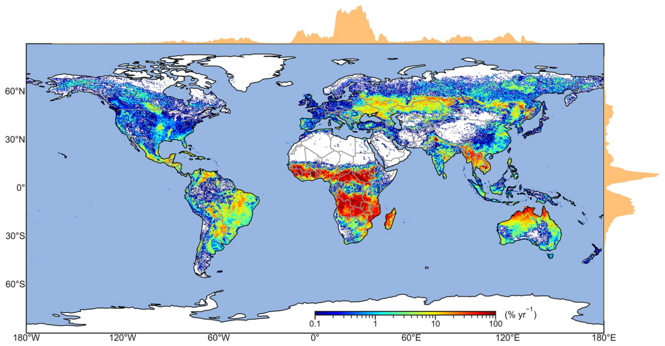


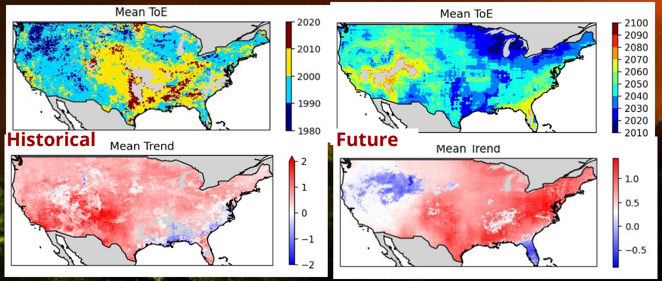




Wildfire Research

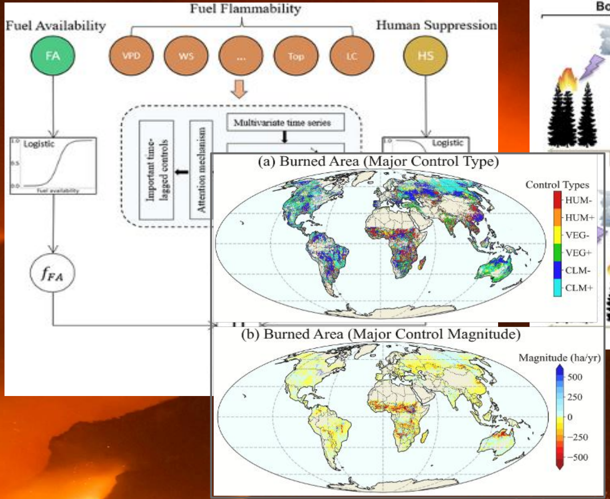
Satellite Wildfire Fluxes & Extent



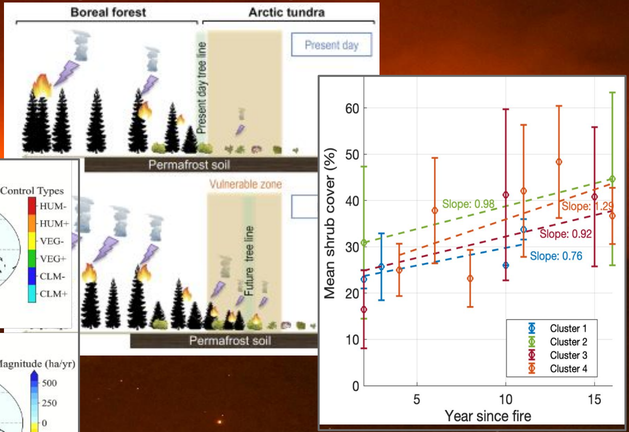


Changing Fire Weather

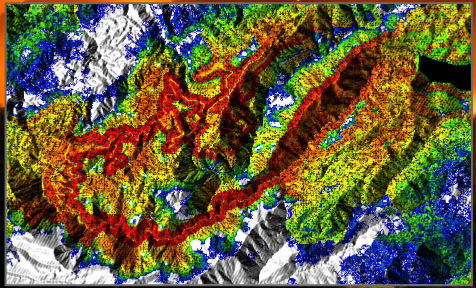
AI/ML Wildfire Prediction



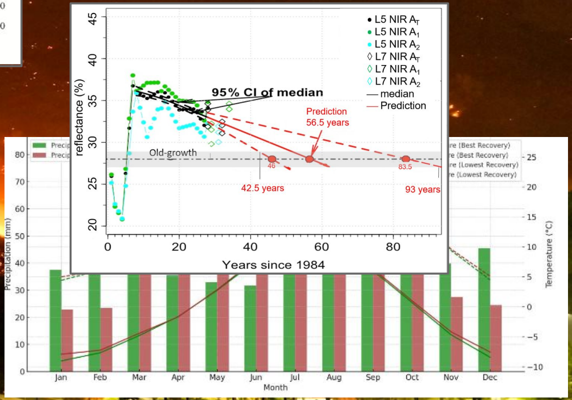
Vegetation Responses to Wildfire



Drivers of Fire Risk

