



INV14A-02 - Artificial Intelligence for Exploring Climate Change Mitigation Strategies and Advancing Earth System Prediction

Forrest M. Hoffman^{1,2}, Jitendra Kumar¹, Zachary L. Langford¹, V. Shashank Konduri³, Auroop R. Ganguly⁴, Cheng-En Yang², Nathan Collier¹, Min Xu¹, William W. Hargrove⁵, Nicki L. Hickmon⁶, Scott M. Collis⁶, Charuleka Varadharajan⁷, and Haruko Wainwright⁷

December 12, 2022

¹Oak Ridge National Laboratory, Oak Ridge, TN, USA

²University of Tennessee, Knoxville, TN, USA

³National Ecological Observatory Network, Boulder, CO, USA

⁴Northeastern University, Boston, MA, USA

⁵US Department of Agriculture - Forest Service, Asheville, NC, USA

⁶Argonne National Laboratory, Lemont, IL, USA

⁷Lawrence Berkeley National Laboratory, Berkeley, CA, USA



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](#) List and the first supercomputer to break the exaflop barrier (May 30, 2022).

*This article is the second in a two-part series.
The first part, "How to Build a Hypercomputer," by
Thomas Sterling, appeared in the July 2001 issue.*

The Do-It-Yourself Supercomputer

By William W. Hargrove,
Forrest M. Hoffman and
Thomas Sterling

Photographs by Kay Chernush

Scientists have
found a cheaper
way to solve
tremendously
difficult
computational
problems:
connect ordinary
PCs so that they
can work together

CLUSTER OF PCs at the
Oak Ridge National
Laboratory in Tennessee
has been dubbed the
Stone SouperComputer.

Hargrove, W. W., F. M. Hoffman, and T. Sterling (2001), The
Do-It-Yourself Supercomputer, *Sci. Am.*, 265(2):72-79,
<https://www.scientificamerican.com/article/the-do-it-yourself-superpc/>

Multivariate Geographic Clustering

- Ecoregions have traditionally been created by experts
- Our approach has been to objectively create ecoregions using continuous continental-scale data and clustering
- We developed a highly scalable *k*-means cluster analysis code that uses distributed memory parallelism
- Originally developed on a 486/Pentium cluster, the code now runs on the largest hybrid CPU/GPU architectures on Earth

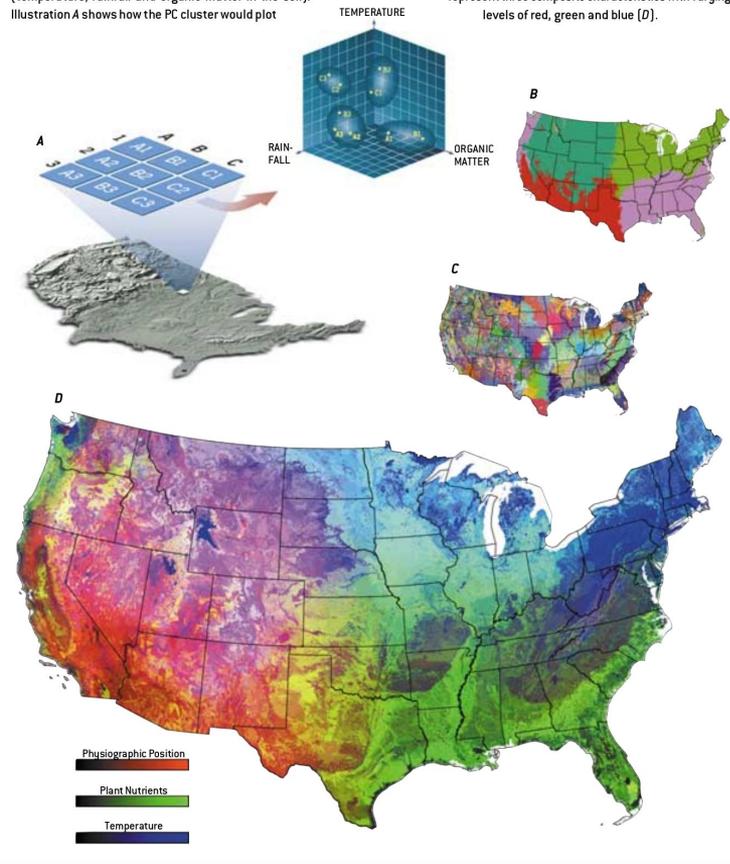
Hargrove, W. W., F. M. Hoffman, and T. Sterling (2001), The Do-It-Yourself Supercomputer, *Sci. Am.*, 265(2):72–79,

<https://www.scientificamerican.com/article/the-do-it-yourself-superc/>

MAKING MAPS WITH THE STONE SOUPERCOMPUTER

TO DRAW A MAP of the ecoregions in the continental U.S., the Stone SouperComputer compared 25 environmental characteristics of 7.8 million one-square-kilometer cells. As a simple example, consider the classification of nine cells based on only three characteristics (temperature, rainfall and organic matter in the soil). Illustration A shows how the PC cluster would plot

the cells in a three-dimensional data space and group them into four ecoregions. The four-region map divides the U.S. into recognizable zones (Illustration B); a map dividing the country into 1,000 ecoregions provides far more detail (C). Another approach is to represent three composite characteristics with varying levels of red, green and blue (D).



New Analysis Reveals Representativeness of the AmeriFlux Network

PAGES 529, 535

The AmeriFlux network of eddy flux covariance towers was established to quantify variation in carbon dioxide and water vapor exchange between terrestrial ecosystems and the atmosphere, and to understand the underlying mechanisms responsible for observed fluxes and carbon pools. The network is primarily funded by the U.S. Department of Energy, NASA, the National Oceanic and Atmospheric Administration, and the National Science Foundation. Similar regional networks elsewhere in the world—for example, CarboEurope, AsiaFlux, OzFlux, and Fluxnet Canada—participate in

synthesis activities across larger geographic areas [Baldocchi et al., 2001; Law et al., 2002]. The existing AmeriFlux network will also form a backbone of “Tier 4” intensive measurement sites as one component of a four-tiered carbon observation network within the North American Carbon Program (NACP). The NACP seeks to provide long-term, mechanistically detailed, spatially resolved carbon fluxes across North America [Wolry and Harris, 2002]. For both of these roles, the AmeriFlux network should be ecologically representative of the environments contained within the geographic boundaries of the program. A new ecoregion-scale analysis of the existing AmeriFlux network reveals that, while central continental environments are well-represented, additional flux towers are needed to represent environmental

BY WILLIAM W. HARGROVE, FORREST M. HOFFMAN, AND BEVERLY E. LAW

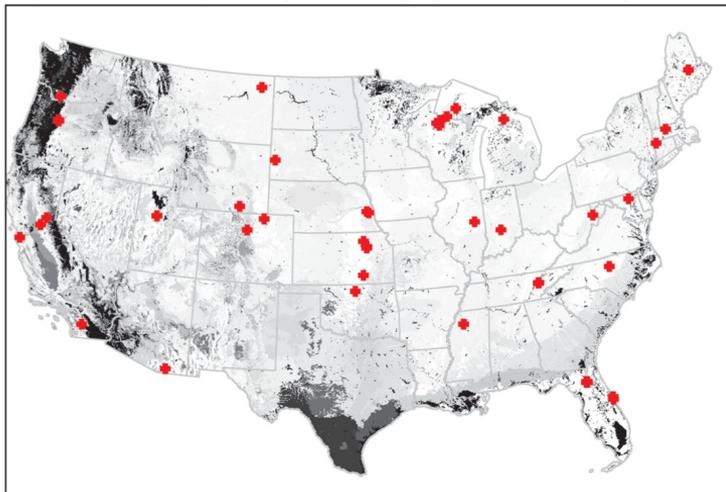


Fig. 1. The representativeness of an existing spatial array of sample locations or study sites—for example, the AmeriFlux network of carbon dioxide eddy flux covariance towers—can be mapped relative to a set of quantitative ecoregions, suggesting locations for additional samples or sites. Distance in data space to the closest ecoregion containing a site quantifies how well an existing network represents each ecoregion in the map. Environments in darker ecoregions are poorly represented by this network.

