The Role of Earth System Science in Sustainable Rural Development

Forrest M. Hoffman
forrest@openearthscience.org

Oak Ridge National Laboratory and University of Tennessee, USA

International Conference on Redefining Rural Development: Multidisciplinary Perspectives and Global Insights

Vivekananda Global University (VGU)

July 30, 2026

Forrest M. Hoffman, Computational Earth System Scientist

- Group Leader for the ORNL Integrated Computational Earth Sciences Group
- 38 years at ORNL: 15 years in Environmental Sciences Division, 14 years in Computer Science and Mathematics Division, and 9 years in Computational Sciences and Engineering Division
- Develop and apply Earth system models to study global biogeochemical cycles, including terrestrial & marine carbon cycle
- Investigate methods for reconciling uncertainties in carbon–Earth system feedbacks through comparison with observations
- Apply artificial intelligence methods (machine learning and data mining) to environmental characterization, simulation, & analysis
- Joint Faculty, University of Tennessee, Knoxville, Department of Civil & Environmental Engineering

Oak Ridge National Laboratory

The University of Tennessee Knoxville

Introduction

- **Earth system science (ESS)** plays a key role in sustainable urban development
  - ESS offers *scientific understanding* needed to manage land, water, and ecosystems to support human well-being and environmental resilience
  - ESS uses a *systems approach* to analyze interactions among agriculture, forestry, water resources, and biodiversity as an interconnected system
  - ESS applies *advanced technologies*, like artificial intelligence, remote sensing, and high-performance computing to address questions of societal relevance
  - ESS provides an *integrated framework* for balancing objectives & tradeoffs for
    - Food production
    - Water security
    - Biodiversity conservation
    - Environmental resilience
    - Economic opportunity
    - Human health

Examples Where ESS Contributes to Urban Development

- **Sustainable agriculture**
  - Improves understanding of soil health, nutrient cycling, water availability, crop yield
  - Supports precision agriculture, crop monitoring, regenerative practices that increase productivity while reducing environmental impacts
- **Water resources management**
  - Integrates hydrology, weather, and land-use science to improve water allocation, irrigation efficiency, flood prediction, and drought preparedness
  - Informs watershed management to protect drinking water and crop productivity
- **Extreme weather adaptation and resilience**
  - Provides forecasts and risk assessments that help communities adapt cropping systems, infrastructure, and natural resource management
  - Drought monitoring systems and early warning systems reduce losses from extreme weather events

Examples Where ESS Contributes to Urban Development

- **Ecosystem conservation**
  - Informs approaches for maintaining healthy ecosystems, which provide natural services such as pollination, erosion control, and water purification
  - Identifies how land-use changes affect biodiversity and ecosystem function
- **Renewable energy and carbon management**
  - Identifies suitable locations for solar, wind, and bioenergy production
  - Quantifies carbon storage in forests, soils, wetlands, and carbon sequestration projects that may provide income opportunities
- **Evidence-based policy**
  - Combines satellite observations, field measurements, environmental models, and socioeconomic data to inform rural development policies
  - Enables decision-makers to evaluate tradeoffs among agricultural production, conservation, water use, and economic growth

Introduction

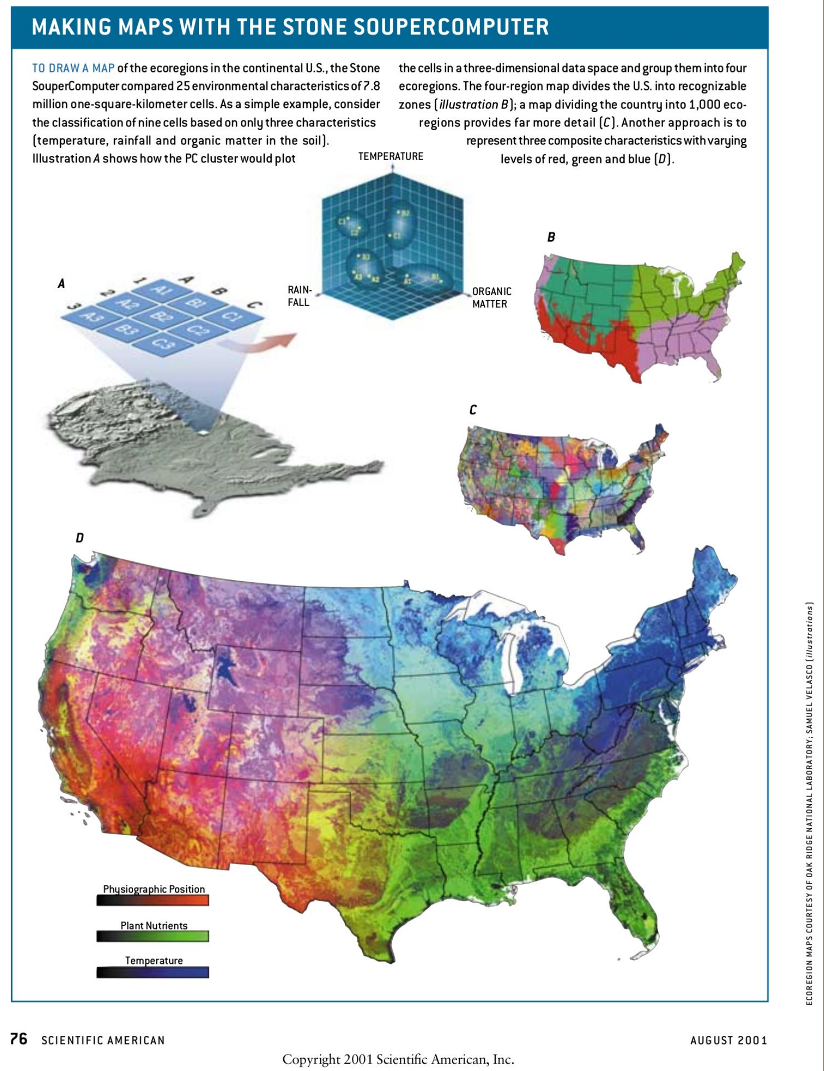
- Observations of the Earth system are increasing in spatial resolution and temporal frequency, and will grow exponentially over the next 5–10 years
- With Exascale computing, simulation output is growing even faster, outpacing our ability to analyze, interpret and evaluate model results
- Explosive 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 #3 fastest supercomputer on the [TOP500](TOP500) List (June 2026) and was the first supercomputer to break the exaflop barrier .

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*).

Eos, Vol. 84, No. 48, 2 December 2003

EOS

EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 84 NUMBER 48
2 DECEMBER 2003
PAGES 529–544

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 [*Wofsy and Harriss*, 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

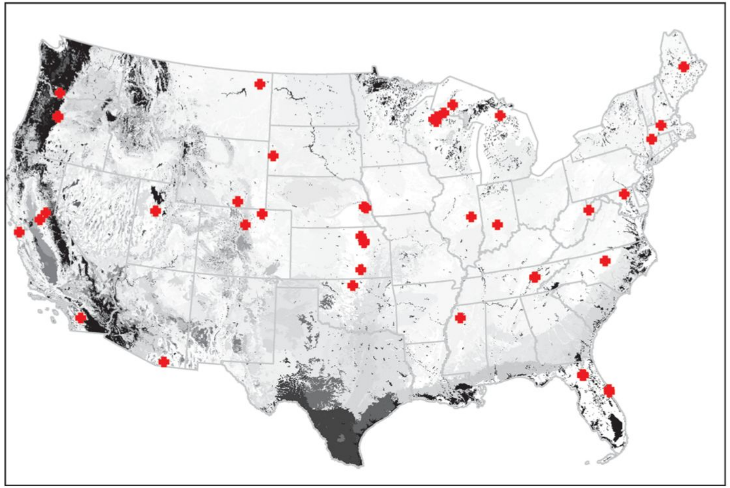


Fig. 1. The representativeness of an existing spatial array of sample locations or study sites—for example, the AmeriFlux network of carbon dioxide eddy flux covariance towers—can be mapped relative to a set of quantitative ecoregions, suggesting locations for additional samples or sites. Distance in 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.

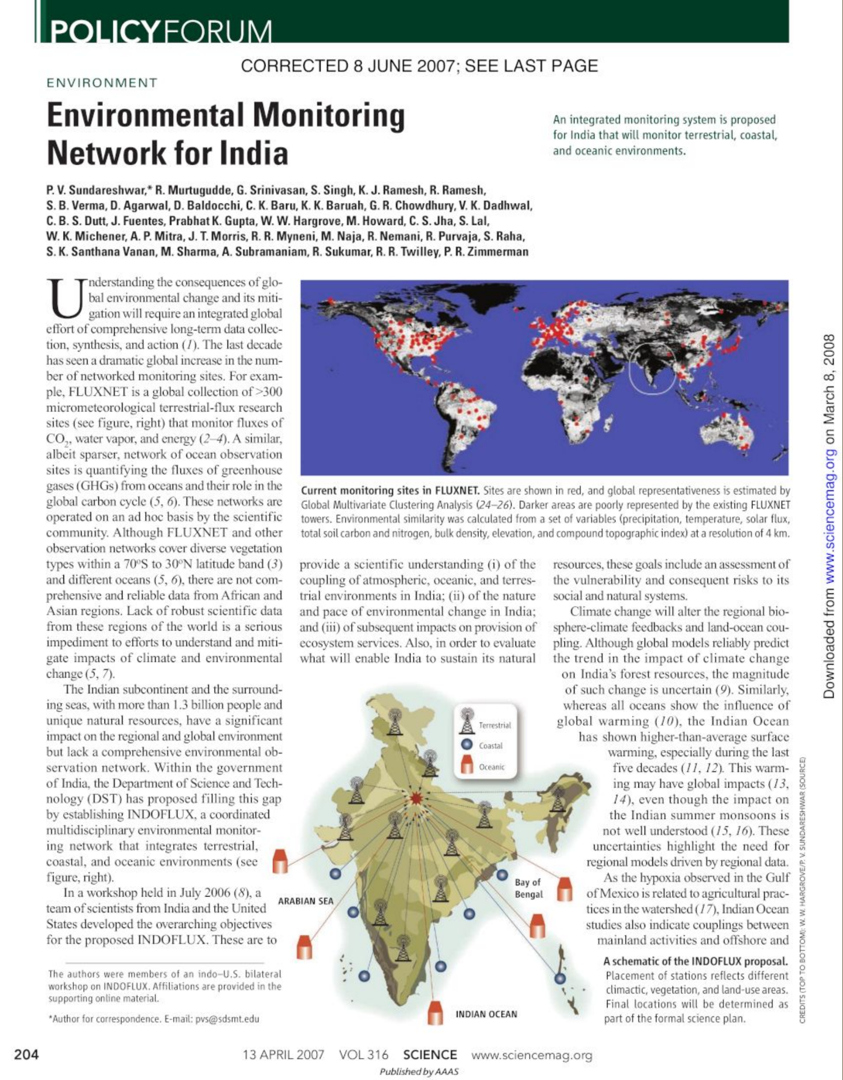
Network Representativeness

- 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.

Optimizing Sampling Networks

- Our group produced this network representativeness map for the authors from global climate, edaphic, and elevation and topography data
- Dark areas, including most of the Indian subcontinent, were poorly represented by the constellation of eddy covariance flux towers participating in FLUXNET in the year 2007



POLICYFORUM

CORRECTED 8 JUNE 2007; SEE LAST PAGE

ENVIRONMENT

Environmental Monitoring Network for India

An integrated monitoring system is proposed for India that will monitor terrestrial, coastal, and oceanic environments.

P. V. Sundareshwar,* R. Murtugudde, G. Srinivasan, S. Singh, K. J. Ramesh, R. Ramesh, S. B. Verma, D. Agarwal, D. Baldocchi, C. K. Baru, K. K. Baruah, G. R. Chowdhury, V. K. Dadhwal, C. B. S. Dutt, J. Fuentes, Prabhat K. Gupta, W. W. Hargrove, M. Howard, C. S. Jha, S. Lal, W. K. Michener, A. P. Mitra, J. T. Morris, R. R. Myneni, M. Naja, R. Nemani, R. Purvaja, S. Raha, S. K. Santhana Vanan, M. Sharma, A. Subramaniam, R. Sukumar, R. R. Twilley, P. R. Zimmerman

Understanding the consequences of global environmental change and its mitigation will require an integrated global effort of comprehensive long-term data collection, synthesis, and action (1). The last decade has seen a dramatic global increase in the number of networked monitoring sites. For example, FLUXNET is a global collection of >300 micrometeorological terrestrial-flux research sites (see figure, right) that monitor fluxes of CO_2, water vapor, and energy (2–4). A similar, albeit sparser, network of ocean observation sites is quantifying the fluxes of greenhouse gases (GHGs) from oceans and their role in the global carbon cycle (5, 6). These networks are operated on an ad hoc basis by the scientific community. Although FLUXNET and other observation networks cover diverse vegetation types within a 70°S to 30°N latitude band (3) and different oceans (5, 6), there are not comprehensive and reliable data from African and Asian regions. Lack of robust scientific data from these regions of the world is a serious impediment to efforts to understand and mitigate impacts of climate and environmental change (5, 7).

The Indian subcontinent and the surrounding seas, with more than 1.3 billion people and unique natural resources, have a significant impact on the regional and global environment but lack a comprehensive environmental observation network. Within the government of India, the Department of Science and Technology (DST) has proposed filling this gap by establishing INDOFLUX, a coordinated multidisciplinary environmental monitoring network that integrates terrestrial, coastal, and oceanic environments (see figure, right).

In a workshop held in July 2006 (8), a team of scientists from India and the United States developed the overarching objectives for the proposed INDOFLUX. These are to provide a scientific understanding (i) of the coupling of atmospheric, oceanic, and terrestrial environments in India; (ii) of the nature and pace of environmental change in India; and (iii) of subsequent impacts on provision of ecosystem services. Also, in order to evaluate what will enable India to sustain its natural resources, these goals include an assessment of the vulnerability and consequent risks to its social and natural systems.