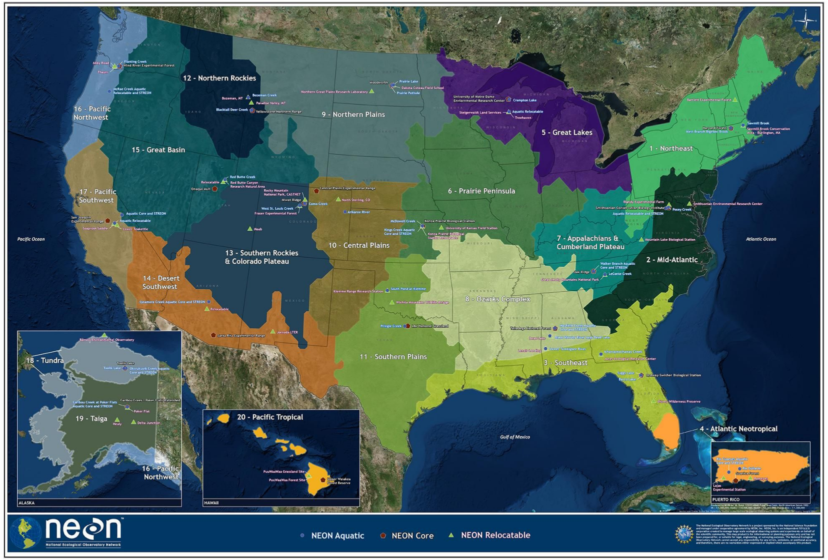


Fuel Connectivity & Fire Spread

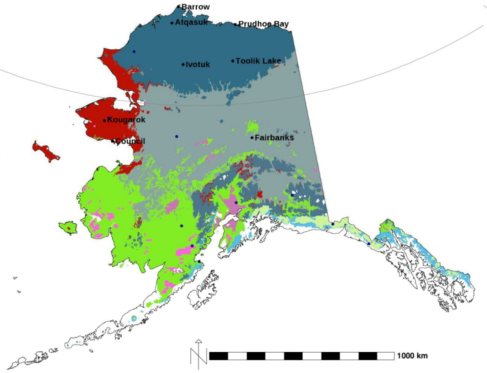


Vegetation Regrowth & Recovery

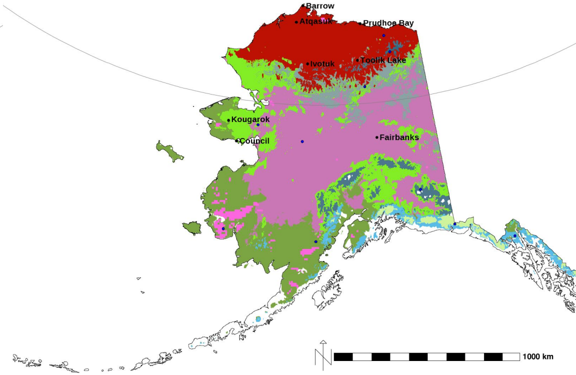
Sampling Network Design



NSF's NEON Sampling Domains



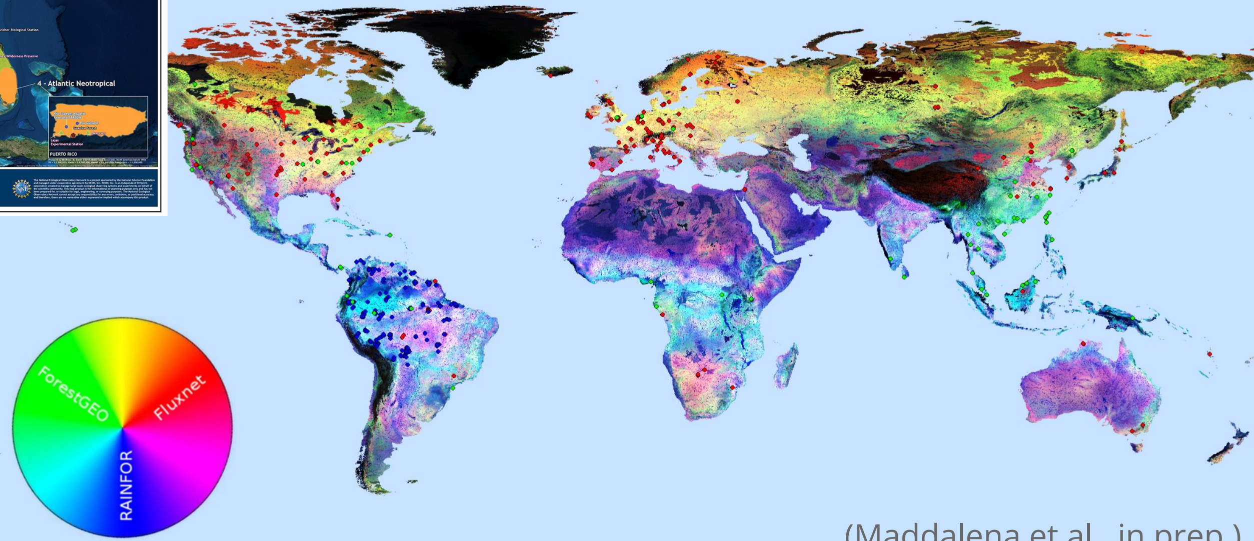
2000–2009



2090–2000

(Hoffman et al., 2013)

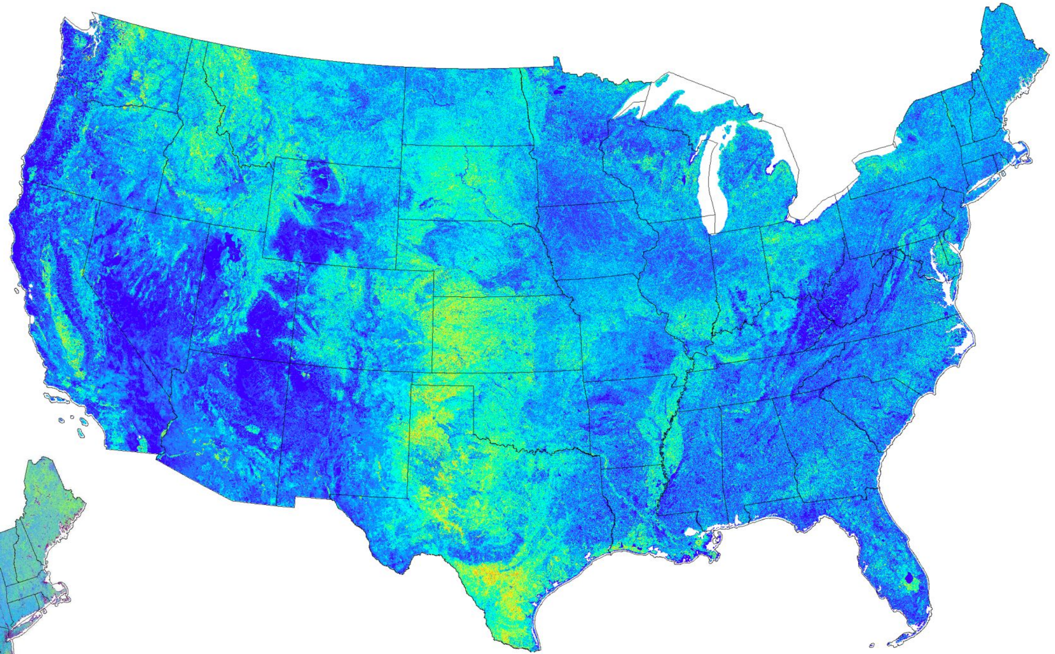
Triple-Network Global Representativeness