Network Representativeness

- The n -dimensional space formed by the data layers offers a natural framework for estimating representativeness of individual sampling sites
- The Euclidean distance between individual sites in data space is a metric of similarity or dissimilarity
- Representativeness across multiple sampling sites can be combined to produce a map of network representativeness

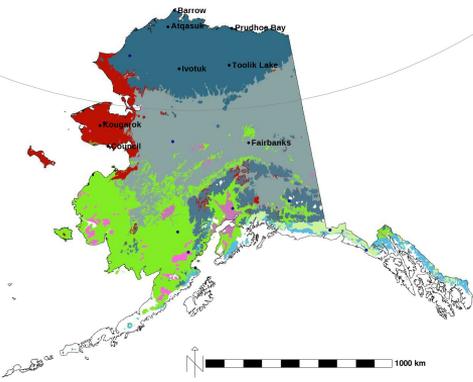
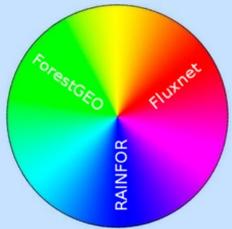
Hargrove, W. W., and F. M. Hoffman (2003), New Analysis Reveals Representativeness of the AmeriFlux Network, *Eos Trans. AGU*, 84(48):529, 535, doi:[10.1029/2003EO480001](https://doi.org/10.1029/2003EO480001).

Sampling Network Design

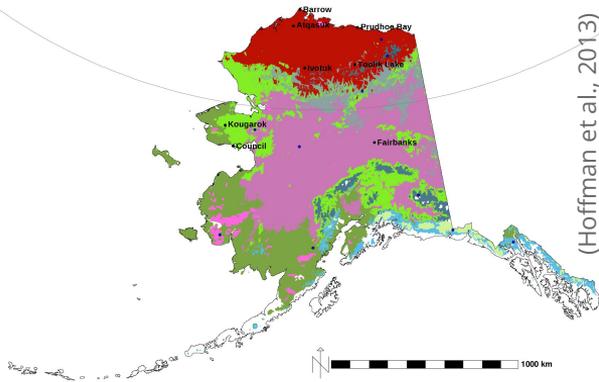


NSF's NEON Sampling Domains

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

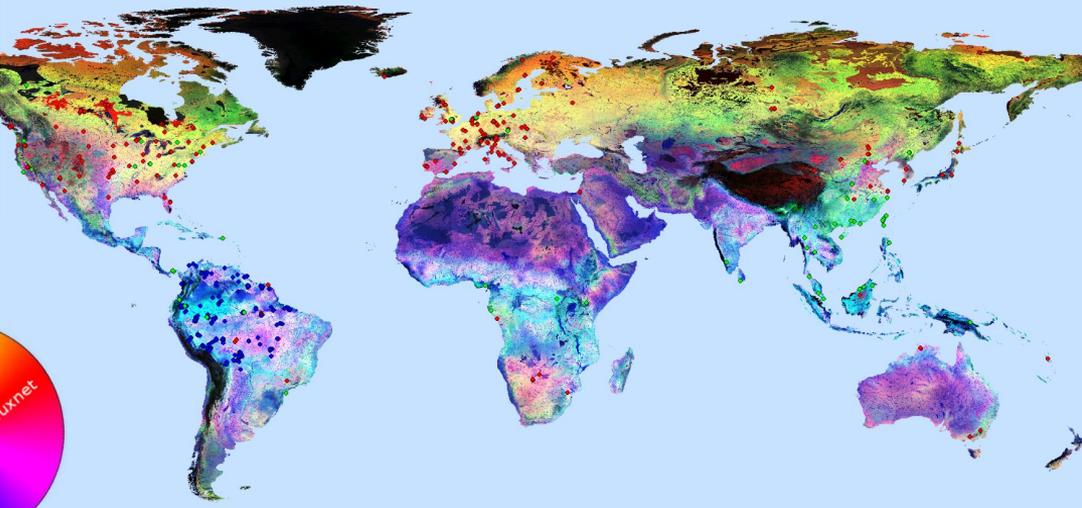


2000-2009



2090-2000

Triple-Network Global Representativeness



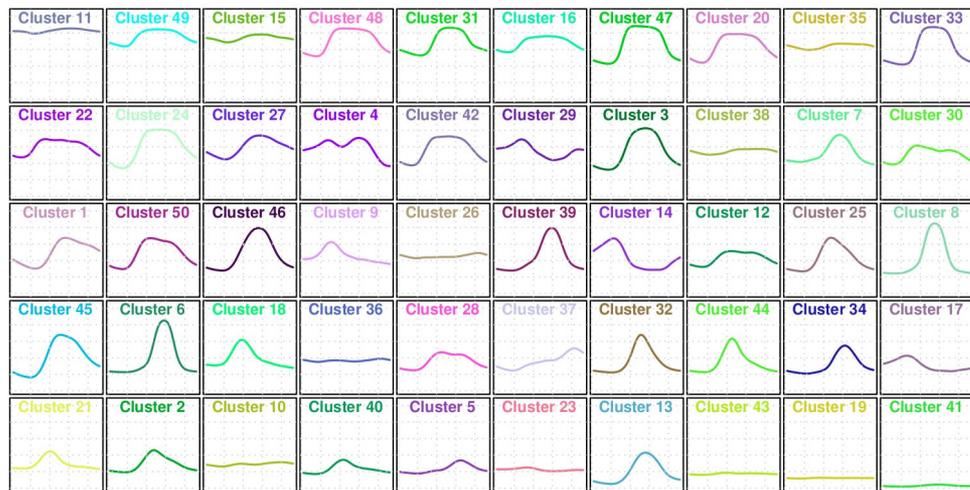
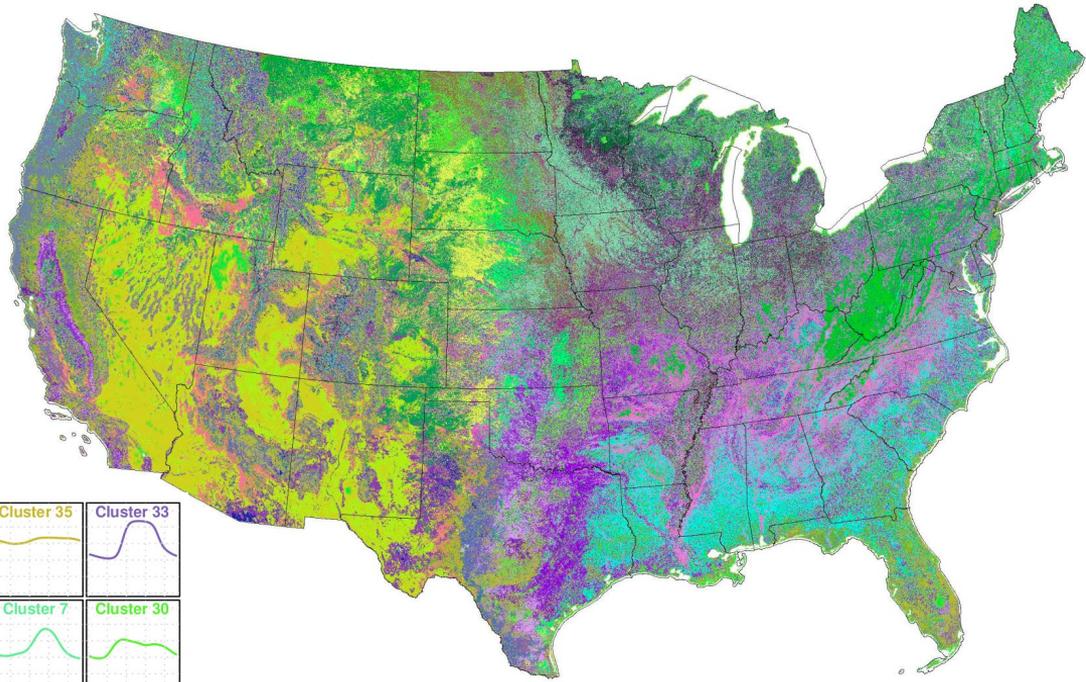
(Maddalena et al., in prep.)