Climate change will alter the regional biosphere-climate feedbacks and land-ocean coupling. Although global models reliably predict the trend in the impact of climate change on India's forest resources, the magnitude of such change is uncertain (9). Similarly, whereas all oceans show the influence of global warming (10), the Indian Ocean has shown higher-than-average surface warming, especially during the last five decades (11, 12). This warming may have global impacts (13, 14), even though the impact on the Indian summer monsoons is not well understood (15, 16). These uncertainties highlight the need for regional models driven by regional data. As the hypoxia observed in the Gulf of Mexico is related to agricultural practices in the watershed (17), Indian Ocean studies also indicate couplings between mainland activities and offshore and

Current monitoring sites in FLUXNET. Sites are shown in red, and global representativeness is estimated by Global Multivariate Clustering Analysis (24–26). Darker areas are poorly represented by the existing FLUXNET towers. Environmental similarity was calculated from a set of variables (precipitation, temperature, solar flux, total soil carbon and nitrogen, bulk density, elevation, and compound topographic index) at a resolution of 4 km.

A schematic of the INDOFLUX proposal. Placement of stations reflects different climactic, vegetation, and land-use areas. Final locations will be determined as part of the formal science plan.

The authors were members of an indo–U.S. bilateral workshop on INDOFLUX. Affiliations are provided in the supporting online material.

*Author for correspondence. E-mail: pvs@sdsmt.edu

CREDITS (TOP TO BOTTOM): W. W. HARGROVE/P. V. SUNDARESHWAR (SOURCE)

Sundareshwar, P. V., et al. (2007), Environmental Monitoring Network for India, *Science*, 316(5822):204–205, doi:10.1126/science.1137417.

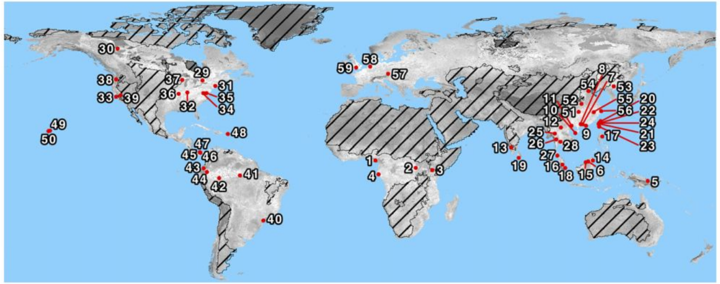


Fig. 1 Map of the CTFS-ForestGEO network illustrating its representation of bioclimatic, edaphic, and topographic conditions globally. Site numbers correspond to ID# in Table 2. Shading indicates how well the network of sites represents the suite of environmental factors included in the analysis; light-colored areas are well-represented by the network, while dark colored areas are poorly represented. Stippling covers nonforest areas. The analysis is described in Appendix S1.

Table 1 Attributes of a CTFS-ForestGEO census

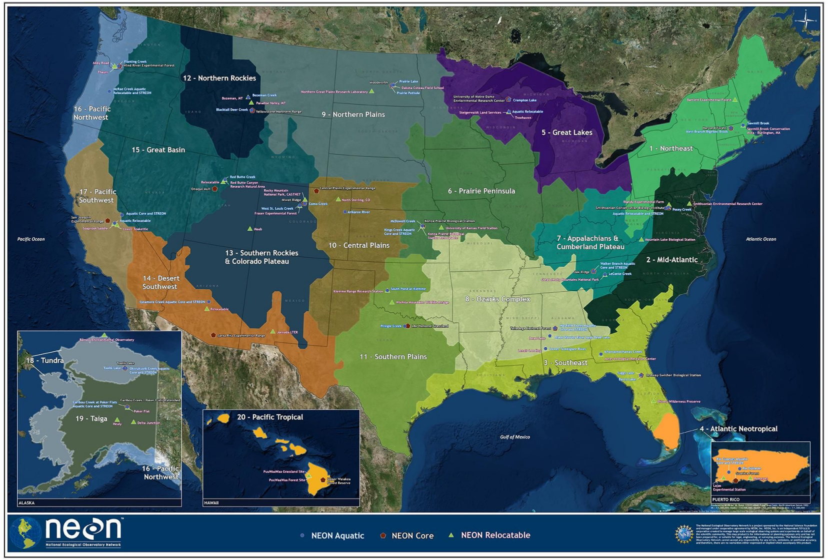
Attribute	Utility
Very large plot size	Resolve community and population dynamics of highly diverse forests with many rare species with sufficient sample sizes (Losos & Leigh, 2004; Condit *et al.*, 2006); quantify spatial patterns at multiple scales (Condit *et al.*, 2000; Wiegand *et al.*, 2007a,b; Detto & Muller-Landau, 2013; Lutz *et al.*, 2013); characterize gap dynamics (Feeley *et al.*, 2007b); calibrate and validate remote sensing and models, particularly those with large spatial grain (Mascaro *et al.*, 2011; Réjou-Méchain *et al.*, 2014)
Includes every freestanding woody stem ≥1 cm DBH	Characterize the abundance and diversity of understory as well as canopy trees; quantify the demography of juveniles (Condit, 2000; Muller-Landau *et al.*, 2006a,b).
All individuals identified to species	Characterize patterns of diversity, species-area, and abundance distributions (Hubbell, 1979, 2001; He & Legendre, 2002; Condit *et al.*, 2005; John *et al.*, 2007; Shen *et al.*, 2009; He & Hubbell, 2011; Wang *et al.*, 2011; Cheng *et al.*, 2012); test theories of competition and coexistence (Brown *et al.*, 2013); describe poorly known plant species (Gereau & Kenfack, 2000; Davies, 2001; Davies *et al.*, 2001; Sonké *et al.*, 2002; Kenfack *et al.*, 2004, 2006)
Diameter measured on all stems	Characterize size-abundance distributions (Muller-Landau *et al.*, 2006b; Lai *et al.*, 2013; Lutz *et al.*, 2013); combine with allometries to estimate whole-ecosystem properties such as biomass (Chave *et al.*, 2008; Valencia *et al.*, 2009; Lin *et al.*, 2012; Ngo *et al.*, 2013; Muller-Landau *et al.*, 2014)
Mapping of all stems and fine-scale topography	Characterize the spatial pattern of populations (Condit, 2000); conduct spatially explicit analyses of neighborhood influences (Condit *et al.*, 1992; Hubbell *et al.*, 2001; Uriarte *et al.*, 2004, 2005; Rüger *et al.*, 2011, 2012; Lutz *et al.*, 2014); characterize microhabitat specificity and controls on demography, biomass, etc. (Harms *et al.*, 2001; Valencia *et al.*, 2004; Chuyong *et al.*, 2011); align on the ground and remote sensing measurements (Asner *et al.*, 2011; Mascaro *et al.*, 2011).
Census typically repeated every 5 years	Characterize demographic rates and changes therein (Russo *et al.*, 2005; Muller-Landau *et al.*, 2006a,b; Feeley *et al.*, 2007a; Lai *et al.*, 2013; Stephenson *et al.*, 2014); characterize changes in community composition (Losos & Leigh, 2004; Chave *et al.*, 2008; Feeley *et al.*, 2011; Swenson *et al.*, 2012; Chisholm *et al.*, 2014); characterize changes in biomass or productivity (Chave *et al.*, 2008; Banin *et al.*, 2014; Muller-Landau *et al.*, 2014)

Optimizing Sampling Networks

- The CTFS-ForestGEO global forest monitoring network is aimed at characterizing forest responses to global change
- The figure at left shows the global representativeness of the CTFS-ForestGEO sites in 2014
- Non-forested areas are masked with hatching, and as expected, they are consistently darker than the forested regions, which are represented to varying degrees by the monitoring sites

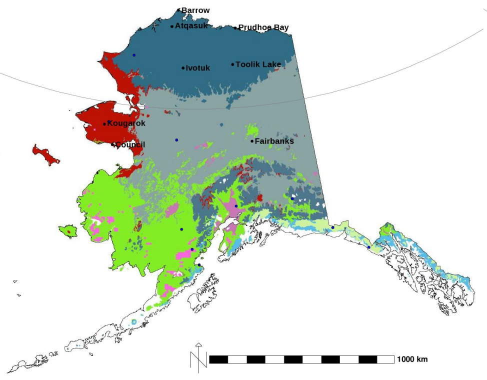
Anderson-Teixeira, K. J., et al. (2015), CTFS-ForestGEO: A Worldwide Network Monitoring Forests in an Era of Global Change, *Glob. Change Biol.*, 21(2):528–549, doi:10.1111/gcb.12712.

Sampling Network Design

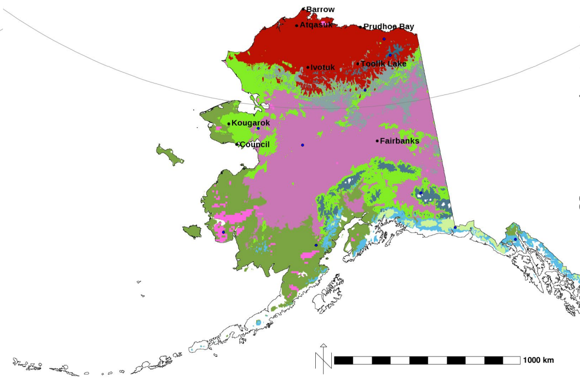


NSF's NEON Sampling Domains

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



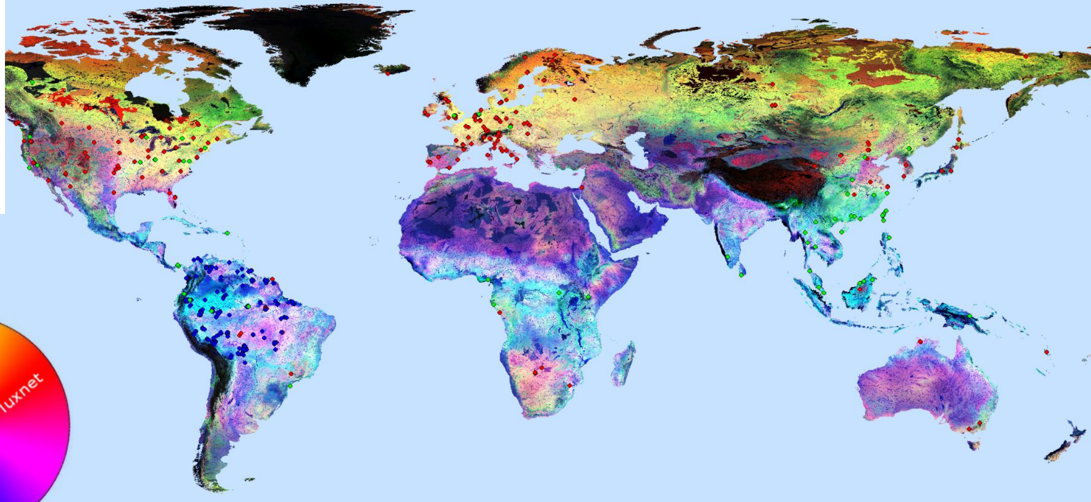
2000–2009

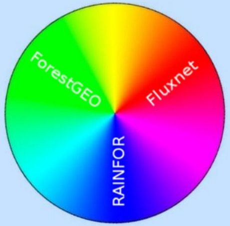


2090–2000

(Hoffman et al., 2013)

Triple-Network Global Representativeness

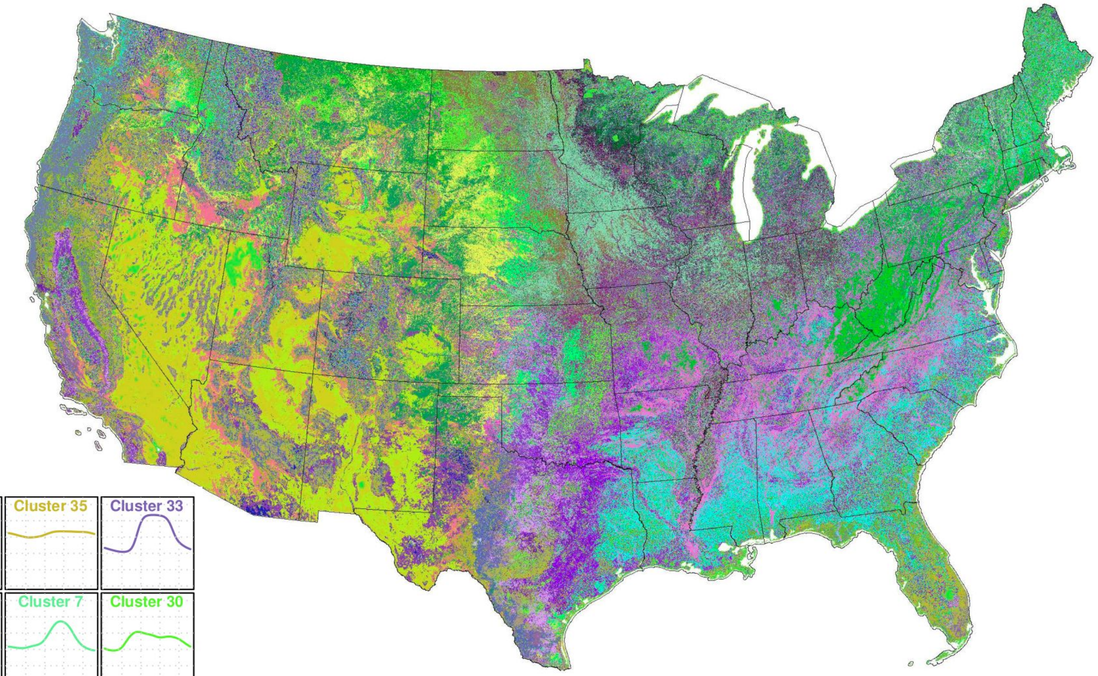


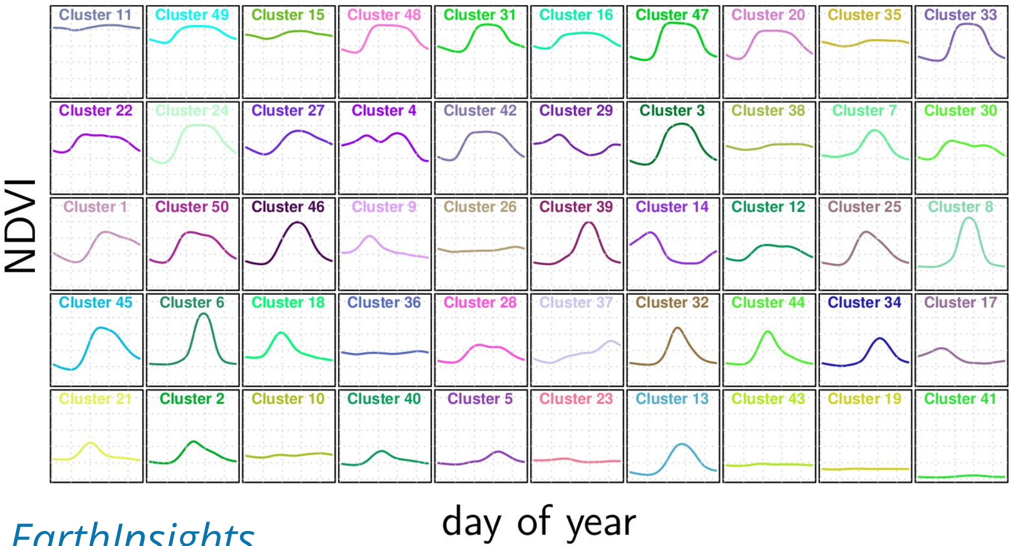


(Maddalena et al., in prep.)