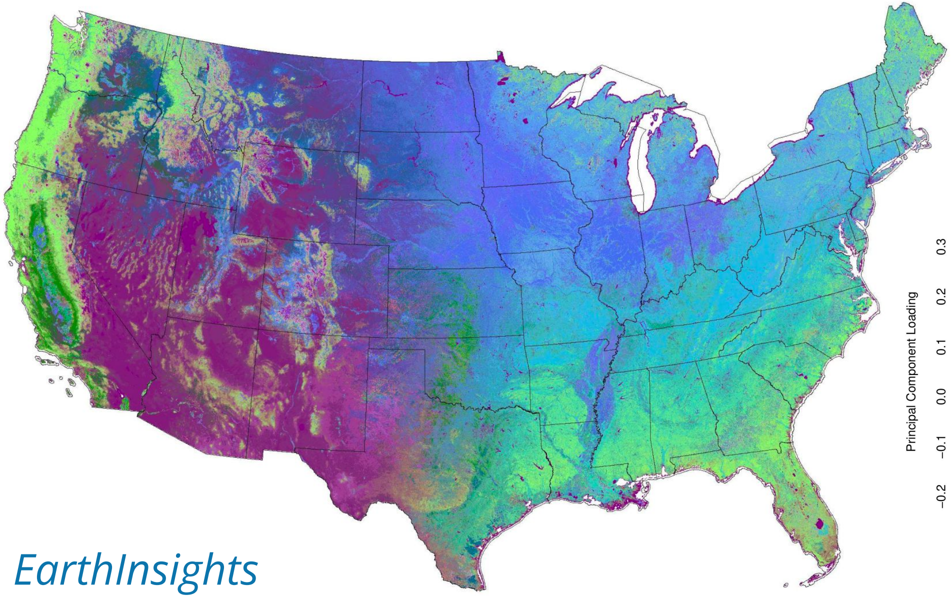


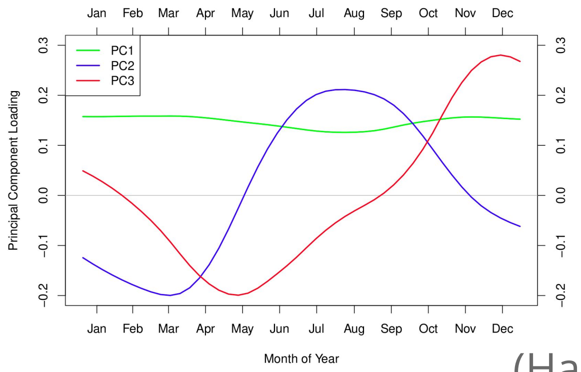
(Maddalena et al., in prep.)

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

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







Principal Components Analysis

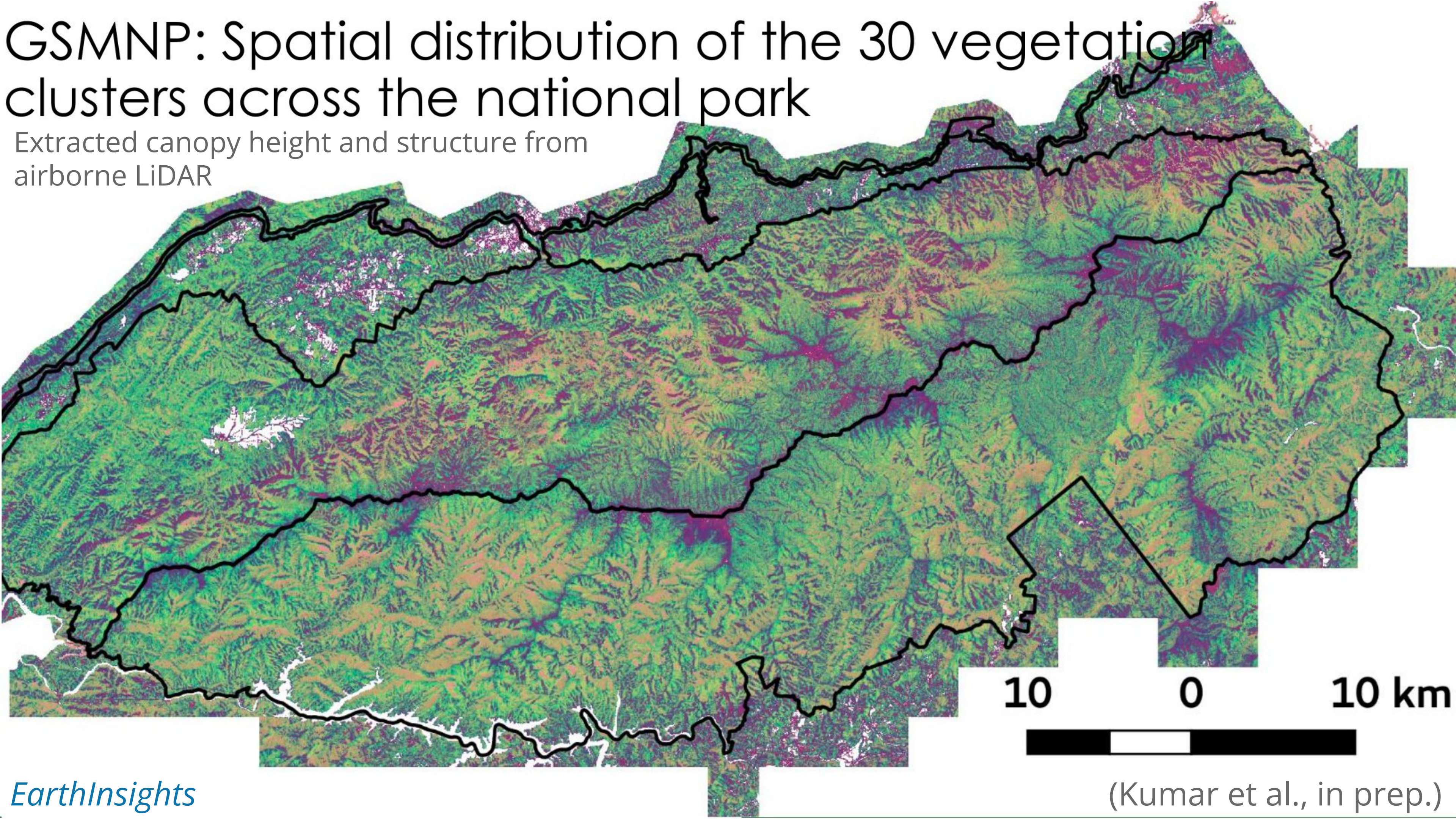
PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