50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI

Every 8 days (46 images/year)

Clustered from year 2000 to present

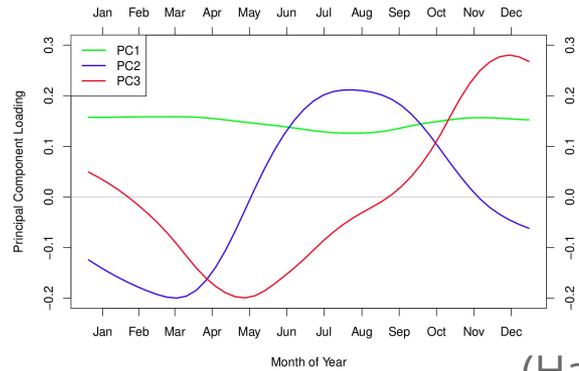
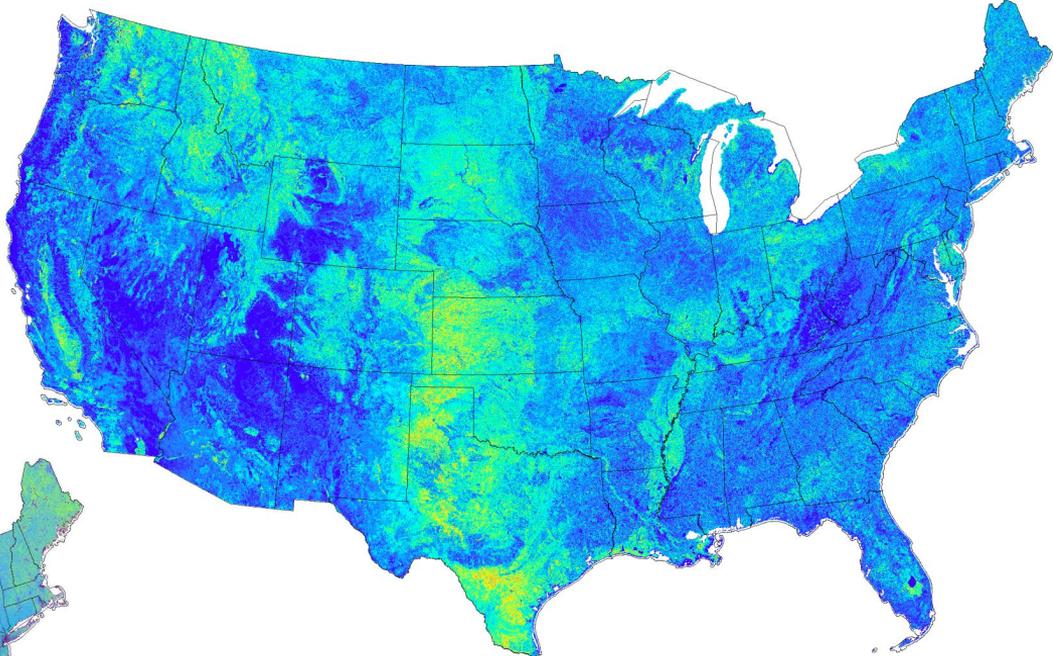
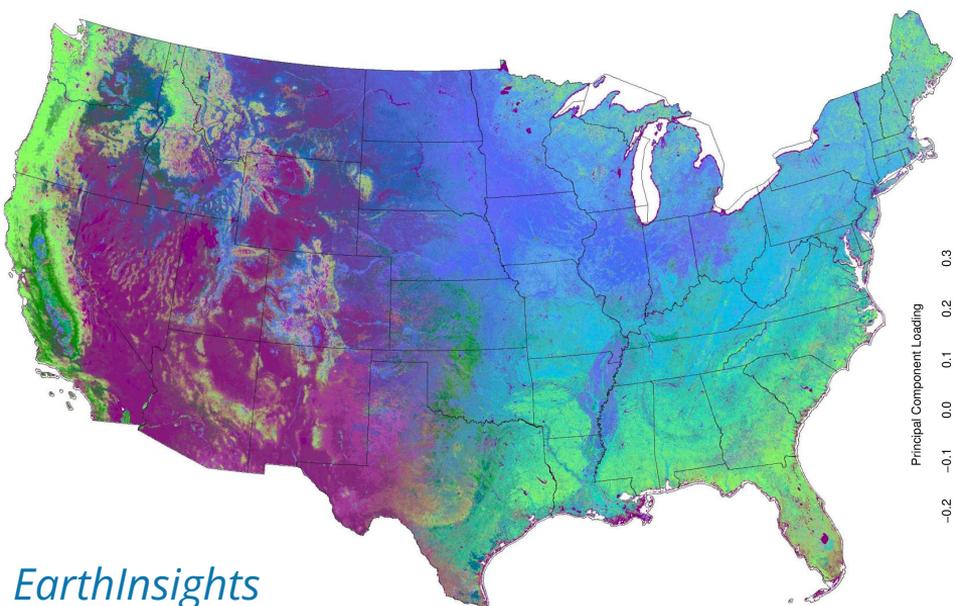