50 Phenoregions for year 2012 (Random Colors)

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

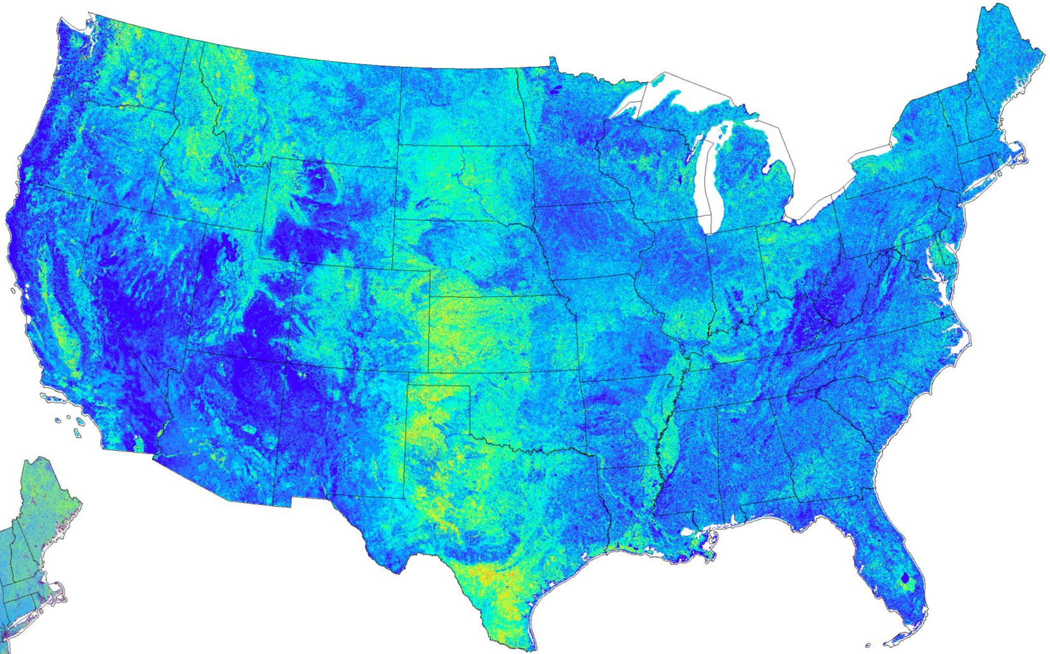


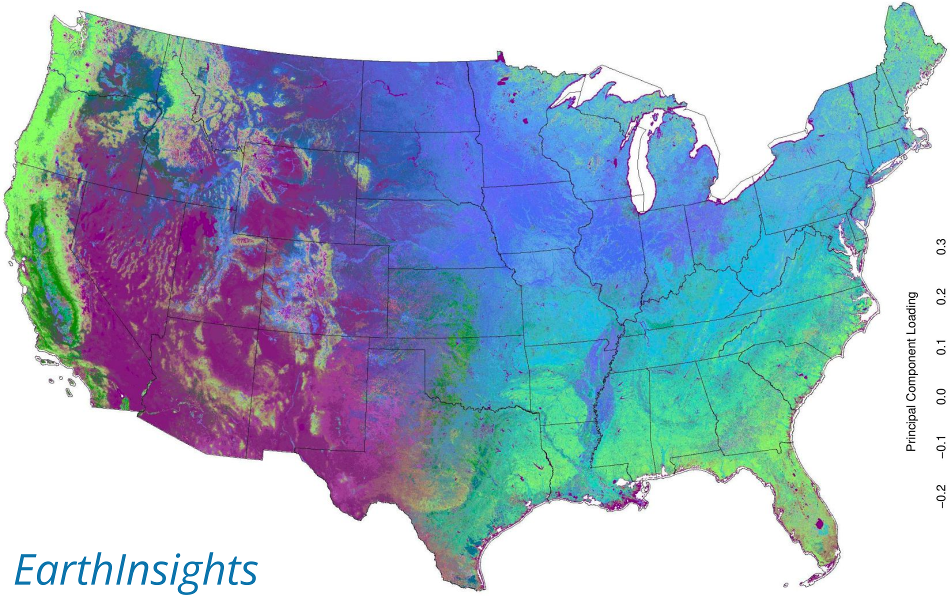


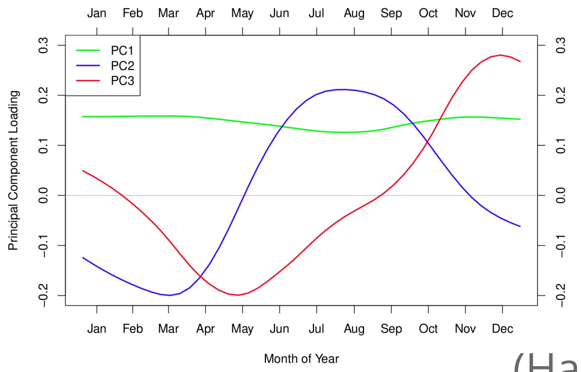
50 Phenoregion Prototypes (Random Colors)

(Hargrove et al., in prep.)

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







Principal Components Analysis

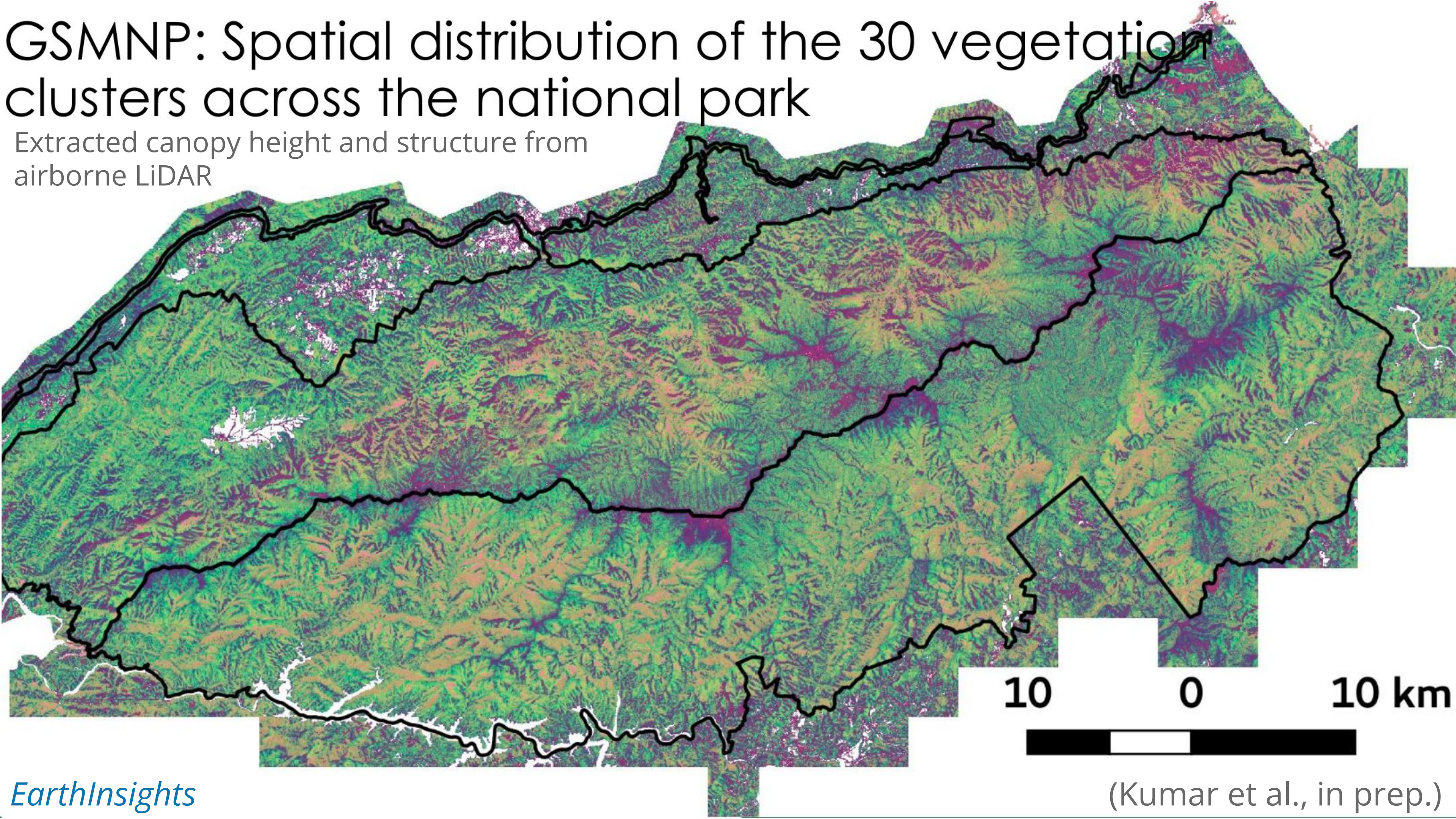
PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

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



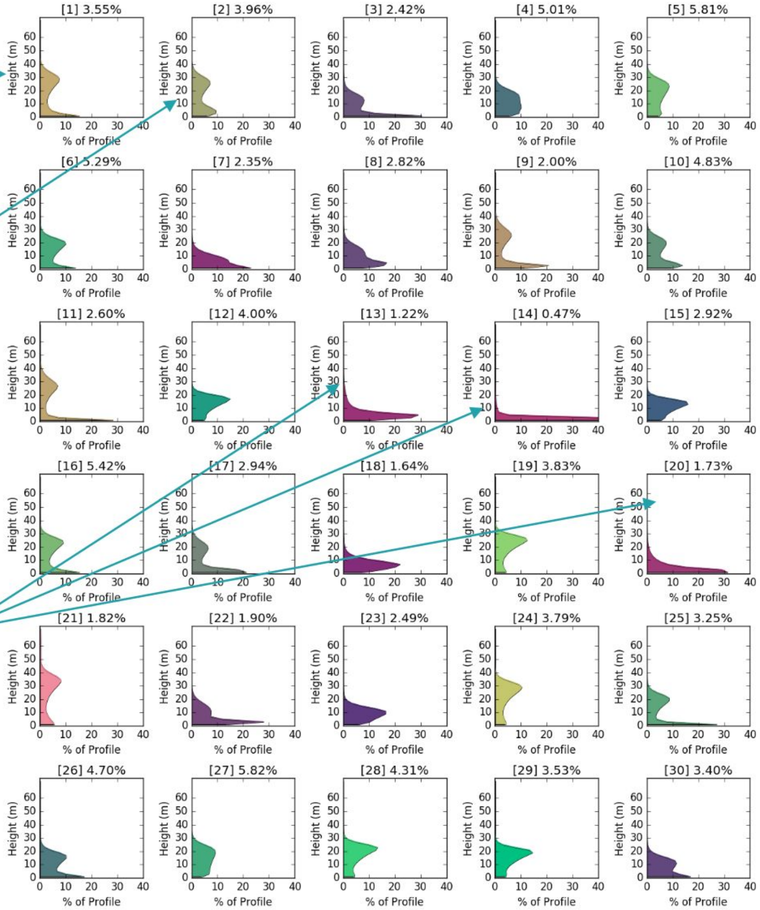
(Kumar et al., in prep.)