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

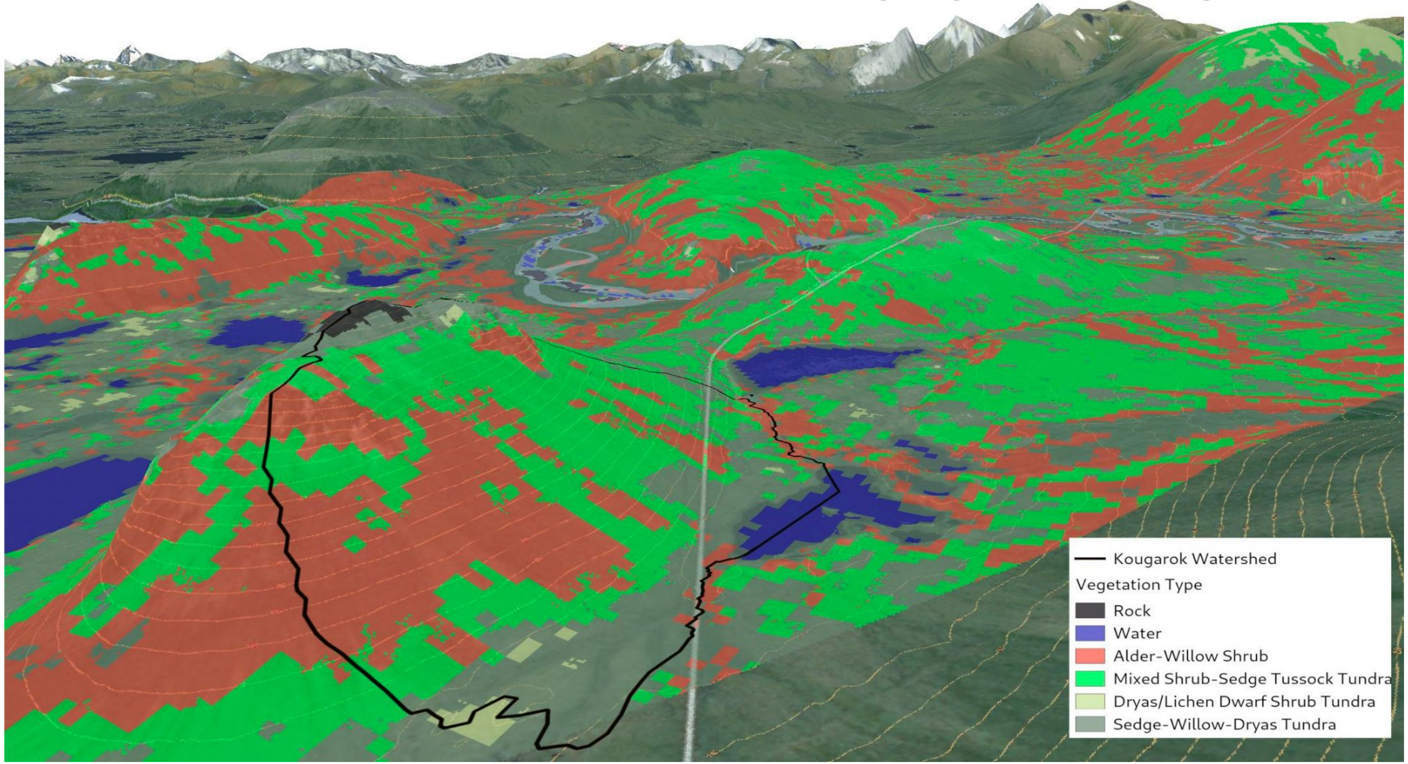


EarthInsights

(Kumar et al., in prep.)

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).

Leveraging Advances in Machine Learning for Earth Sciences

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

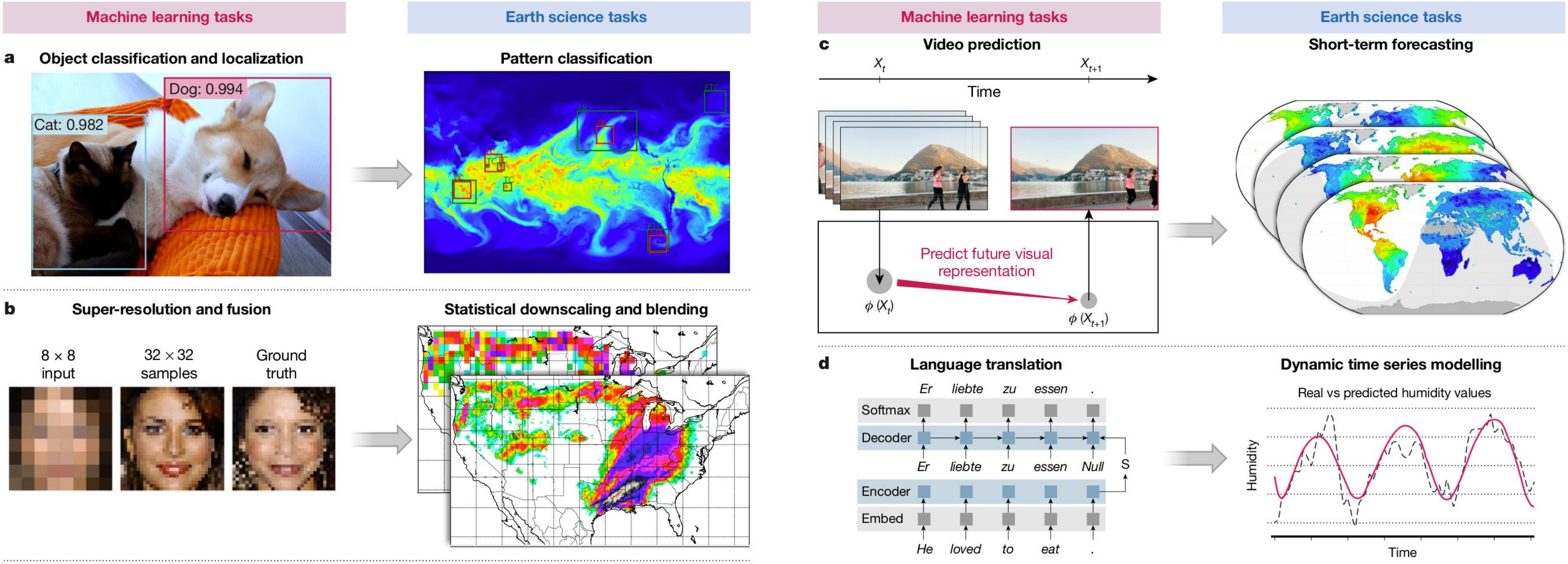
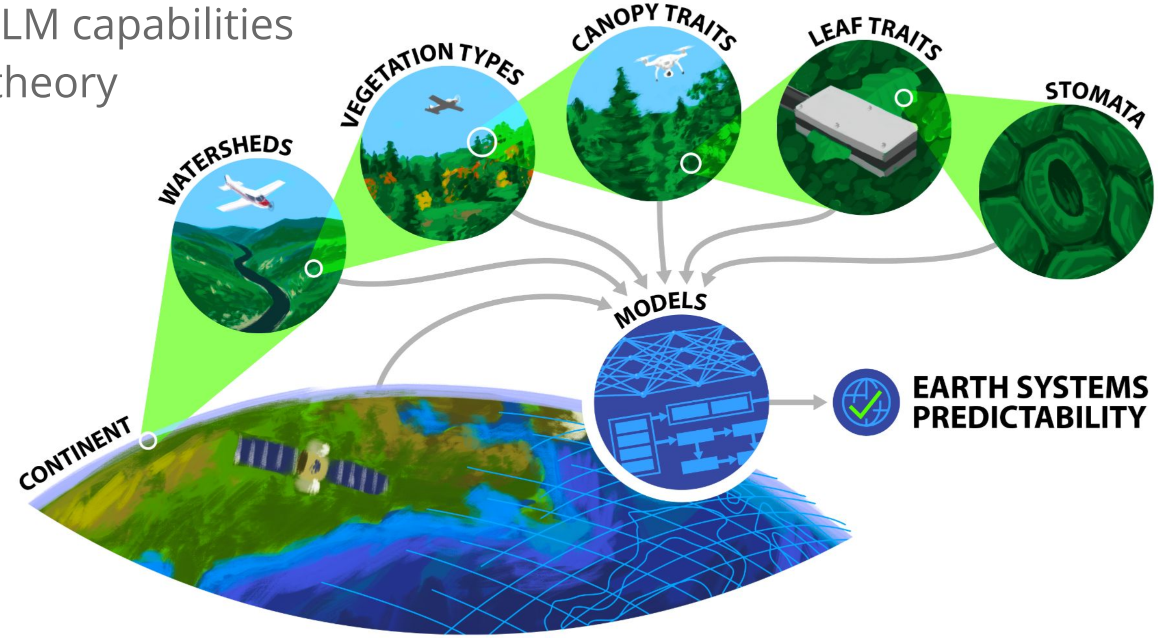