50 Phenoregion Prototypes (Random Colors)

NDVI

day of year

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

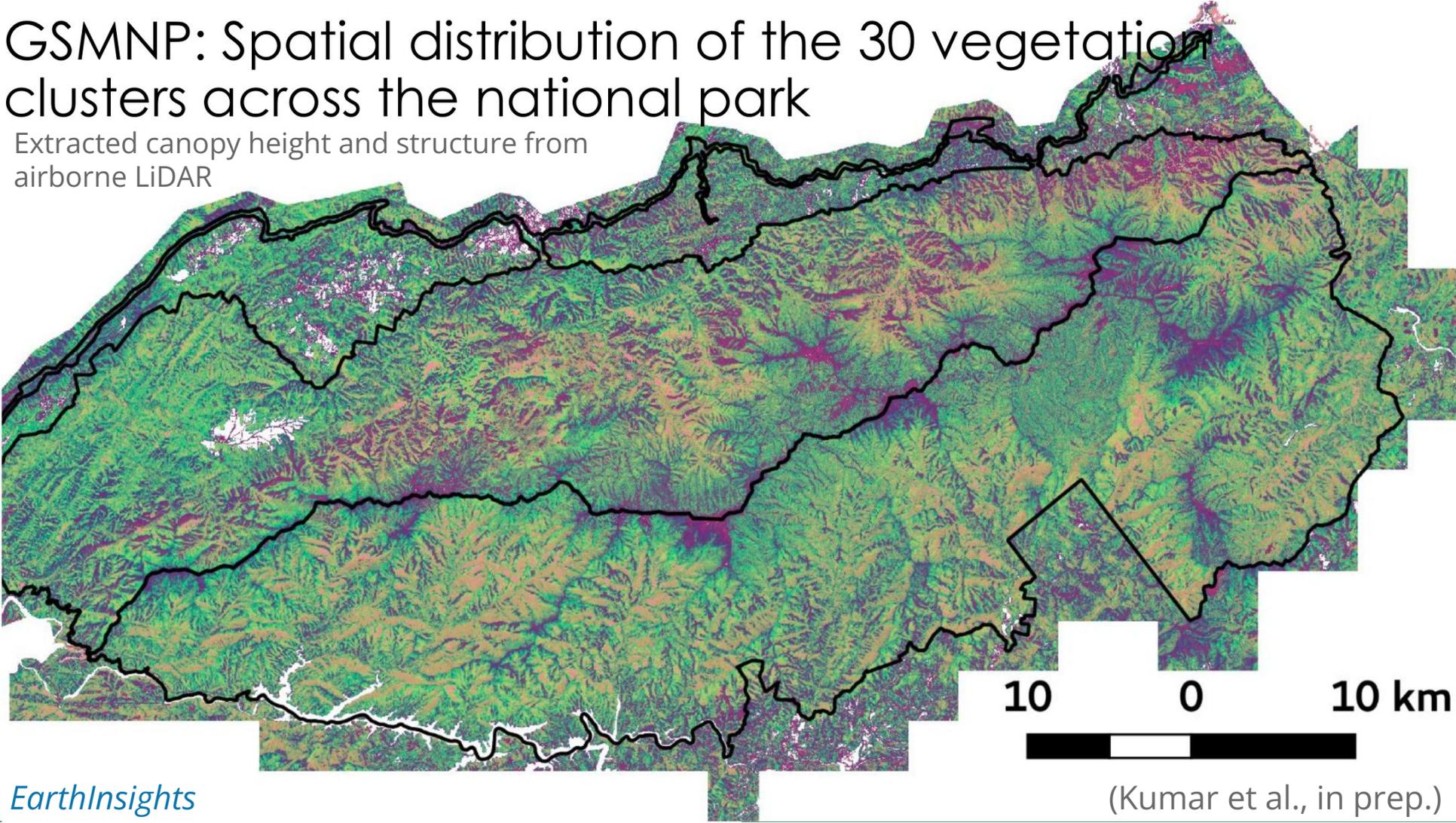


Principal Components Analysis

- PC1 ~ Evergreen
- PC2 ~ Deciduous
- PC3 ~ Dry Deciduous

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

Extracted canopy height and structure from
airborne LiDAR

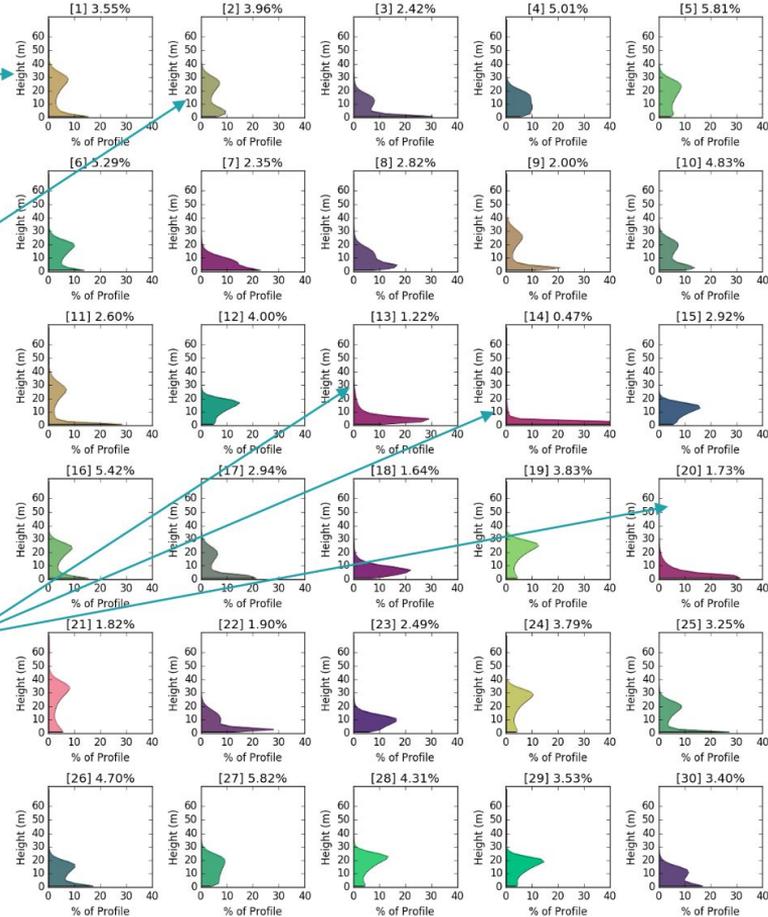


GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

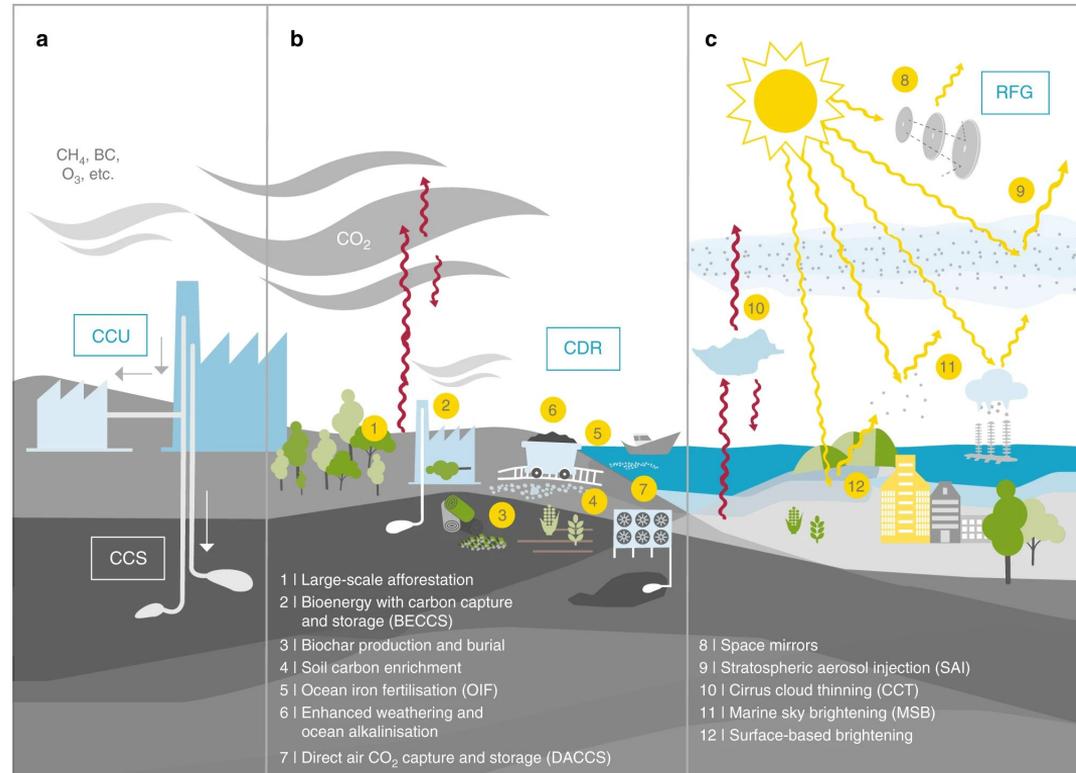
forests with slightly lower mean height with dense understory vegetation

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



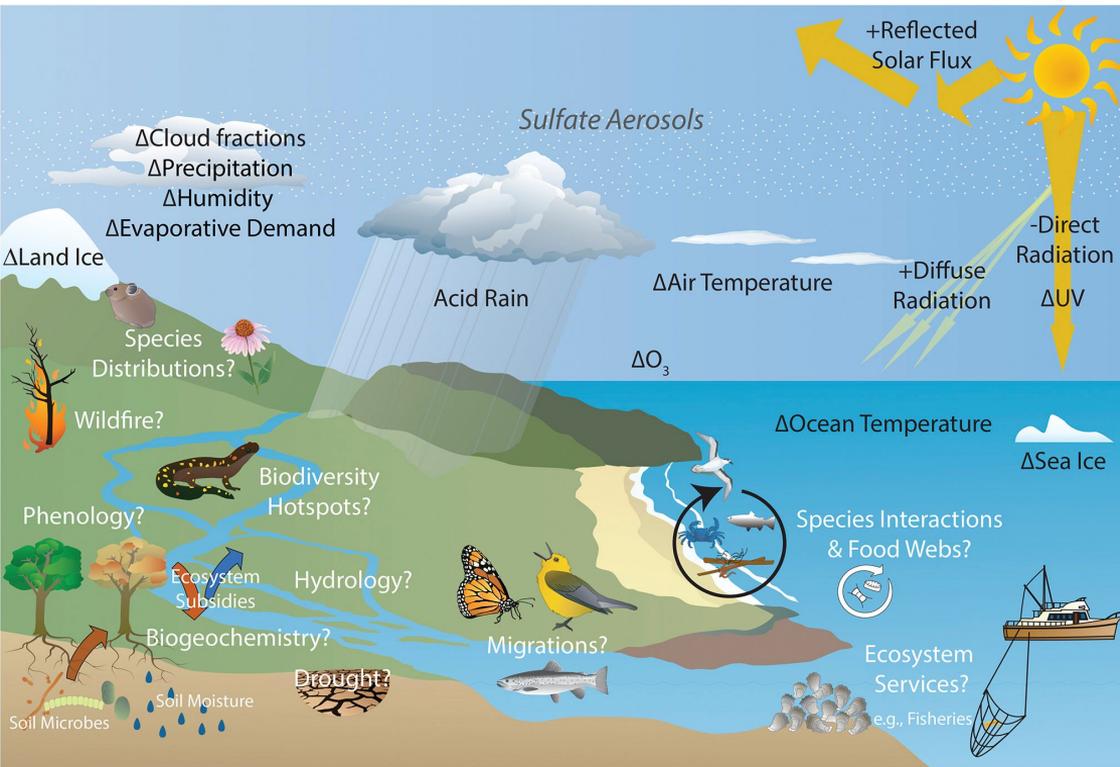
Climate Change Mitigation through Climate Intervention

- 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



- 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

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).

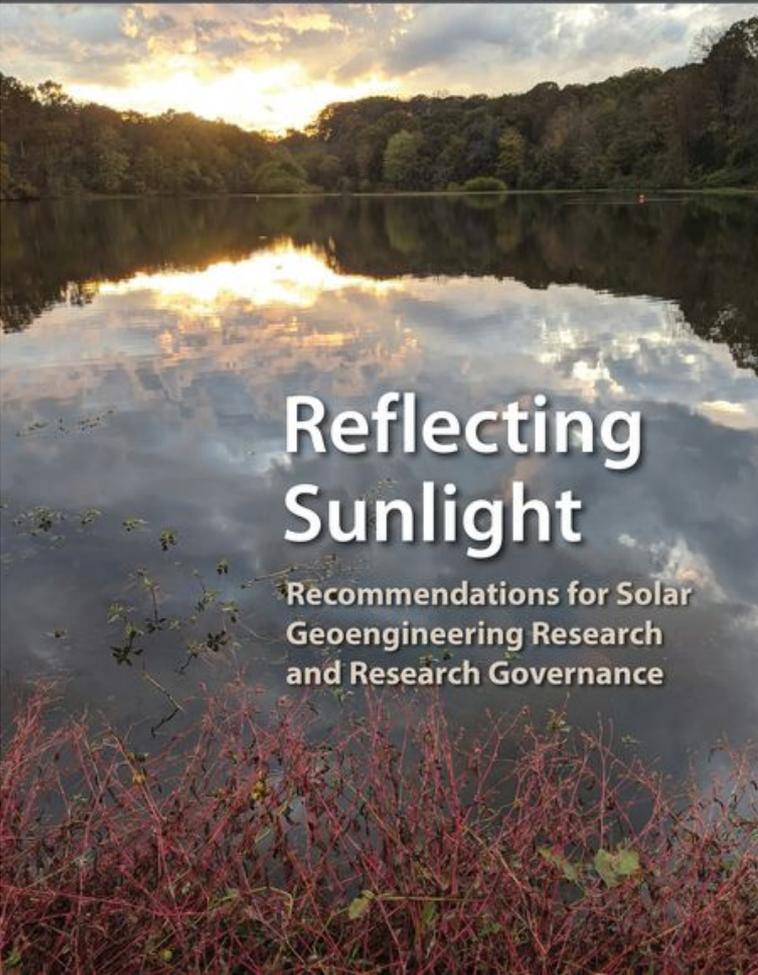
Climate Intervention Research

A 2021 report from the National Academies of Sciences, Engineering, and Medicine (NASEM) concludes **a strategic investment in research is needed** to advance policymakers' understanding of climate response options.