GSMNP: 30 representative vertical structures (cluster centroids) identified

tall forests with low understory vegetation

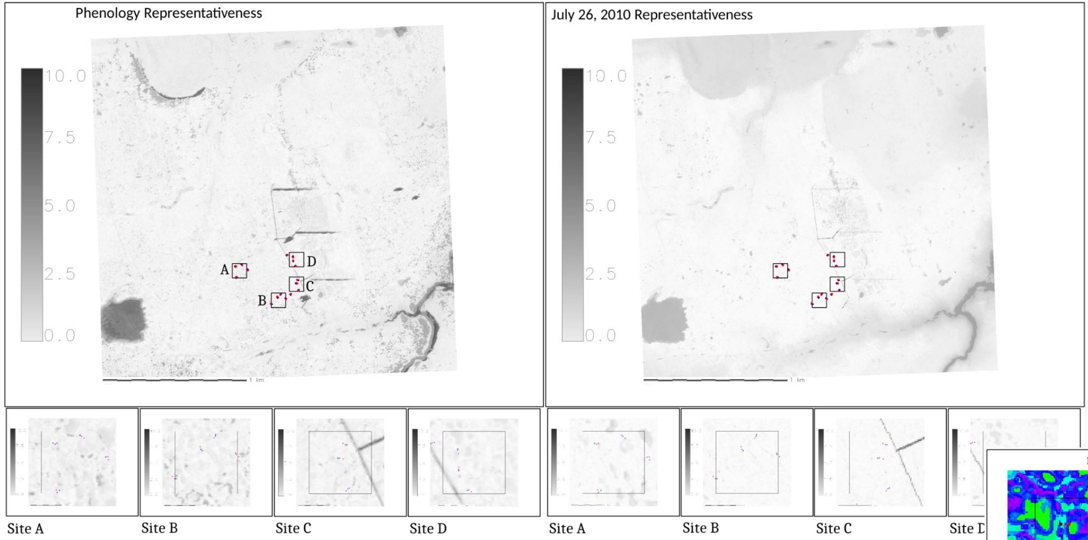
forests with slightly lower mean height with dense understory vegetation

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



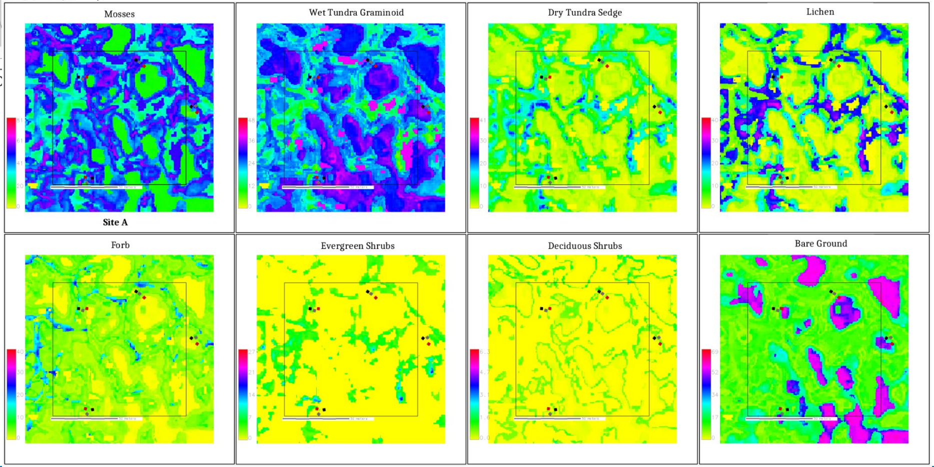
(Kumar et al., in prep.)

Vegetation Distribution at Barrow Environmental Observatory



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

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

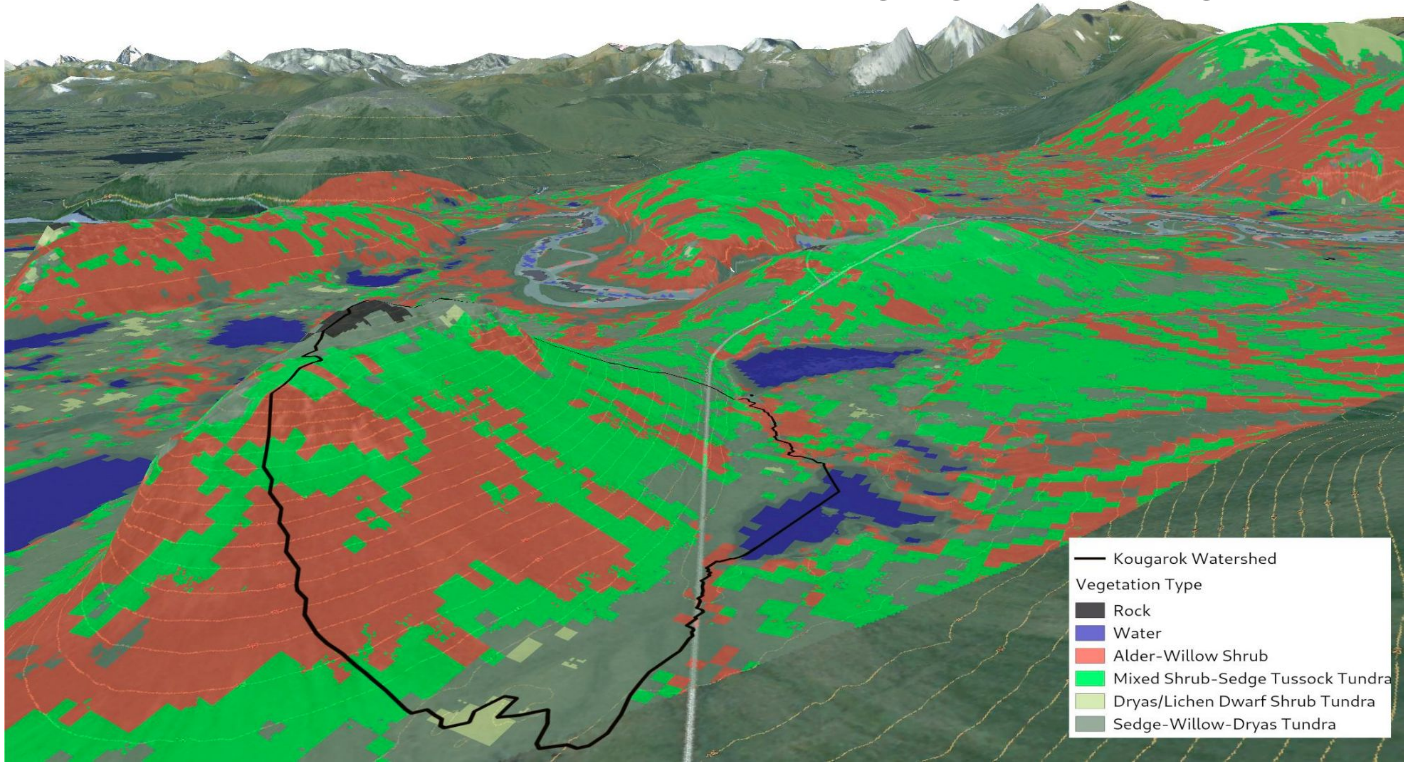


In situ data from field measurement activities inform the development of wide-scale maps of vegetation distribution through inference using remote sensing data as surrogate variables, and relationships with environmental controls can be extracted

Langford, Z. L., et al. (2016), Mapping Arctic Plant Functional Type Distributions in the Barrow Environmental Observatory Using WorldView-2 and LiDAR Datasets, *Remote Sens.*, 8(9):733, doi:10.3390/rs8090733.

Arctic Vegetation Mapping from Multi-Sensor Fusion

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



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

Satellite Data Analytics Enables Within-Season Crop Identification

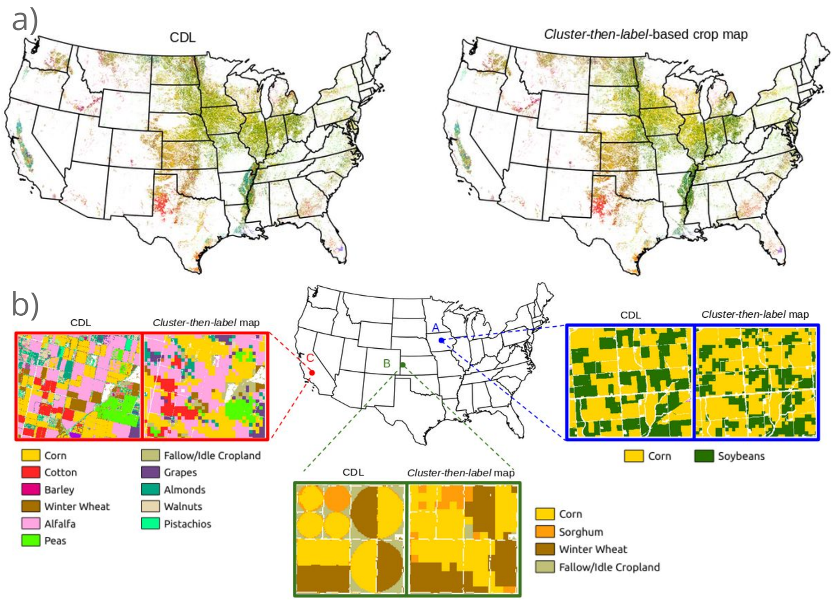
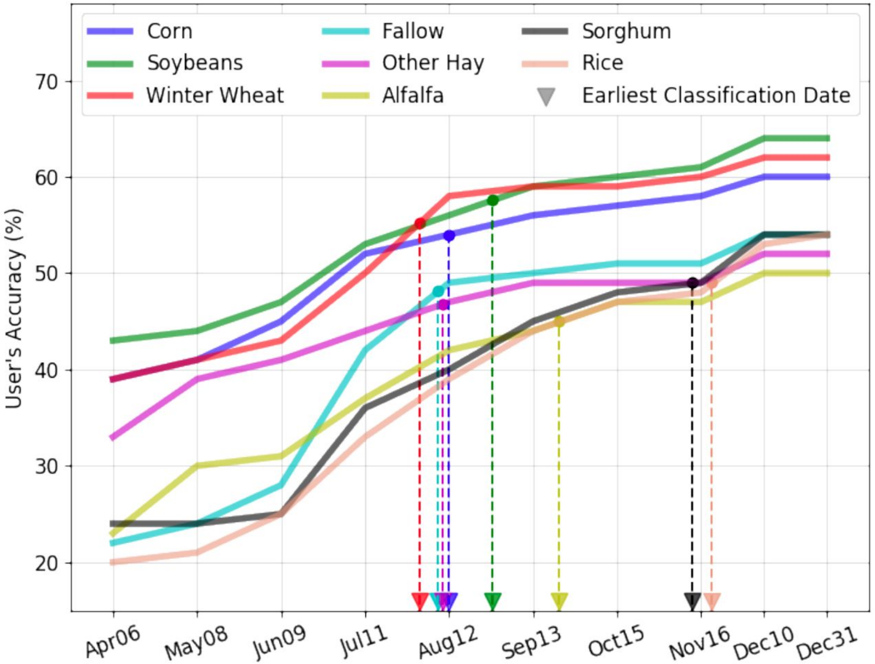


Figure: a) Comparison of cluster-then-label crop map with USDA Crop Data Layer (CDL) shows similar patterns at continental scale. b) Good spatial agreement is found at three selected regions, but cluster-then-label crop maps lack sharpness at field boundaries due to coarser resolution of MODIS data.

Earliest date for crop type classification



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

Leveraging Advances in Machine Learning for Earth Sciences

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

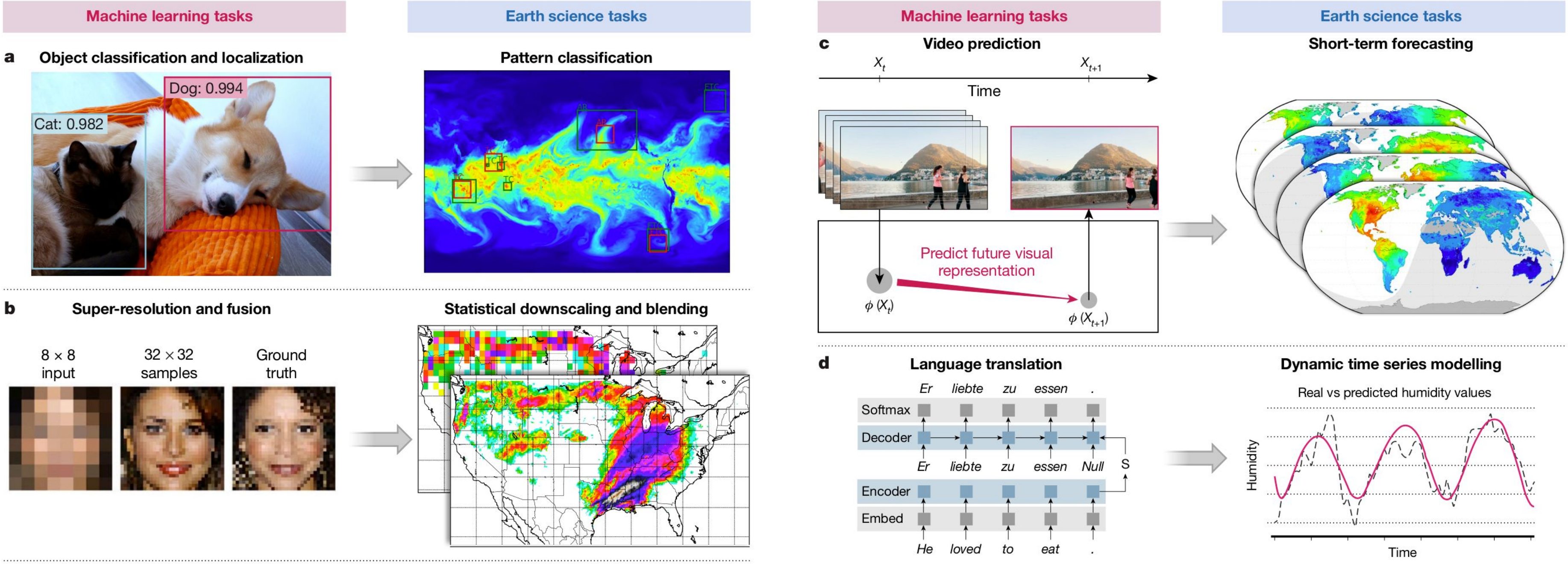
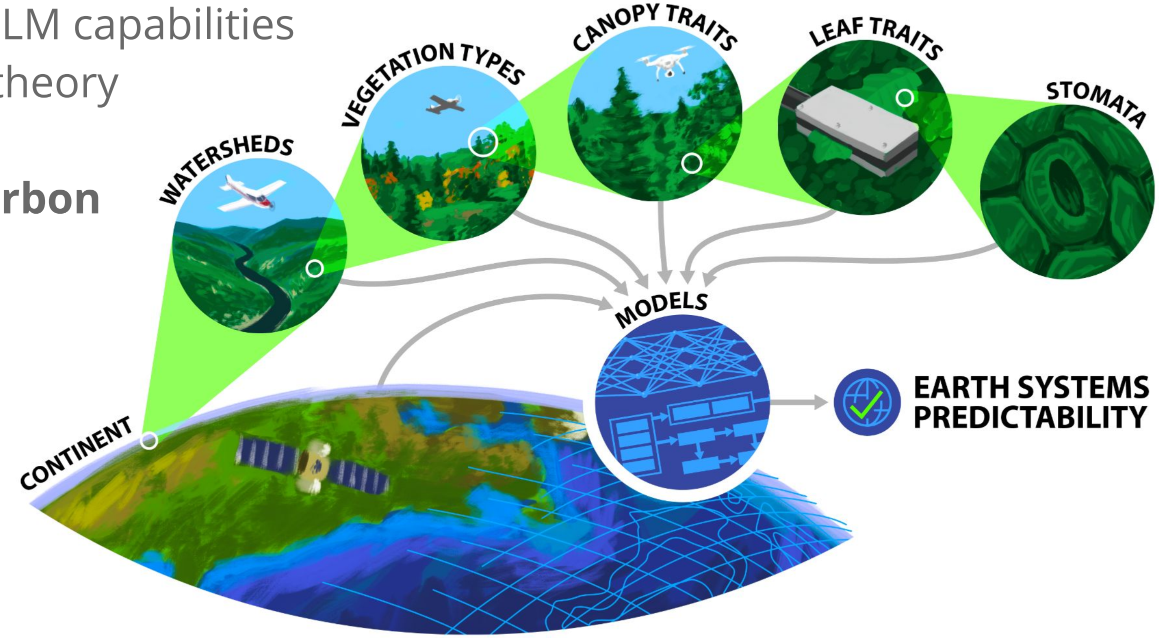