Figure 2 in Reichstein et al. (2019)

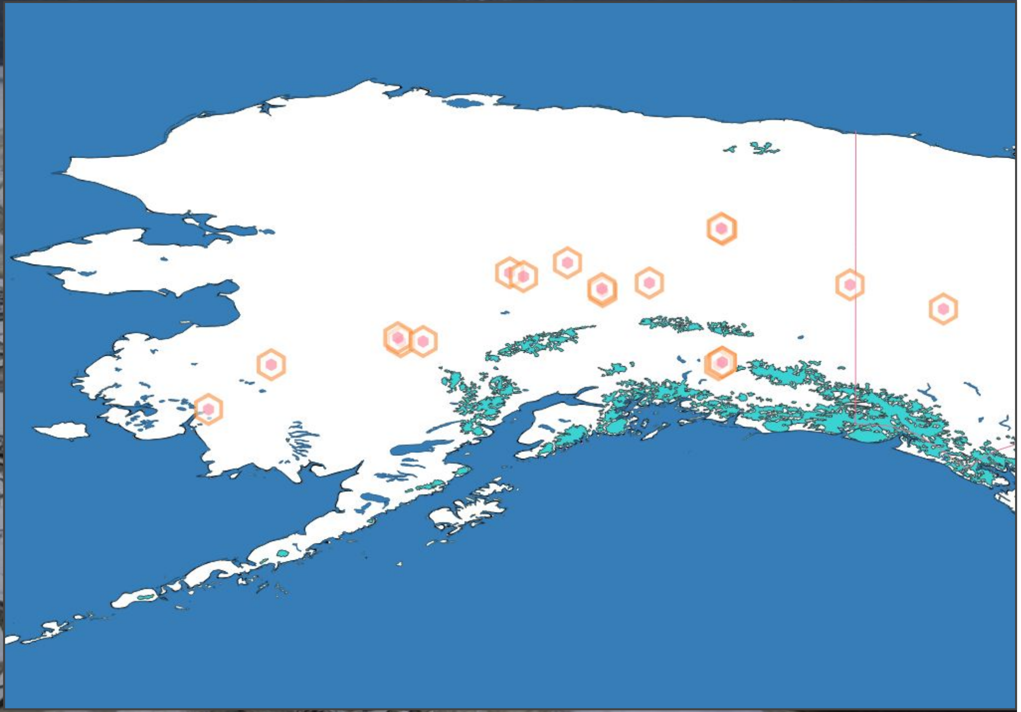
Machine Learning for Understanding Biospheric Processes

- Widening adoption of deep neural networks and growth of Earth science data are fueling interest in AI/ML for use in 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 matter & age**
  - **Wildfire**
  - **Methane fluxes**
  - **Ecohydrology**
- All of these applications involve unresolved, subgrid-scale processes that strongly influence results at the largest scales