The US should develop a transdisciplinary research program, in collaboration with other nations, to advance understanding of solar geoengineering's technical feasibility and effectiveness, possible impacts on society and the environment, and social dimensions such as public perceptions, political and economic dynamics, and ethical and equity considerations.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

CONSENSUS STUDY REPORT



Reflecting Sunlight

Recommendations for Solar
Geoengineering Research
and Research Governance

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 CO₂ 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](https://doi.org/10.1088/1748-9326/abacf7).

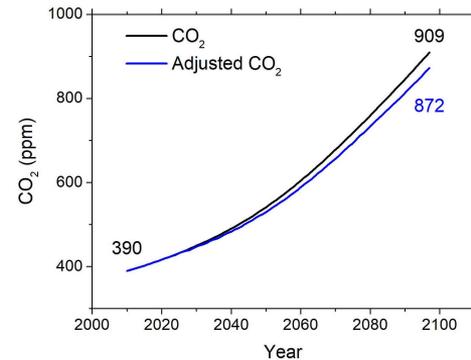
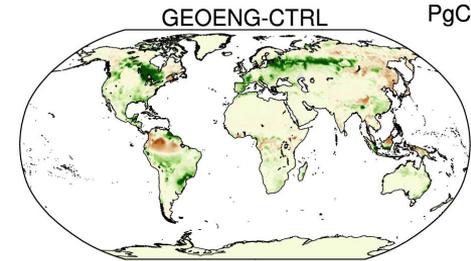
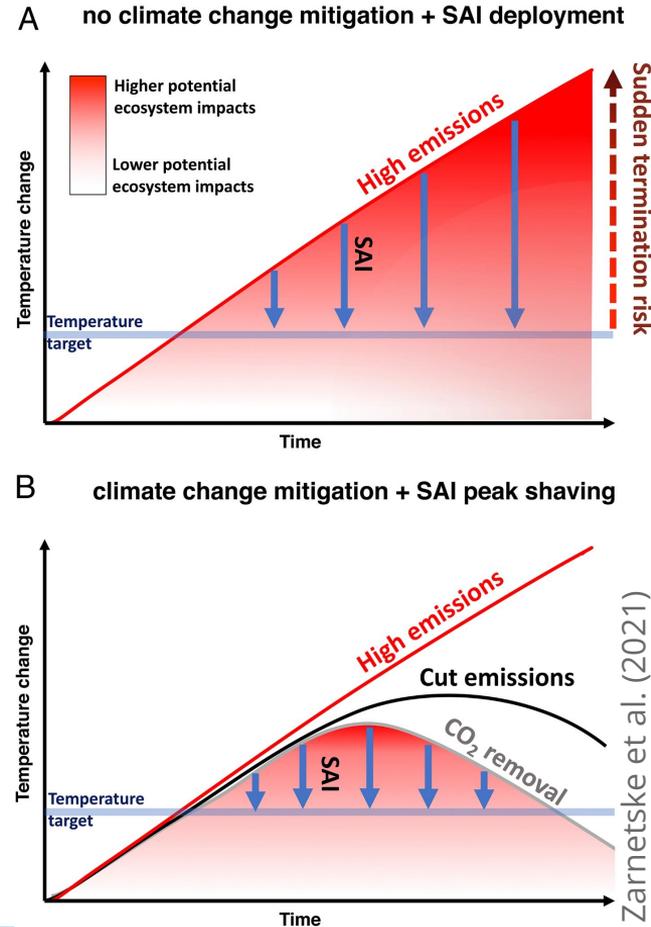


Figure: The larger sink under SAI increased land C storage by 79 Pg C by 2097, which would reduce the projected atmospheric CO₂ level.

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 CO₂ trajectory from CMIP6, which prescribes a drawdown of CO₂
- 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 CO₂, **but may not capture all the as yet unobserved processes**



Leveraging Advances in Machine Learning for Earth Sciences

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

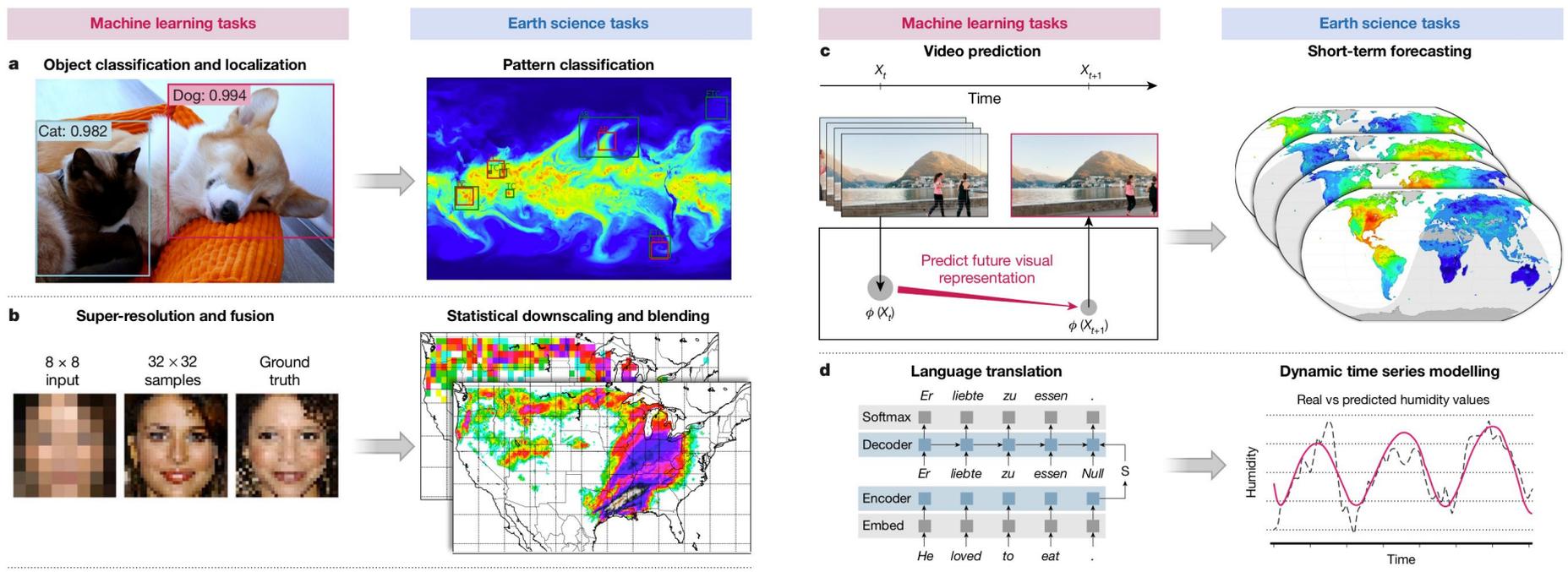
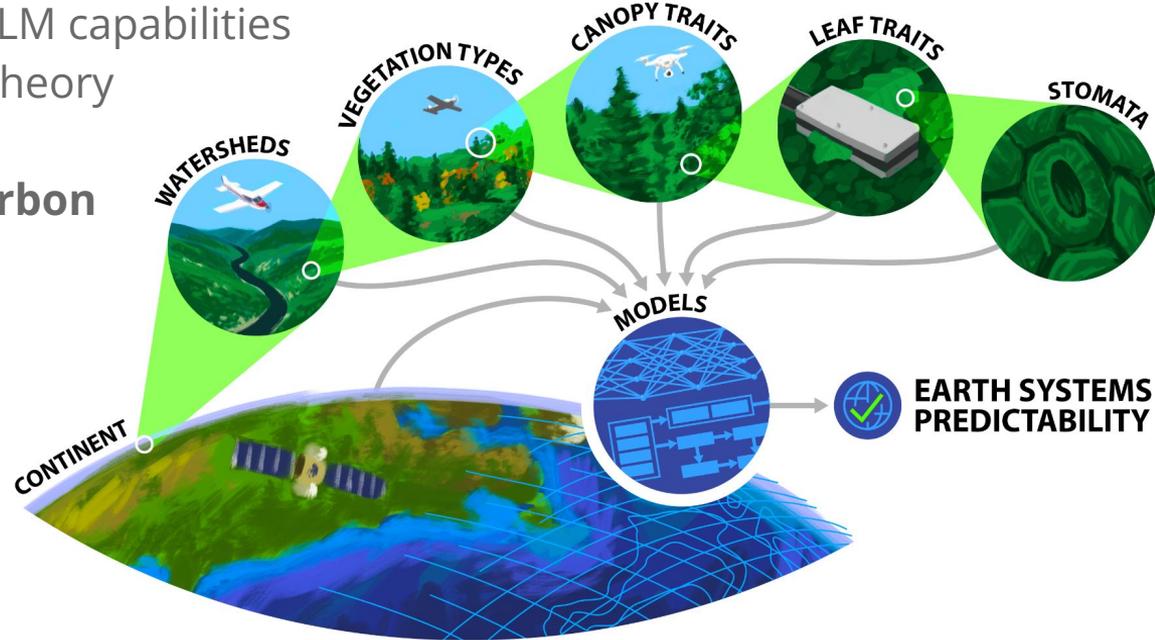