Figure 2 in Reichstein et al. (2019)

Machine Learning for Understanding Biospheric Processes

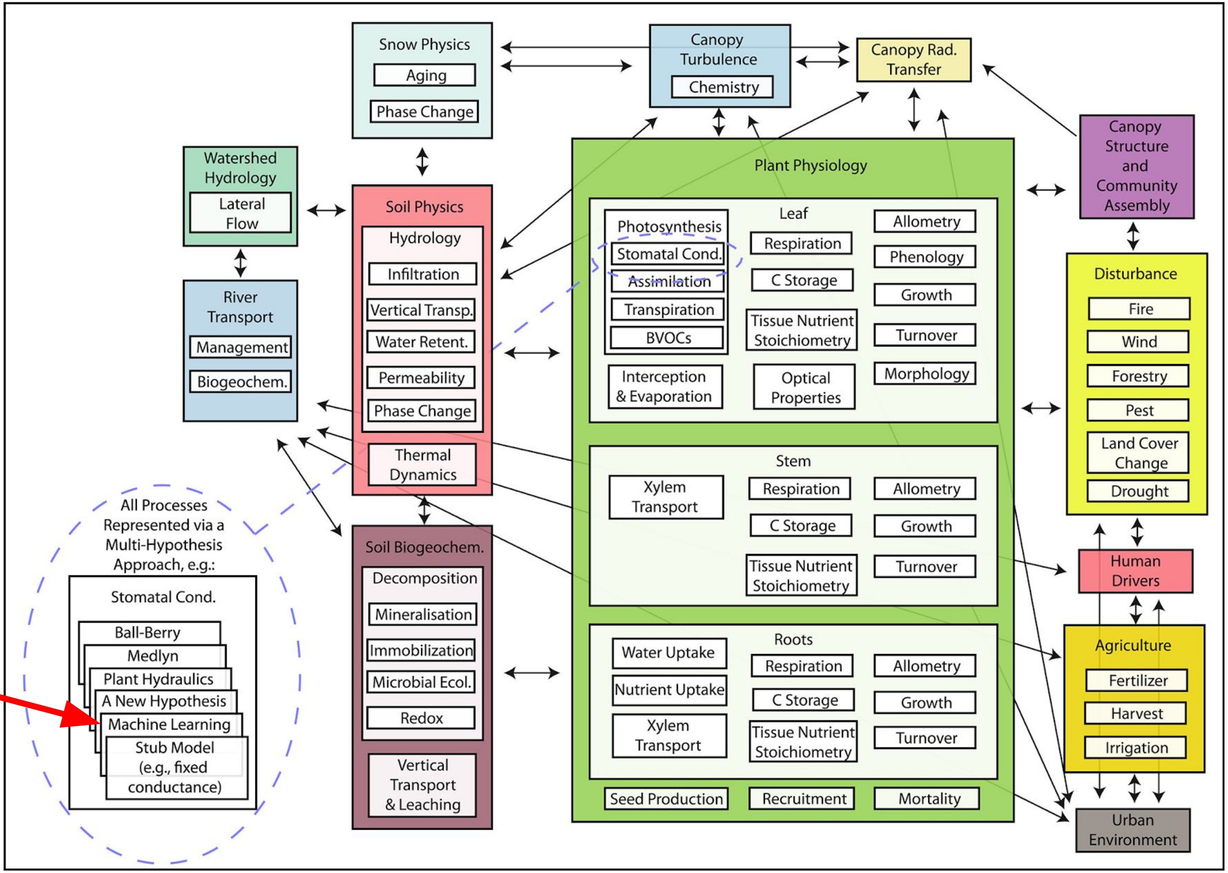
- Widening adoption of deep neural networks and growth of climate data are fueling interest in AI/ML for use in weather and climate and Earth system models
- ML potential is high for improving predictability when (1) *sufficient data are available for process representations* and (2) *process representations 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 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

AI-Constrained Ecohydrology for Improving Earth System Predictions

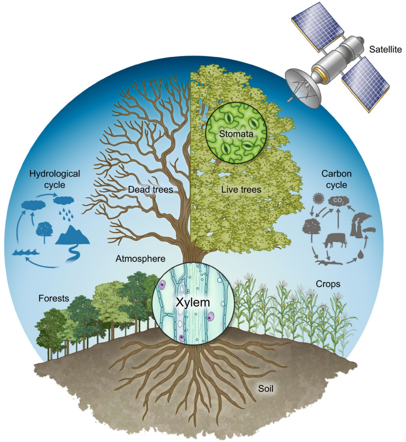
Collaboration among ORNL, LANL, Penn State, et al.

Contact: Forrest M. Hoffman

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

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

Nature



McDowell et al. (2019)

The Future is Bright for AI/ML in Earth System Science

A Convergence of New Technology, Explosive Data Growth, and Free Tools

- High performance computing (exascale in big centers and commercial cloud)
- Large data storage resources (commercial and on-premise cloud)
- High speed networks (e.g., ESnet) and data movement technologies (Globus)
- Satellites (shoebox CubeSats) and airborne (drones) platforms
- Cheap (free!) and easy-to-use ML tools (PyTorch, Keras, Scikit-Learn)

Future Applications Could Revolutionize Our Understanding and Ability to Predict

- Poorly understood processes and mechanisms can be mimicked with adequate amounts of data and advanced ML techniques
- Explainable AI and systematic approaches to modeling could lead to new scientific discoveries and improved understanding of the Earth system
- Predictions of complex, nonlinear, large-scale phenomena and natural hazards could be predicted with increasing accuracy

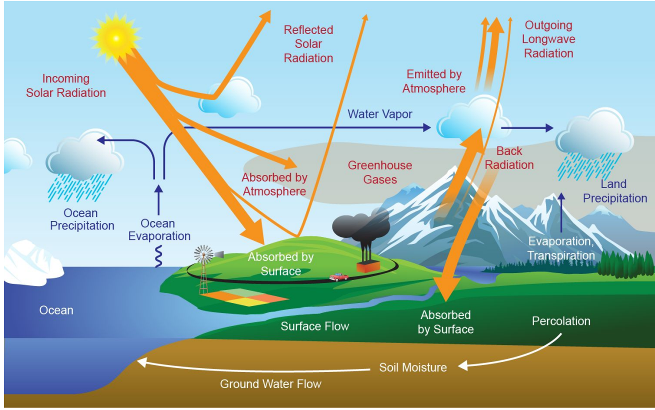
International Land Model Benchmarking (ILAMB)

What is ILAMB?

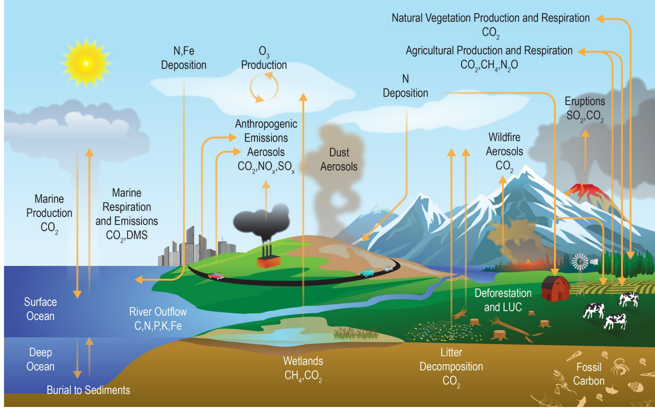
RUBISCO

A community coordination activity created to:

- **Develop internationally accepted benchmarks** for land model performance by drawing upon collaborative expertise
- **Promote the use of these benchmarks** for model intercomparison
- **Strengthen linkages between experimental, remote sensing, and Earth system modeling communities** in the design of new model tests and new measurement programs
- **Support the design and development of open source benchmarking tools**



Energy and Water Cycles



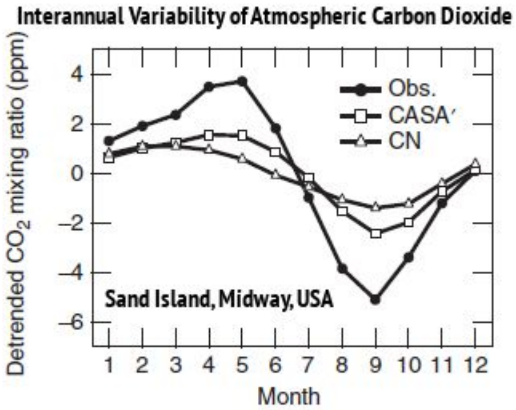
Carbon and Biogeochemical Cycles

BERKELEY LAB | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH | OAK RIDGE National Laboratory | Sandia National Laboratories | THE UNIVERSITY OF CALIFORNIA IRVINE | MICHIGAN

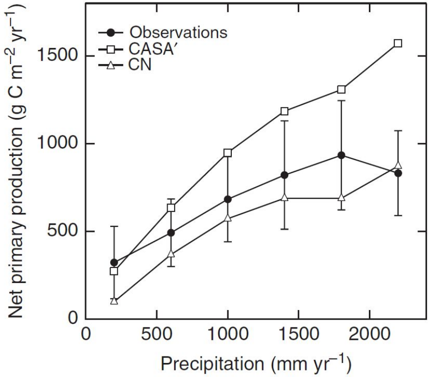
What is a Benchmark?

RUBISCO

- A **benchmark** is a quantitative test of model function achieved through comparison of model results with observational data
- Acceptable performance on a benchmark **is a necessary but not sufficient condition** for a fully functioning model
- **Functional relationship benchmarks** offer tests of model responses to forcings and yield insights into ecosystem processes
- Effective benchmarks must draw upon **a broad set of independent observations** to evaluate model performance at multiple scales



Models often fail to capture the amplitude of the seasonal cycle of atmospheric CO_2



Models may reproduce correct responses over only a limited range of forcing variables

(Randerson et al., 2009)

Berkeley Lab · NCAR National Center for Atmospheric Research · Oak Ridge National Laboratory · Sandia National Laboratories · The University of California Irvine · Michigan

Why Benchmark Models?