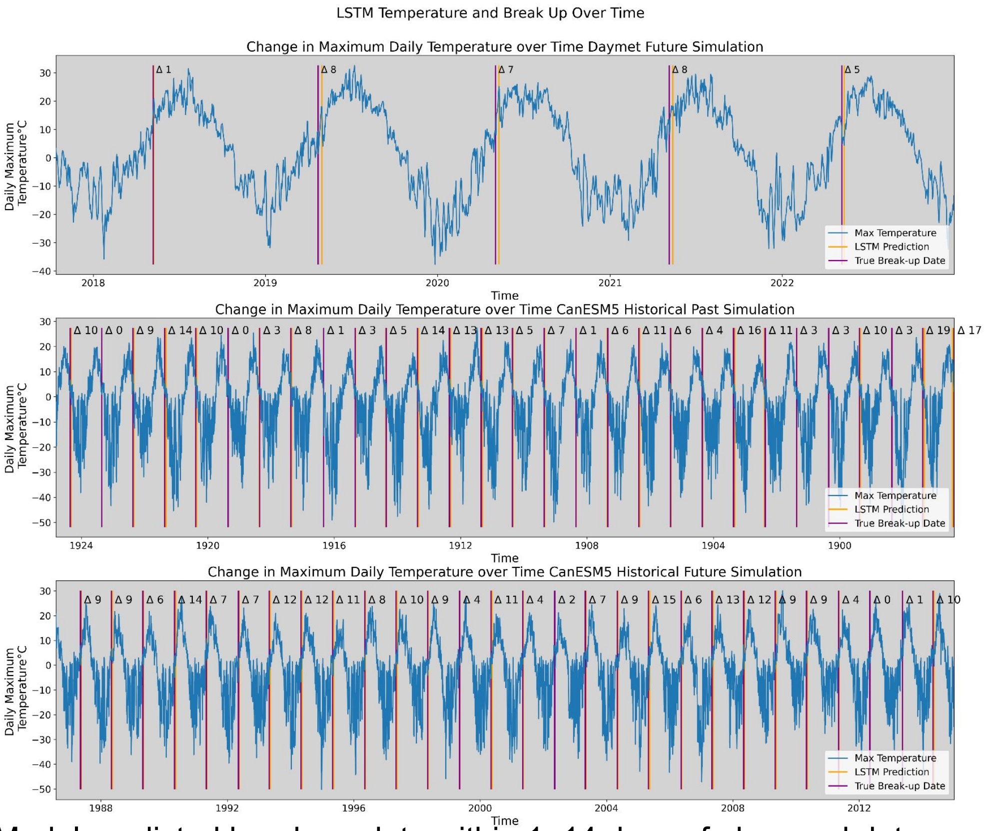


orecasting River Ice Breakup using Machine Learning

- Study sites were selected at long term river ice monitoring stations in Yukon river basin.
- We developed Long Short Term Memory (LSTM) and transformer models to predict river ice breakups.
- Primary predictor variables: daily min/max air temp., precipitation, snow water eq., shortwave radiation
- Datasets: DAYMET, CanESM5 (Historical, SSP119, SSP370, SSP585, SSP534-over)

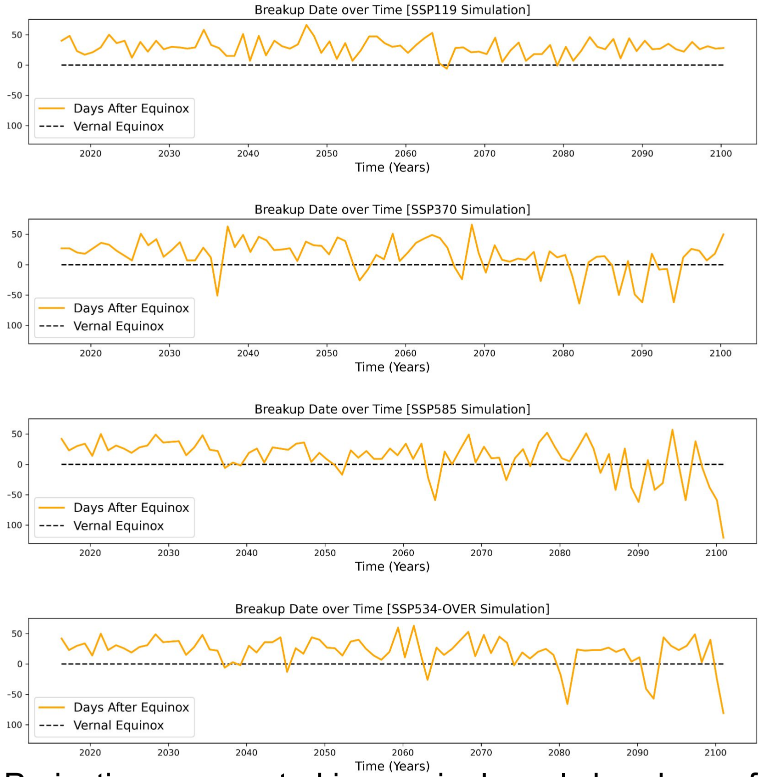


Break-up date predictions for historical period



Model predicted break-up date within 1–14 days of observed dates.

Break-up date predictions under future scenarios



Projections suggested increasingly early break-up of river ice under warming scenarios.

ESS Observations & AI/ML Are Essential for Skillful Predictions

ARTIFICIAL INTELLIGENCE FOR EARTH SYSTEM PREDICTABILITY (AI4ESP): CHALLENGES AND OPPORTUNITIES

FORREST M. HOFFMAN
Oak Ridge National Laboratory

CHARULEKA VARADHARAJAN
HARUKO WAINWRIGHT
Lawrence Berkeley National Laboratory

NICKI L. HICKMON
SCOTT M. COLLIS
Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY

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**

U.S. DEPARTMENT OF ENERGY

Earth System Predictability Sessions

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

Cross-Cut Sessions

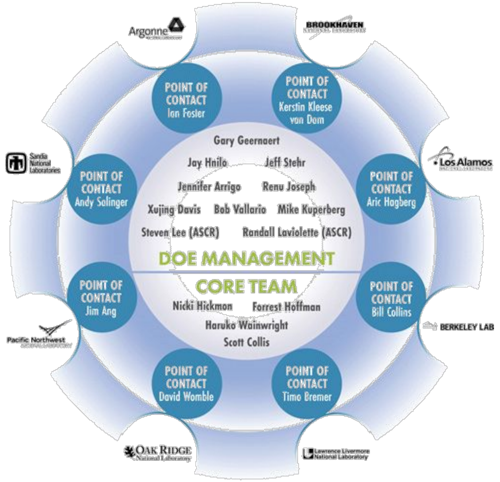
- Data Acquisition
- Neural Networks
- Surrogate models and emulators
- Knowledge-Informed Machine Learning
- Hybrid Modeling
- Explainable/Interpretable/Trustworthy AI
- Knowledge Discovery & Statistical Learning
- AI Architectures and Co-design