Figure 2 in Reichstein et al. (2019)

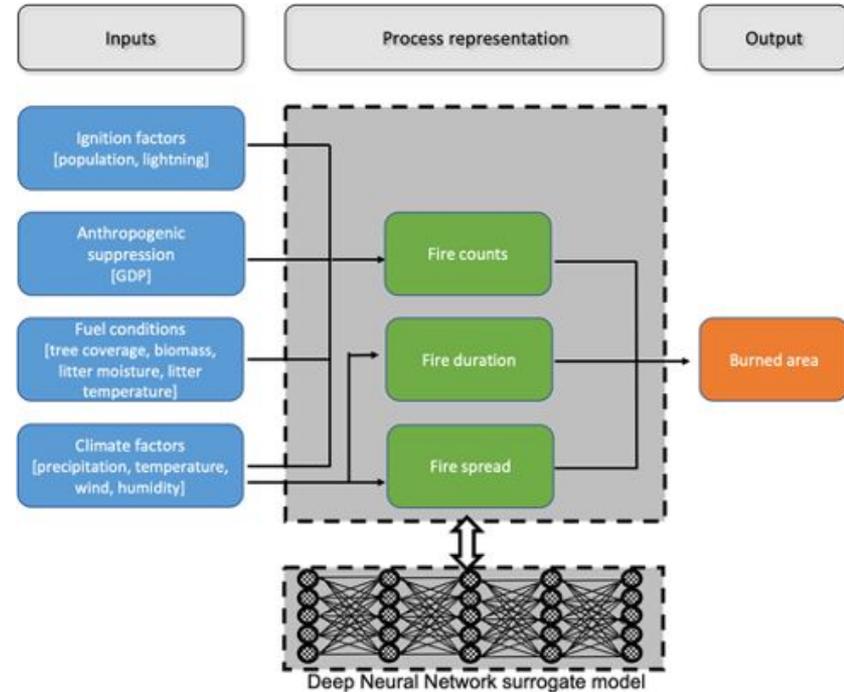
Machine Learning for Understanding Biospheric Processes

- 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



Hybrid Modeling of Wildfire Activities

- Improve model simulations of **wildfire processes**, including ignition, fire duration, and spread rate with Deep Neural Network models
- Improve simulated **wildfire emissions** and their impacts on atmospheric properties, including aerosols, greenhouse gases, phosphorus transport, and pollutants
- Improve the projection of near-future and long-term dynamics of wildfire activities
- Accelerate E3SM coupled land-atmosphere modeling activities for wildfire research
- Explore online ML training/validation strategy for E3SM coupled model simulations

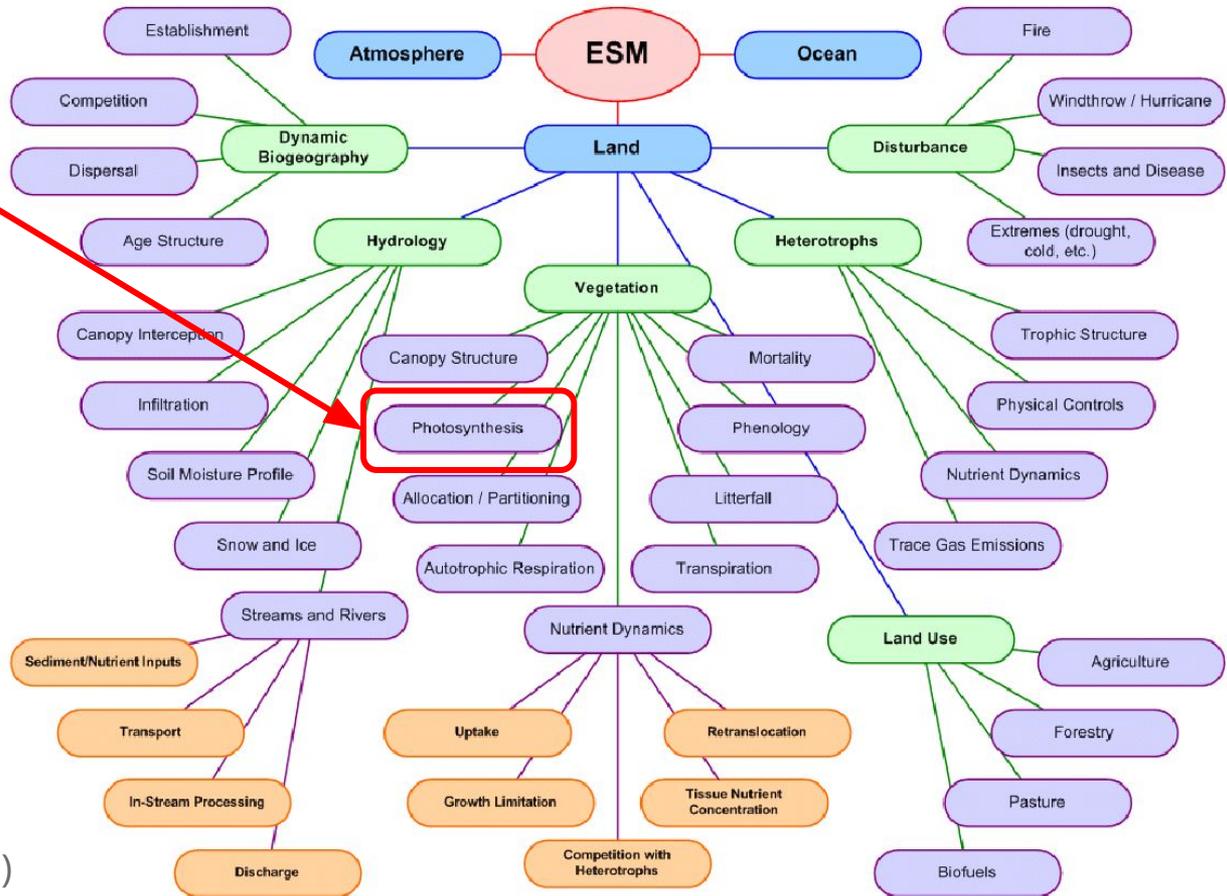


Zhu et al. (2022)

Hybrid ML/Process-based Modeling for Terrestrial Modeling

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

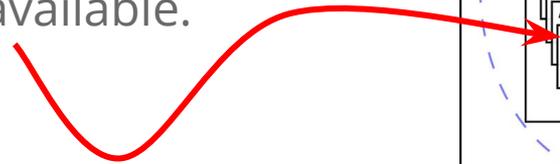
- Multiple hypotheses
- Many leaf-level measurements
- Most computationally intensive part of the land model