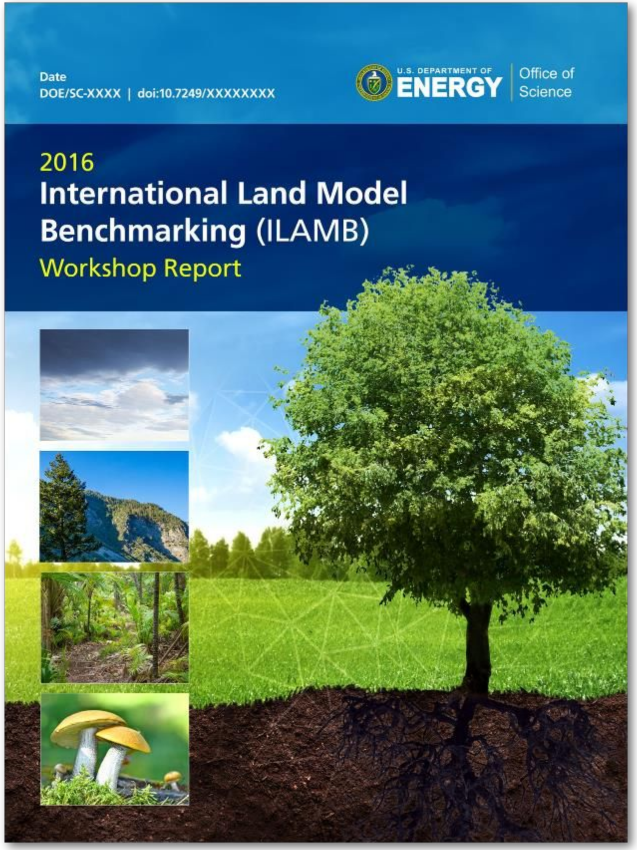
- To **quantify and reduce uncertainties** in carbon cycle feedbacks to improve projections of future climate change (Eyring et al., 2019; Collier et al., 2018)
- To **diagnose impacts of process-based or machine learning model development** on process representations and their interactions
- To **guide synthesis efforts**, such as the Intergovernmental Panel on Climate Change (IPCC), by determining which models are broadly consistent with observations (Eyring et al., 2019)
- To **increase scrutiny of key datasets** used for model evaluation
- To **identify gaps in existing observations** needed to inform model development
- To **accelerate delivery of new measurement datasets** for rapid and widespread use in model assessment

- **First ILAMB Workshop** was held in Exeter, UK, on June 22–24, 2009
- **Second ILAMB Workshop** was held in Irvine, CA, USA, on January 24–26, 2011
  - ~45 researchers participated from the US, Canada, UK, Netherlands, France, Germany, Switzerland, China, Japan, and Australia
  - Developed methodology for model-data comparison and baseline standard for performance of land model process representations (Luo et al., 2012)

2016 International Land Model Benchmarking (ILAMB) Workshop
May 16–18, 2016, Washington, DC

Third ILAMB Workshop was held May 16–18, 2016

- Workshop Goals
  - Design of new metrics for model benchmarking
  - Model Intercomparison Project (MIP) evaluation needs
  - Model development, testbeds, and workflow processes
  - Observational datasets and needed measurements
- Workshop Attendance
  - 60+ participants from Australia, Japan, China, Germany, Sweden, Netherlands, UK, and US (10 modeling centers)
  - ~25 remote attendees at any time



(Hoffman et al., 2017)

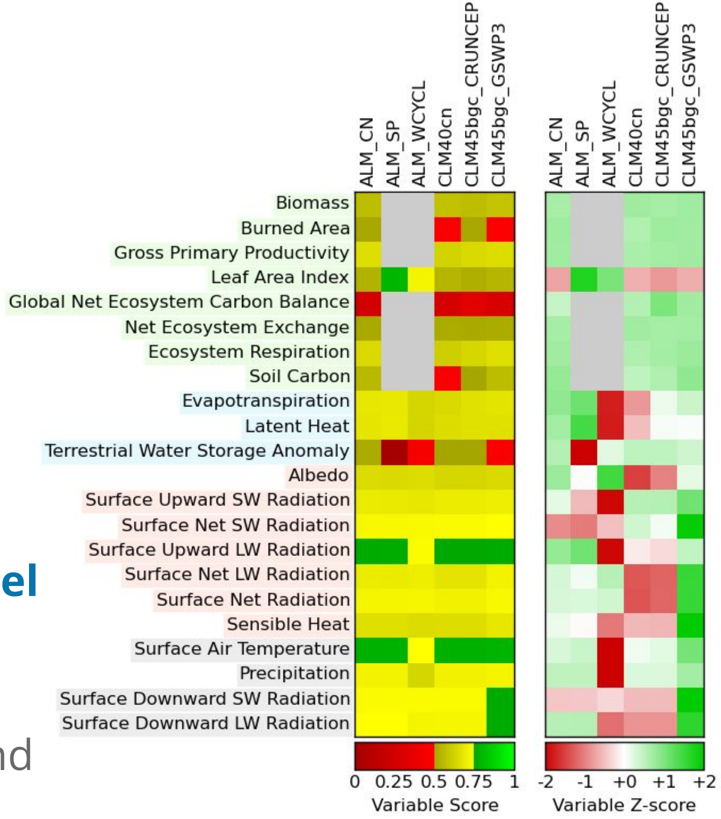
BERKELEY LAB
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE National Laboratory
Sandia National Laboratories
THE UNIVERSITY OF CALIFORNIA IRVINE
MICHIGAN

2025 ILAMB Meeting in New Orleans (December 2025)

60+ participants from 6 countries — Report coming soon!

Development of ILAMB Packages

- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall, doi:[10.18139/ILAMB.v001.00/1251597](https://doi.org/10.18139/ILAMB.v001.00/1251597)
- **ILAMBv2** released at 2016 ILAMB Workshop, doi:[10.18139/ILAMB.v002.00/1251621](https://doi.org/10.18139/ILAMB.v002.00/1251621)
- **ILAMBv3** released at 2025 ILAMB Meeting
- **Open Source software** written in Python; **runs in parallel** on laptops, clusters, and supercomputers
- Routinely used for land model evaluation during development of ESMs, including the **E3SM Land Model** (Zhu et al., 2019) and the **CESM Community Land Model** (Lawrence et al., 2019)
- Used to evaluate **TRENDY models** for annual GCB
- **Models are scored** based on statistical comparisons and functional response metrics



ILAMB Produces Diagnostics and Scores Models

- ILAMB generates a top-level **portrait plot** of models scores
- For every variable and dataset, ILAMB can automatically produce
  - **Tables** containing individual metrics and metric scores (when relevant to the data), including
    - Benchmark and model **period mean**
    - **Bias** and **bias score** (S_{bias})
    - **Root-mean-square error (RMSE)** and **RMSE score** (S_{rmse})
    - **Phase shift** and **seasonal cycle score** (S_{phase})
    - **Interannual coefficient of variation** and **IAV score** (S_{iav})
    - **Spatial distribution score** (S_{dist})
    - **Overall score** ($S_{overall}$) →

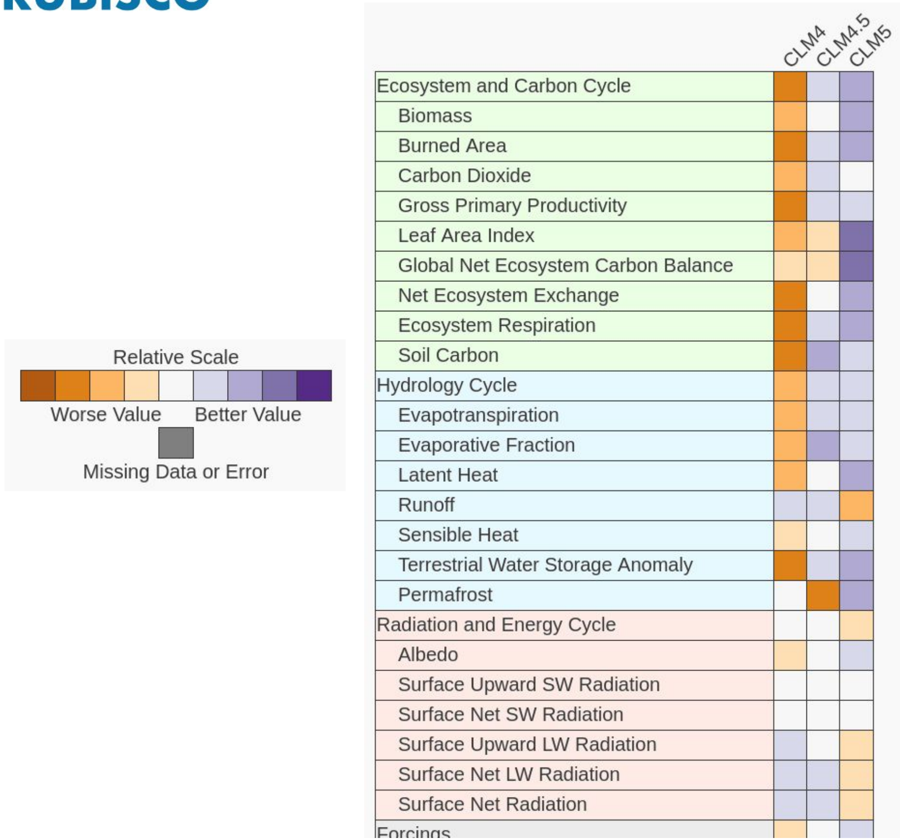
$$S_{\text{overall}} = \frac{S_{\text{bias}} + 2S_{\text{rmse}} + S_{\text{phase}} + S_{\text{iav}} + S_{\text{dist}}}{1+2+1+1+1}$$

  - **Graphical diagnostics**
    - Spatial contour maps
    - Time series line plots
    - Spatial Taylor diagrams (Taylor, 2001)
- Similar **tables** and **graphical diagnostics** for functional relationships

ILAMB Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), CO_2 (NOAA GMD, Mauna Loa), Gross primary production (Fluxnet, GBAF), Leaf area index (AVHRR, MODIS), Global net ecosystem carbon balance (GCP, Khatiwala/Hoffman), Net ecosystem exchange (Fluxnet, GBAF), Ecosystem Respiration (Fluxnet, GBAF), Soil C (HWSD, NCSCDv22, Koven)
- **Hydrology:** Evapotranspiration (GLEAM, MODIS), Evaporative fraction (GBAF), Latent heat (Fluxnet, GBAF, DOLCE), Runoff (Dai, LORA), Sensible heat (Fluxnet, GBAF), Terrestrial water storage anomaly (GRACE), Permafrost (NSIDC)
- **Energy:** Albedo (CERES, GEWEX.SRB), Surface upward and net SW/LW radiation (CERES, GEWEX.SRB, WRMC.BSRN), Surface net radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)
- **Forcing:** Surface air temperature (CRU, Fluxnet), Diurnal max/min/range temperature (CRU), Precipitation (CMAP, Fluxnet, GPCC, GPCP2), Surface relative humidity (ERA), Surface down SW/LW radiation (CERES, Fluxnet, GEWEX.SRB, WRMC.BSRN)

ILAMB Assessing Multiple Generations of CLM

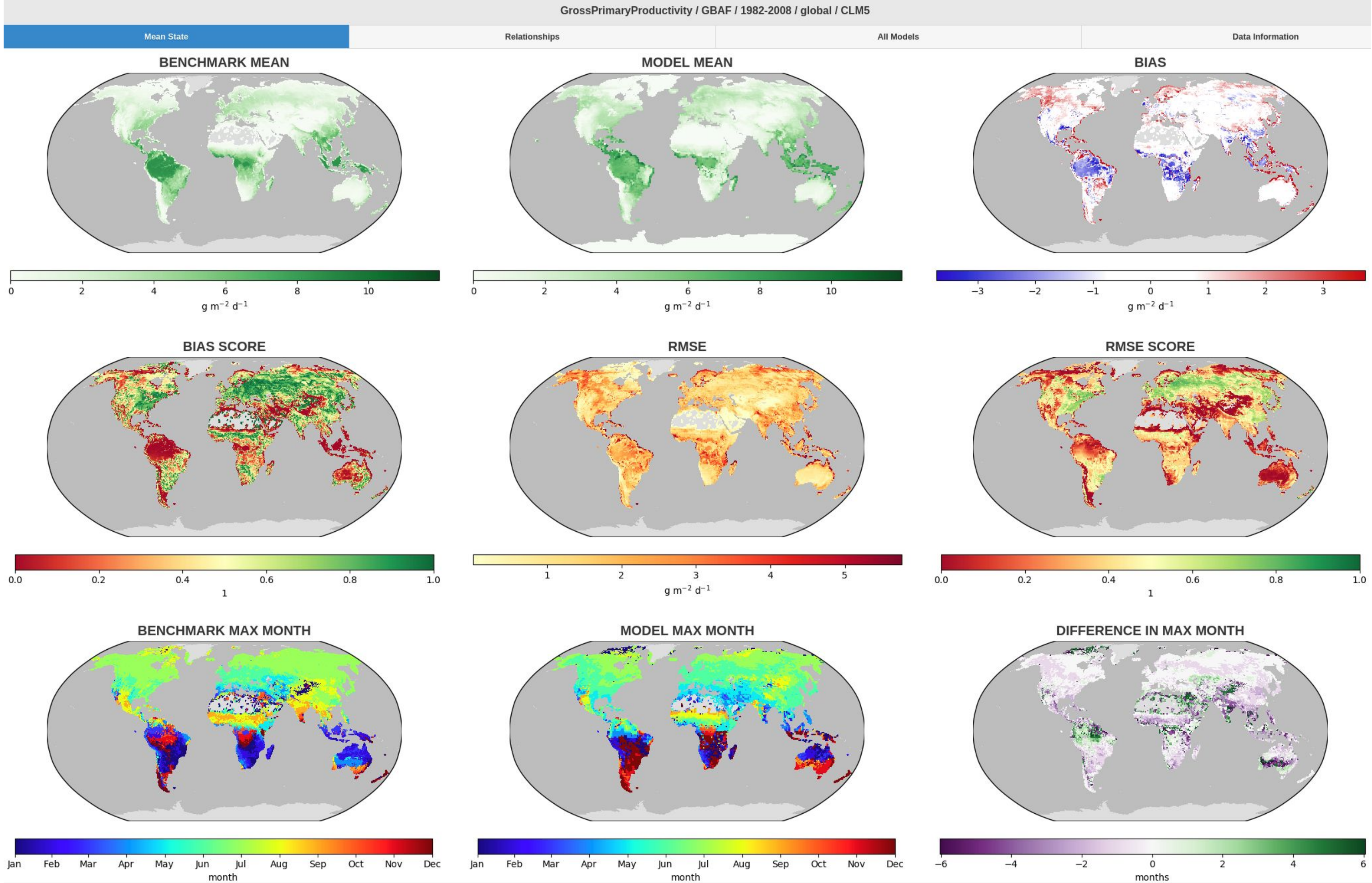


- Improvements in mechanistic treatment of hydrology, ecology, and land use with much more complexity in Community Land Model version 5 (CLM5)
- Simulations improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development

http://webext.cgd.ucar.edu/I20TR/_build_set1F/

(Lawrence et al., 2019)

ILAMB Graphical Diagnostics

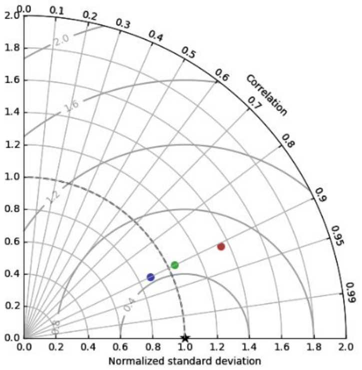


RUBISCO

Berkeley Lab NCAR National Center for Atmospheric Research Oak Ridge National Laboratory Sandia National Laboratories The University of California Irvine Michigan