Workshop Report

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

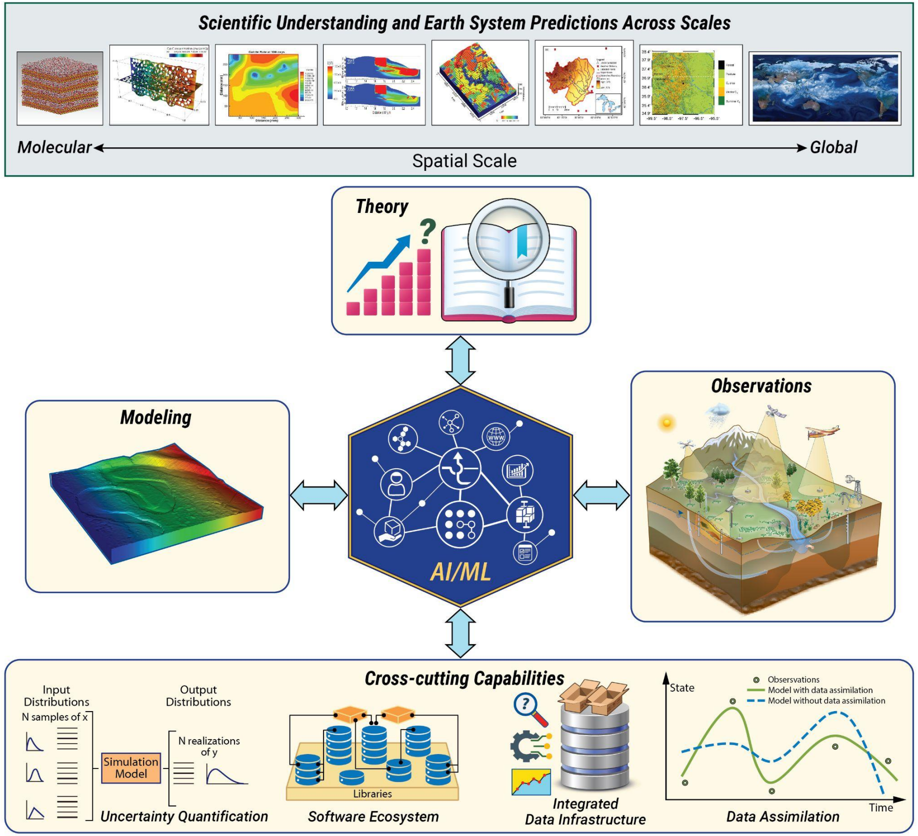
AMS Special Collection

- In thenew AI for the Earth Systems journal



AI4ESP

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

Highlights Across All Sessions

Science

- AI/ML can accelerate next-generation integrated models to support decision-making that incorporate complex natural and human processes at sufficient resolutions
- Broad consensus on need for deep integration of process-based and ML models (hybrid models)
- Challenges: scaling, sub-grid representation, model calibration/UQ, extreme events, human systems
- Data gaps are vast – more observations informed by model needs, AI-ready products
- Results must be robust, explainable, & trustworthy

Data, Software, Infrastructure

- Need benchmark data and model intercomparison approaches
- Computational infrastructure for integration of process & ML models, data assimilation and synthesis
- Use ML to accelerate data-model and model-observation pipelines

Culture

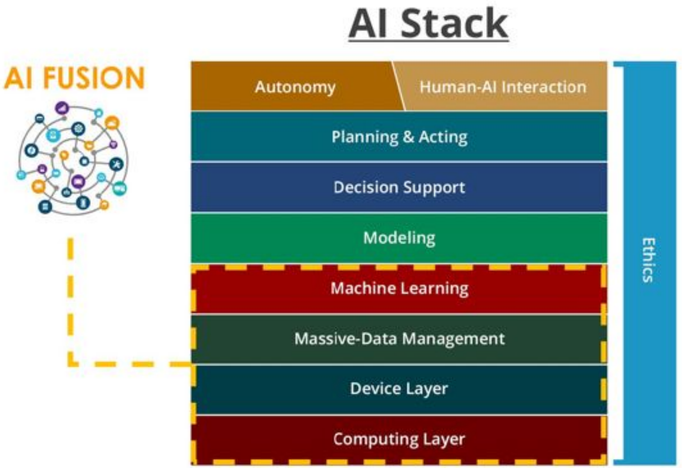
- Workforce development across domain and computational scientists
- Interdisciplinary research centers focused on AI4ESP

AI4ESP WORKSHOP HIGHLIGHTS

Codesign Is Critical

Codesign advanced computing, software, hybrid ML/physical models, observations and future Earth system modeling capabilities

- Common/consistent language & format
- Merged products (standardization, interoperability)
- Adaptive data & parameter selection
- Computation using large datasets without moving
- Specialized AI/ML code & architecture
- Training and benchmarking datasets and hybrid model design



College of Engineering, Carnegie Mellon University

AI4ESP WORKSHOP HIGHLIGHTS

Infrastructure Investment Is Imperative

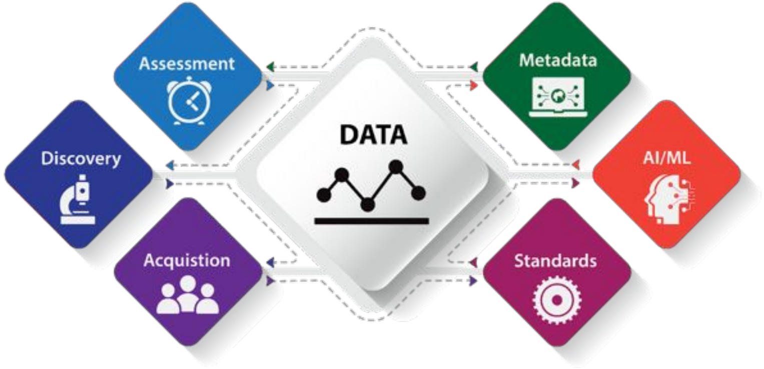
- Workforce development
- Multi-agency/institution coordination, cooperation, collaboration
- Codesign, creation, implementation & maintenance
  - Computational resources
  - Training, benchmarking, & combined datasets
  - AI methodology development
  - Interoperable frameworks for data & hybrid modeling
- FAIR/Equitable data & software practices
- Observations covering normal & capturing rare & extreme events
- Adaptive observatories, data assimilation, & modeling

image from technologynetworks.com

AI4ESP WORKSHOP HIGHLIGHTS

Cultural Change Is Compulsory

- Communities excited to work together
  – need combined purpose and early success
- Existing & upcoming workforce development
- Common terminology across groups & scales in AI4ESP space
- Transfer learning for different domains & scales
- Achieve & maintain FAIR, equitable data access
- Open science community effort pulling in an ultimately singular direction
- Environmental justice throughout the system

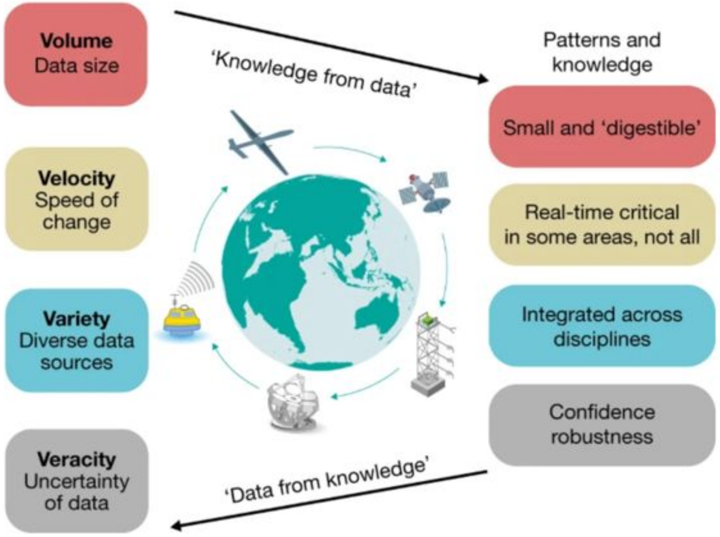


Modular Data Ecosystem to enable data interoperability for AI. Courtesy of Prakash & Serbin