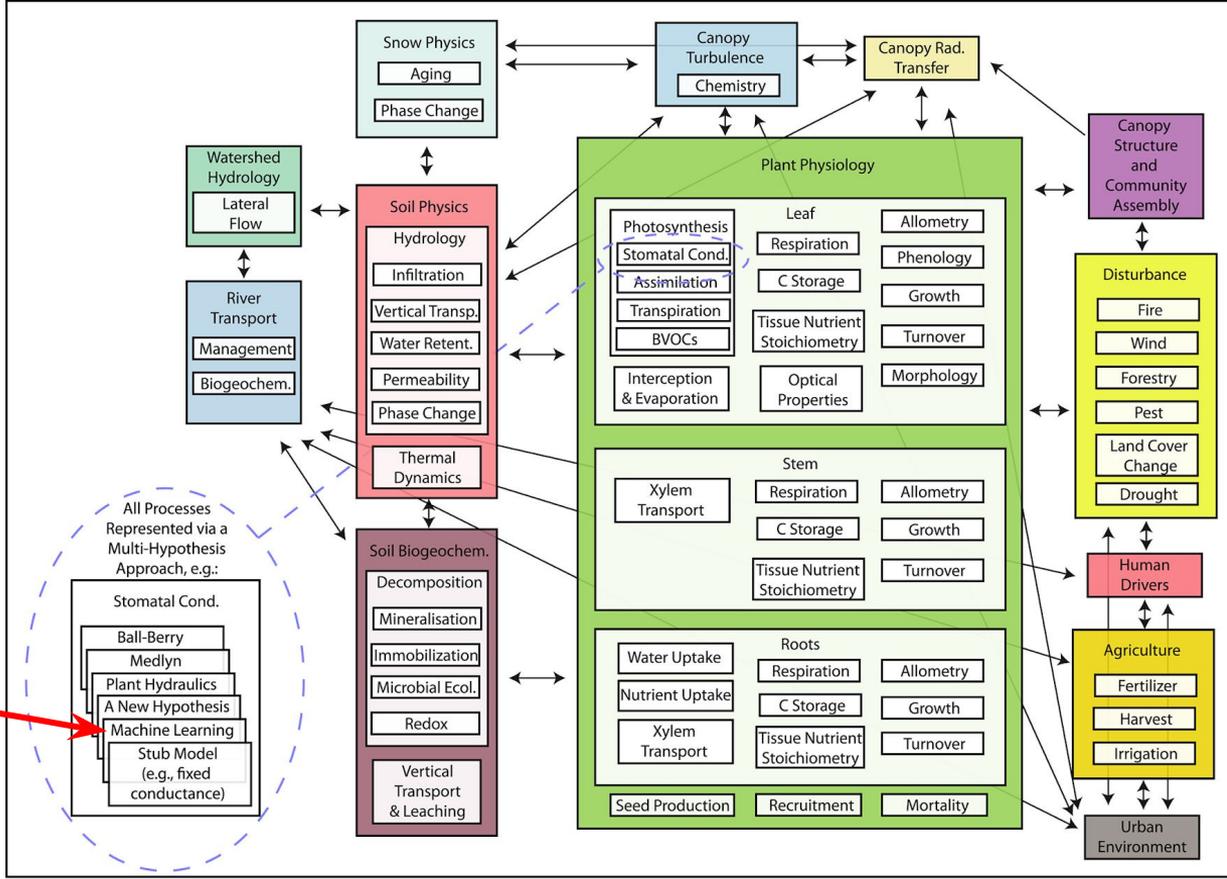
(Figure from P. E. Thornton)

Hybrid ML/Process-based Modeling for Terrestrial Modeling

Individual processes can be represented in a multi-hypothesis approach, and ML provides an opportunities for (1) a model surrogate module or (2) a data-derived module that can be further explored or used to calibrate other hypotheses, when sufficient data are available.



(Fisher and Koven, 2020)

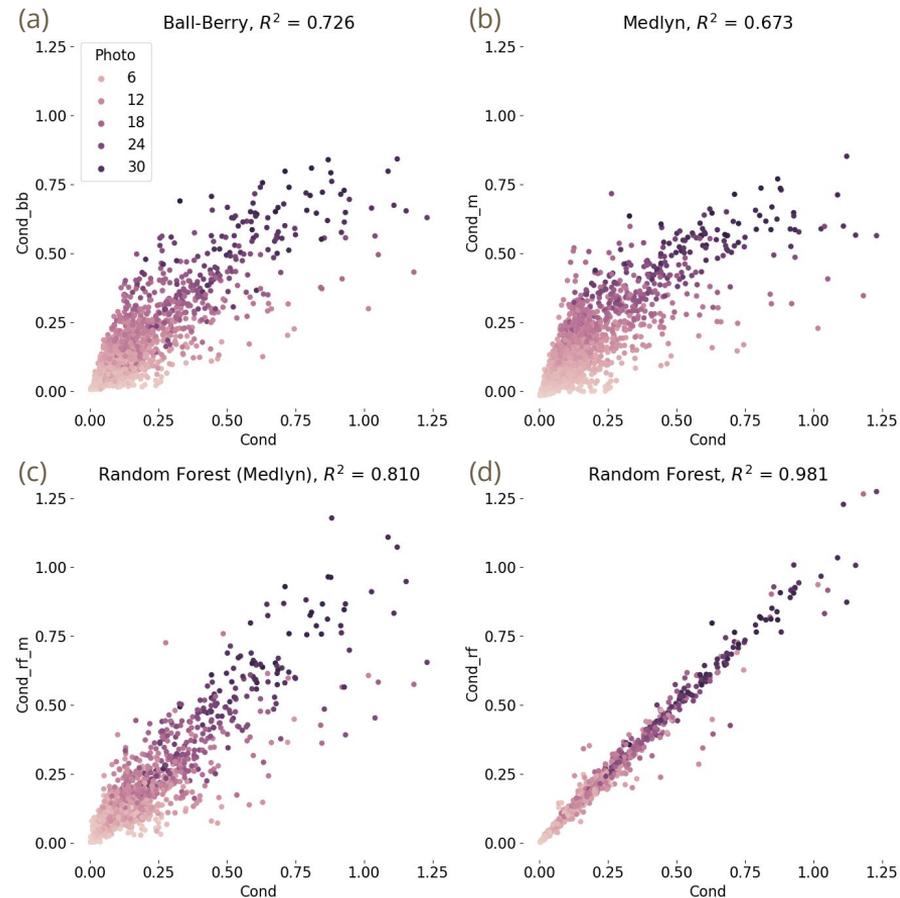


(a) Process Schematic of a Possible Full-Complexity Configuration of a Land Surface Model

Hybrid Modeling of Photosynthesis and Ecohydrology

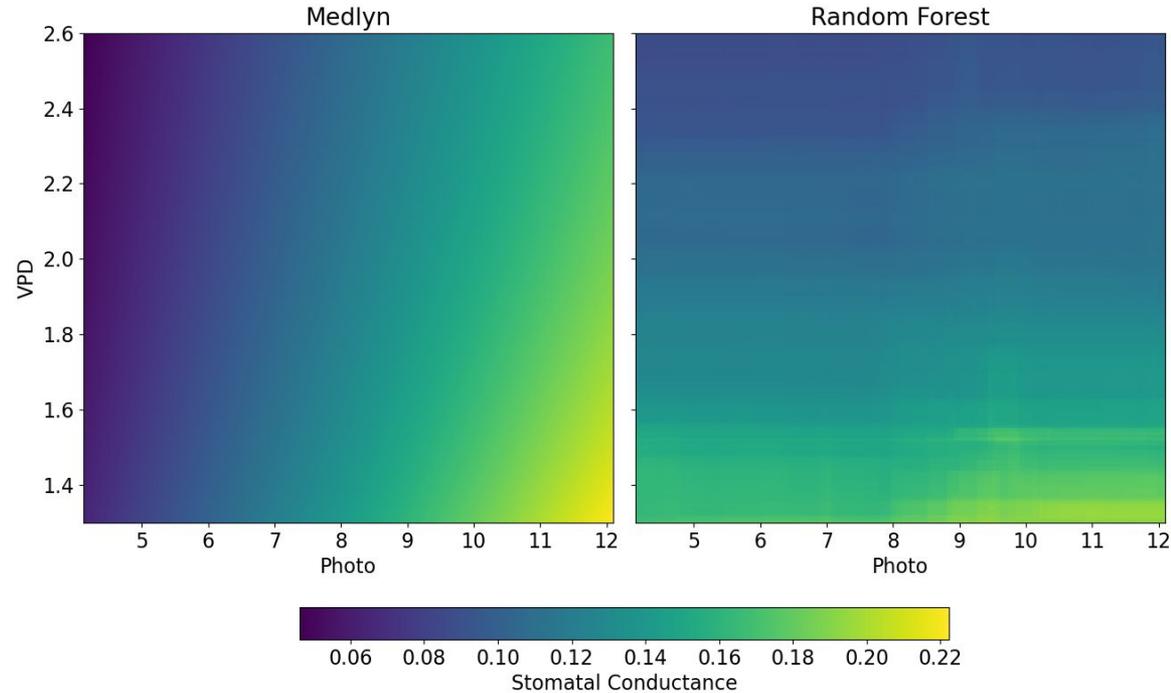
- Significant leaf-level data may be used to train ML parameterizations to **improve accuracy** and **computational performance**
- **Estimated stomatal conductance** vs. measured stomatal conductance for (a) Ball-Berry, (b) Medlyn, (c) Random forest (with Medlyn inputs), and (d) Random forest with all 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)