ILAMB Graphical Diagnostics

SPATIAL TAYLOR DIAGRAM



MODEL COLORS

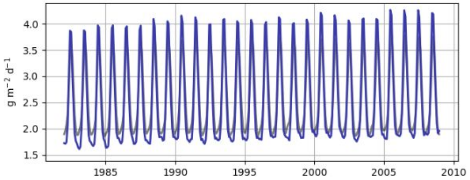
- Benchmark
- CLM4
- CLM4.5
- CLM5

Spatially integrated regional mean

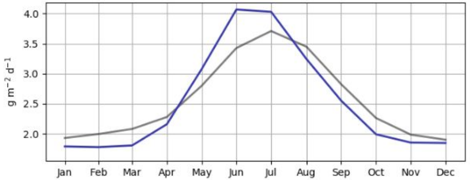
MODEL COLORS

- Benchmark
- CLM4
- CLM4.5
- CLM5

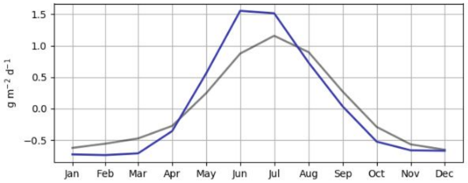
REGIONAL MEAN



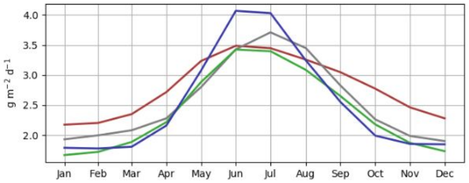
ANNUAL CYCLE



MONTHLY ANOMALY

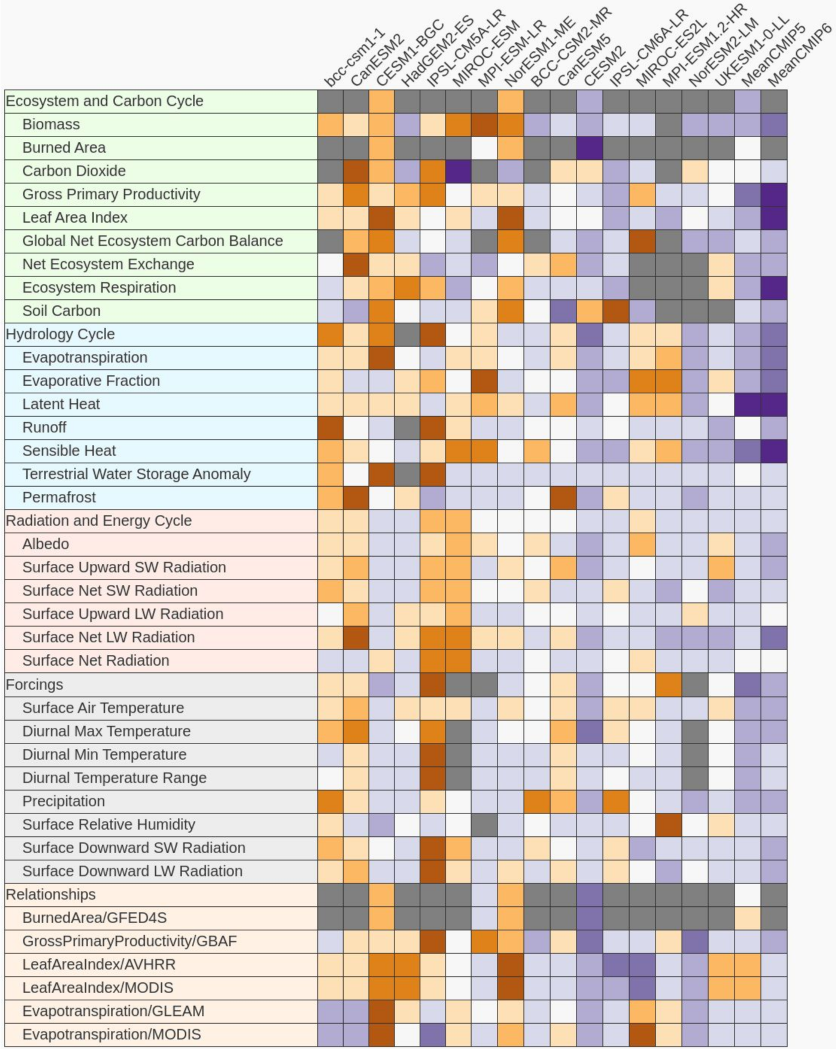


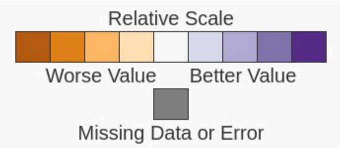
ANNUAL CYCLE



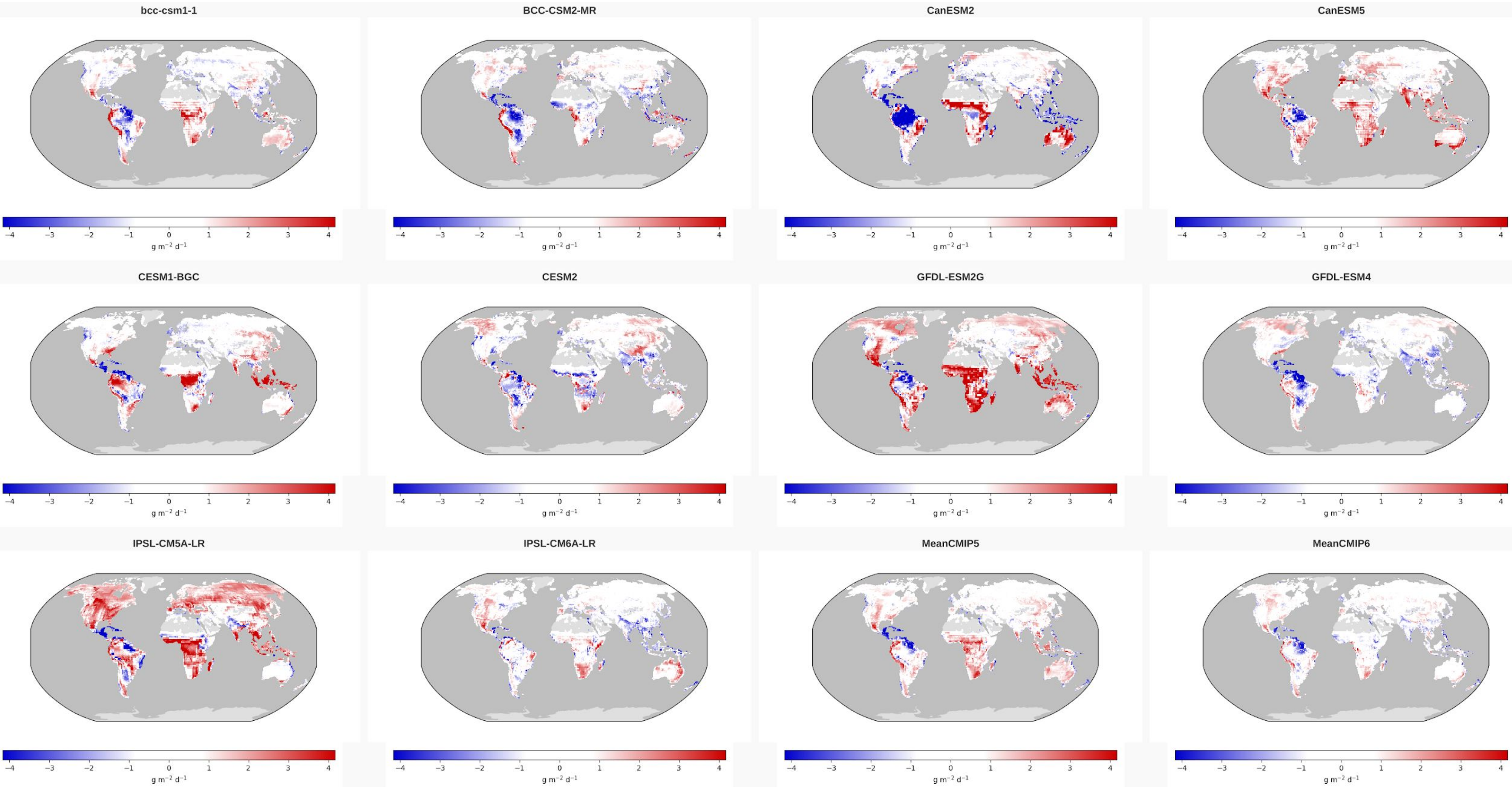
CMIP5 vs. CMIP6 Models

- The CMIP6 suite of land models (right) has improved over the CMIP5 suite of land models (left)
- The multi-model mean outperforms any single model for each suite of models
- The multi-model mean CMIP6 land model is the "best model" overall
- Why did CMIP6 land models improve?





(Hoffman et al., in prep)



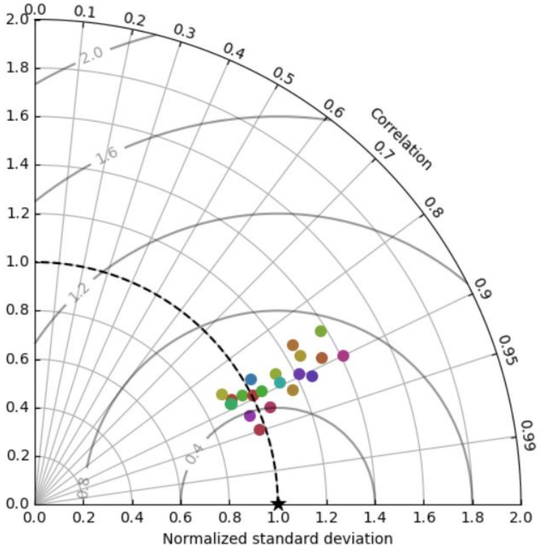
bcc-csm1-1
BCC-CSM2-MR
CanESM2
CanESM5
CESM1-BGC
CESM2
GFDL-ESM2G
GFDL-ESM4
IPSL-CM5A-LR
IPSL-CM6A-LR
MeanCMIP5
MeanCMIP6
−4 −3 −2 −1 0 1 2 3 4
g m−2 d−1

	Download Data	Period Mean (original grids) [Pg yr-1]	Model Period Mean (intersection) [Pg yr-1]	Benchmark Period Mean (intersection) [Pg yr-1]	Model Period Mean (complement) [Pg yr-1]	Benchmark Period Mean (complement) [Pg yr-1]	Bias [g m-2 d-1]	RMSE [g m-2 d-1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Overall Score [1]
Benchmark	[-]	114.												
bcc-csm1-1	[-]	123.	112.	114.	8.79	0.0945	0.238	1.51	1.01	0.484	0.435	0.830	0.955	0.628
BCC-CSM2-MR	[-]	114.	107.	113.	5.88	0.671	-0.0233	1.52	1.11	0.479	0.447	0.817	0.941	0.626
CanESM2	[-]	129.	117.	114.	9.54		0.0601	2.31	2.00	0.388	0.437	0.650	0.836	0.549
CanESM5	[-]	141.	128.	114.	10.1		0.730	1.87	1.60	0.449	0.418	0.710	0.948	0.589
CESM1-BGC	[-]	129.	123.	113.	5.55	0.660	0.379	1.66	1.20	0.426	0.468	0.765	0.889	0.603
CESM2	[-]	110.	104.	113.	5.57	0.642	-0.0542	1.62	1.32	0.458	0.466	0.774	0.933	0.619
GFDL-ESM2G	[-]	167.	152.	114.	12.4		1.26	2.78	1.38	0.377	0.288	0.735	0.897	0.517
GFDL-ESM4	[-]	105.	99.0	114.	6.18		-0.177	1.59	1.49	0.495	0.403	0.702	0.939	0.588
IPSL-CM5A-LR	[-]	165.	150.	113.	11.7	0.515	1.18	2.68	1.20	0.327	0.352	0.781	0.896	0.542
IPSL-CM6A-LR	[-]	115.	109.	113.	5.27	0.708	0.111	1.39	1.14	0.547	0.477	0.790	0.961	0.650
MeanCMIP5	[-]	121.	115.	114.	6.65		0.574	1.41	0.981	0.494	0.502	0.799	0.965	0.652
MeanCMIP6	[-]	116.	110.	114.	6.26		0.129	1.17	0.931	0.572	0.522	0.826	0.956	0.679
MIROC-ESM	[-]	129.	118.	102.	9.04	11.4	0.396	1.90	1.27	0.463	0.435	0.767	0.920	0.604
MIROC-ESM2L	[-]	116.	104.	113.	9.90	0.119	-0.0111	1.95	1.99	0.409	0.379	0.628	0.920	0.543
MPI-ESM-LR	[-]	169.	159.	104.	8.91	9.81	1.36	2.36	1.29	0.402	0.371	0.715	0.930	0.558
MPI-ESM1.2-LR	[-]	141.	133.	104.	6.89	9.81	0.725	2.06	1.13	0.409	0.393	0.769	0.925	0.578
NorESM1-ME	[-]	129.	120.	114.	7.82		0.386	1.86	1.25	0.387	0.456	0.761	0.856	0.583
NorESM2-LM	[-]	107.	97.5	114.	7.59		-0.0828	1.63	1.31	0.443	0.472	0.791	0.938	0.623
UK-HadGEM2-ES	[-]	137.	130.	113.	6.93	0.848	0.602	2.01	1.10	0.389	0.388	0.820	0.855	0.568
UKESM1-0-LL	[-]	126.	119.	113.	7.06	0.825	0.387	1.77	1.16	0.436	0.419	0.791	0.924	0.598