AI4ESP WORKSHOP HIGHLIGHTS

Uncertainty Quantification & Propagation Is Underlying

- Digital twin mindset
- Common understanding of uncertainty
- Defined uncertainty
- Capture beginning with instrument/sensor calibration/operation
- Propagation requires formatting and transfer standards
- Assimilation, parameterization, surrogate, emulator, hybrid modeling

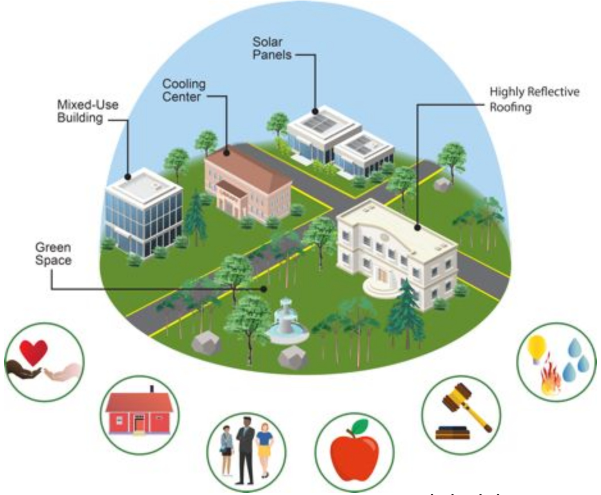


Data challenges in the earth sciences: different data sources, small data / big data challenges, and uncertainty in the data. Figure taken from (Reichstein, M. et al. 2019)

AI4ESP WORKSHOP HIGHLIGHTS

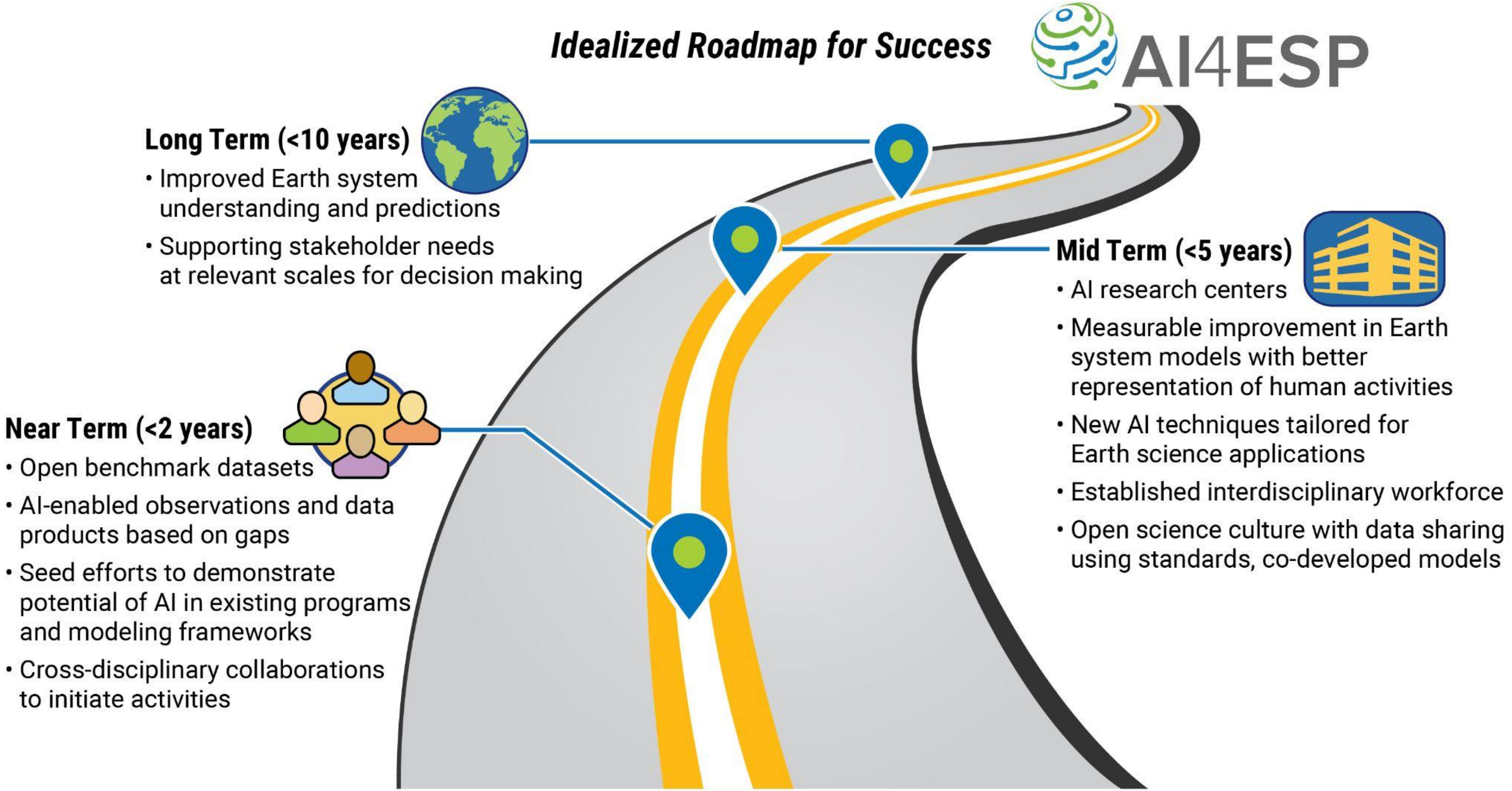
Human System Integration Is Significant

- Inclusion of complex human processes & decisions
- Capture complex feedbacks between all components
- Build decision-relevant process models
- Ethically sensitive data synthesis and gap filling
- Representation of human systems and dynamics in models
- Results must be robust, explainable, & trustworthy
- Results must be shared efficiently (both positive & negative)



globalchange.gov

AI4ESP WORKSHOP HIGHLIGHTS



AI4ESS Breakout #3: AI for ESS Science

AI/ML for ESS Data-Model Integration to Advance Predictive Process Understanding

Question 1: [30 min]

What immediate (0.5–1 year) and intermediate (1–2 year) opportunities exist or could be created to leverage AI to meet the ESS mission, and what is needed to take action?

Question 2: [30 min]

How can AI help ESS contribute to other BER programs (both divisions), other SC offices (e.g., ASCR), and the DOE mission (e.g., secretary 9 priorities https://www.energy.gov/articles/secretary-wright-acts-unleash-golden-era-american-energy-dominance)?