Hybrid Modeling of Photosynthesis and Ecohydrology

- Most process-based or empirical formulations are continuous
- But ML formulations may exhibit discontinuities in the multi-dimensional space of inputs because of out-of-sample data or artifacts of sampling or precision
- For example, we can see such discontinuities at right for Random Forest in the VPD vs. photosynthesis heat map for stomatal conductance
- These discontinuities are likely to have numerical consequences when attempting to couple a ML parameterization into a hybrid empirical / ML Earth system model



(Massoud, Collier, et al. in prep)



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



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

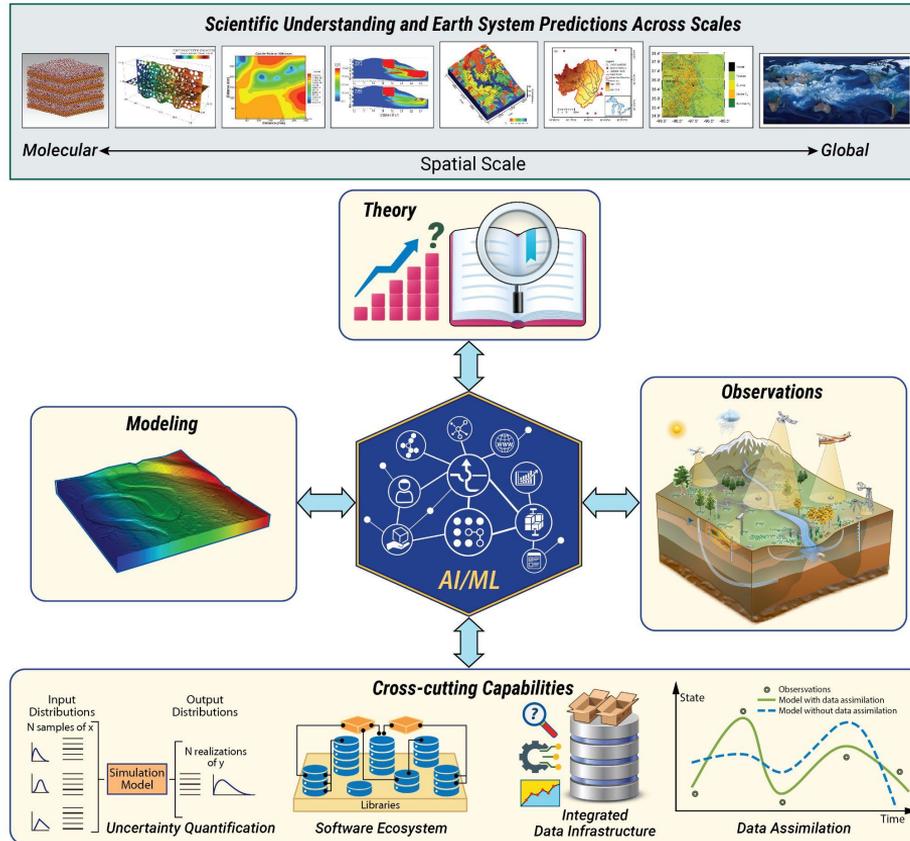
- Posted on ai4esp.org
- Executive Summary
- Long summary
- Earth science chapters
- Computational science chapters

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

AI4ESP WORKSHOP HIGHLIGHTS

Idealized Roadmap for Success



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

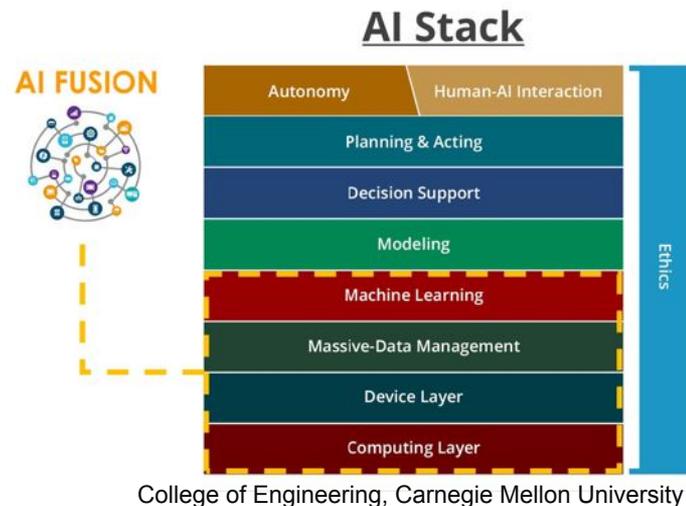
- Workforce development across domain and computational scientists
- Interdisciplinary research centers focused on AI4ESP

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



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

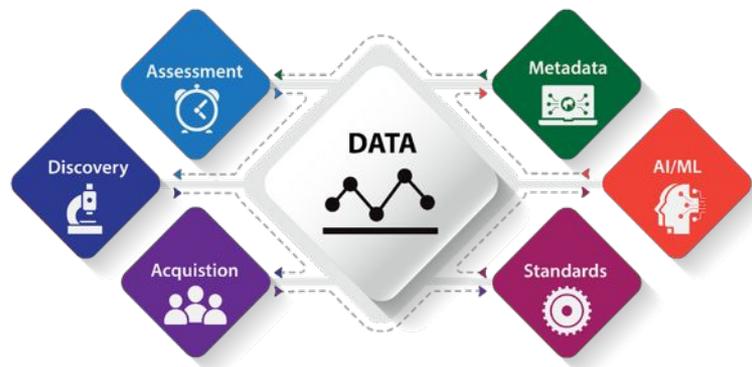


image from [technologynetworks.com](https://www.technologynetworks.com)

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

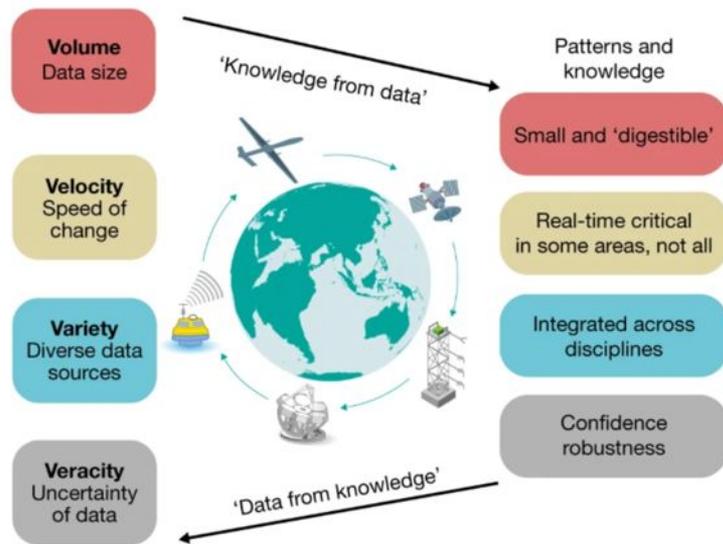


Modular Data Ecosystem to enable data interoperability for AI. Courtesy of Prakash & Serbin

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

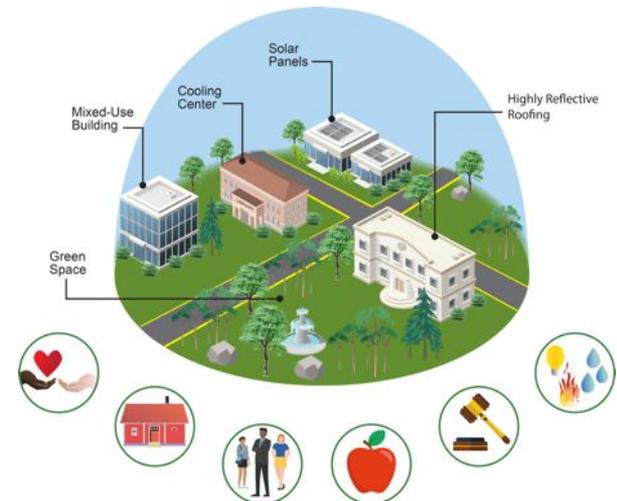


Data challenges in the earth sciences: different data sources, small data / big data challenges, and uncertainty in the data. Figure taken from (Reichstein, M. et al. 2019)

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)





THANK YOU