Gross Primary Productivity

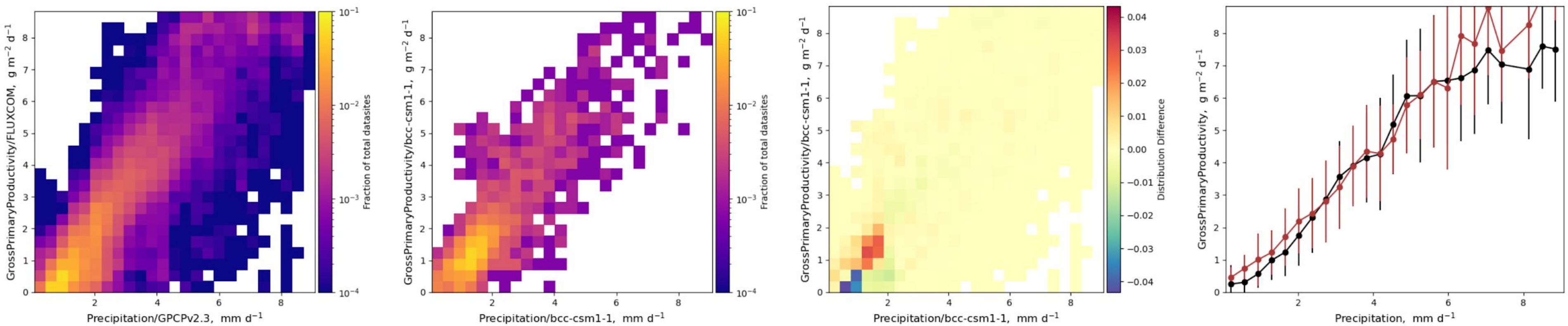
- Multimodel GPP is compared with global seasonal GBAF estimates
- We can see Improvements across generations of models (e.g., CESM1 vs. CESM2, IPSL-CM5A vs. 6A)
- The mean CMIP6 and CMIP5 models perform best

Spatial Taylor Diagram



BERKELEY LAB
NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE
National Laboratory
Sandia National Laboratories
THE UNIVERSITY OF CALIFORNIA IRVINE
MICHIGAN

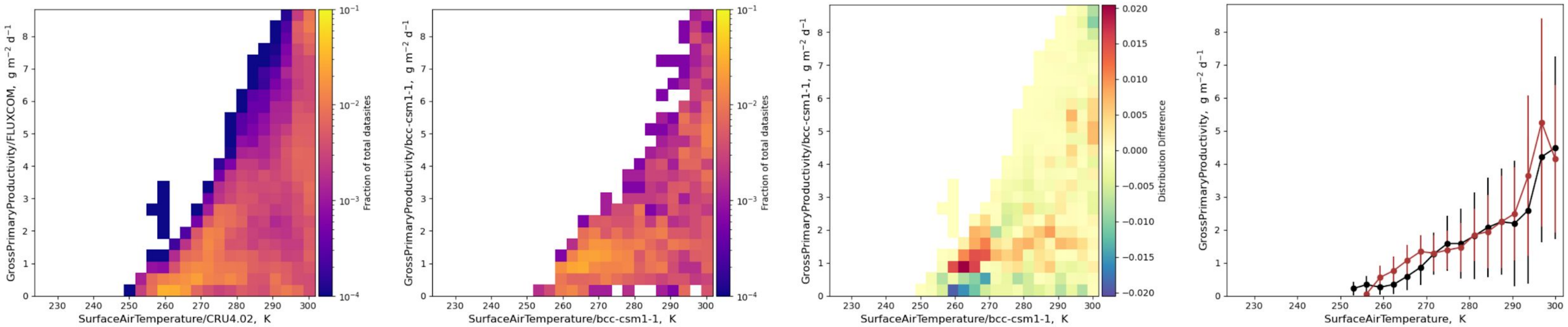
Precipitation/GPCPv2.3



SurfaceDownwardSWRadiation/CERESed4.1

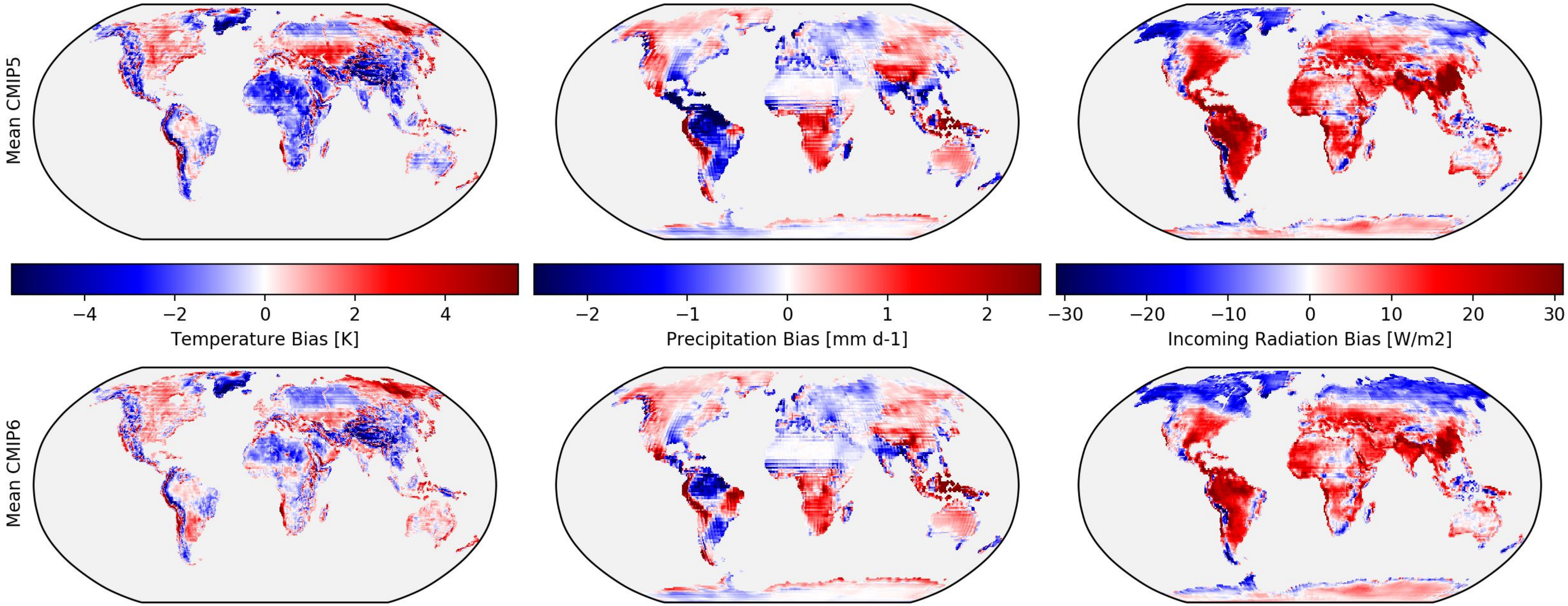
SurfaceNetSWRadiation/CERESed4.1

SurfaceAirTemperature/CRU4.02



Reasons for Land Model Improvements

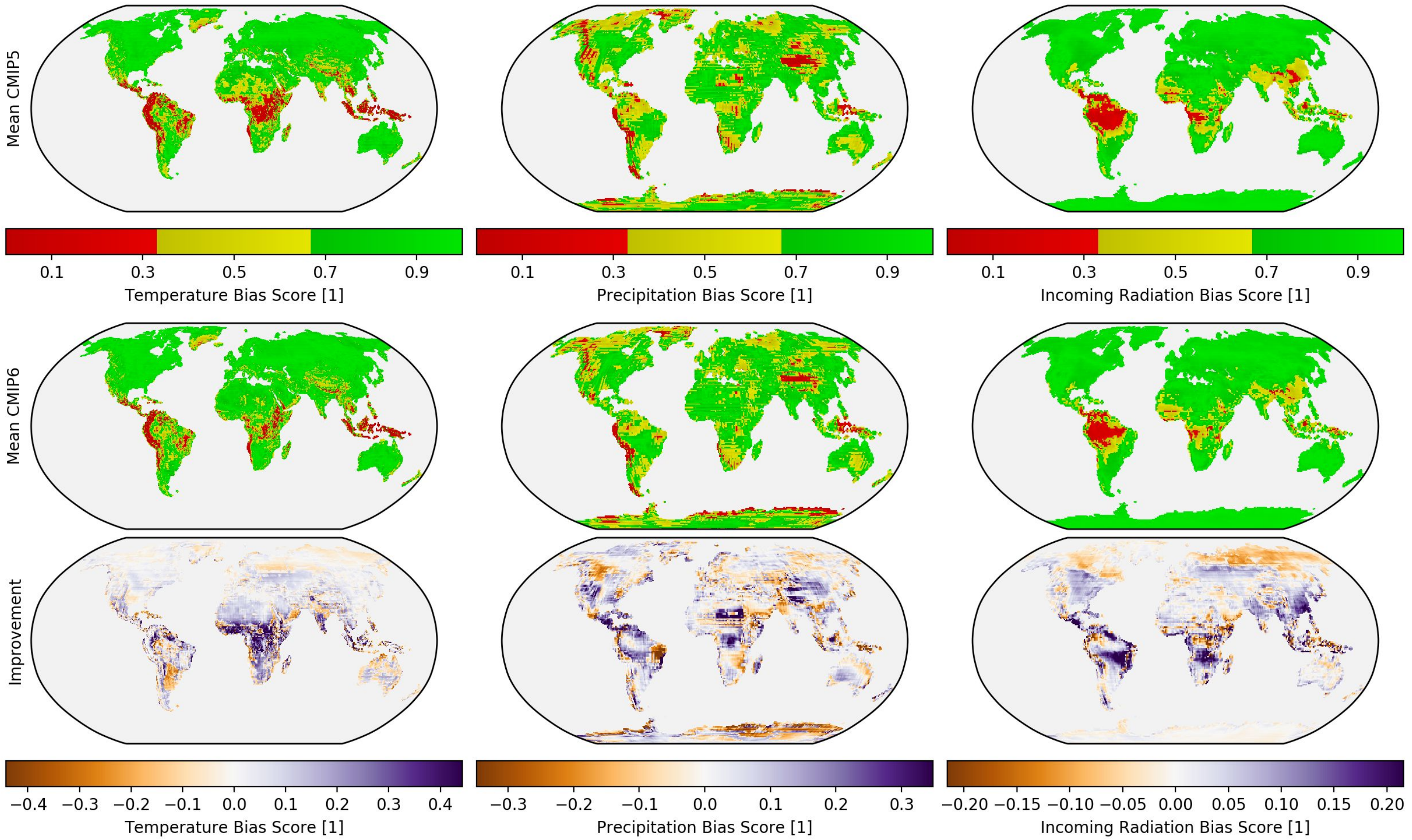
ESM improvements in climate forcings (temperature, precipitation, radiation) likely partially drove improvements exhibited by land carbon cycle models



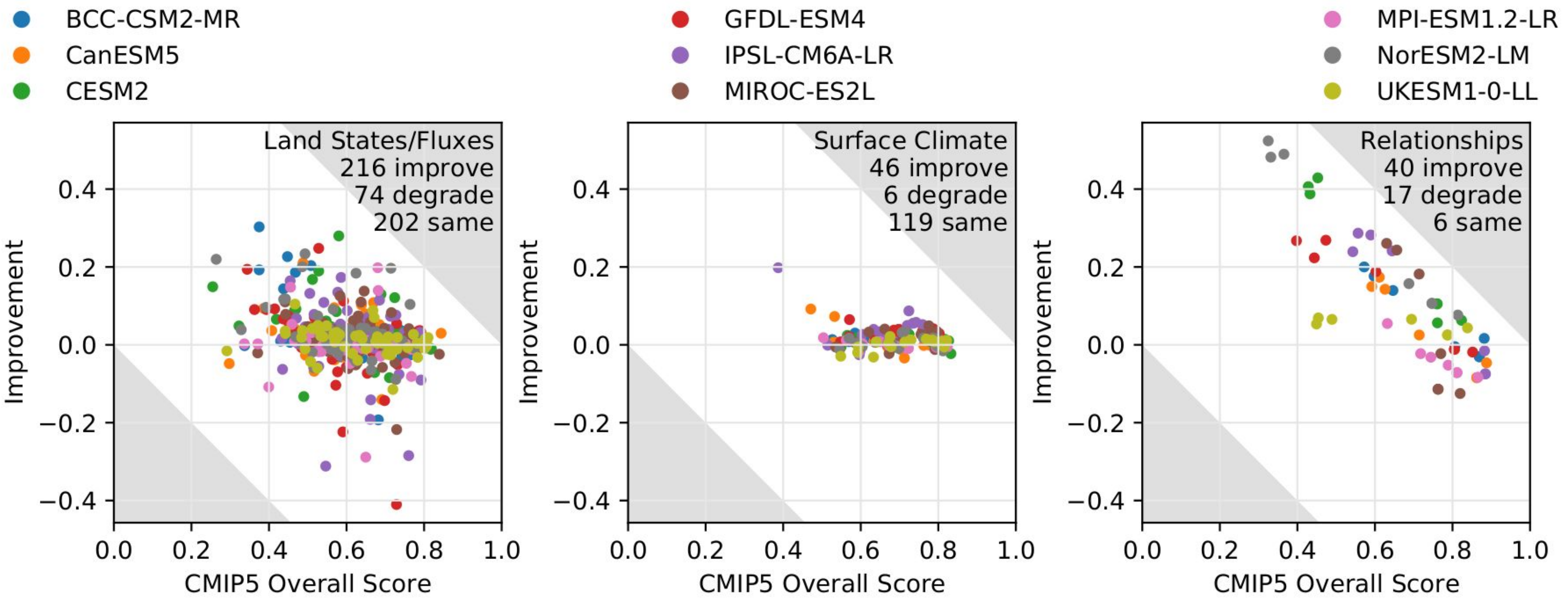
(Hoffman et al., in prep)

Reasons for Land Model Improvements

Differences in bias scores for temperature, precipitation, and incoming radiation were primarily positive, further indicating more realistic climate representation



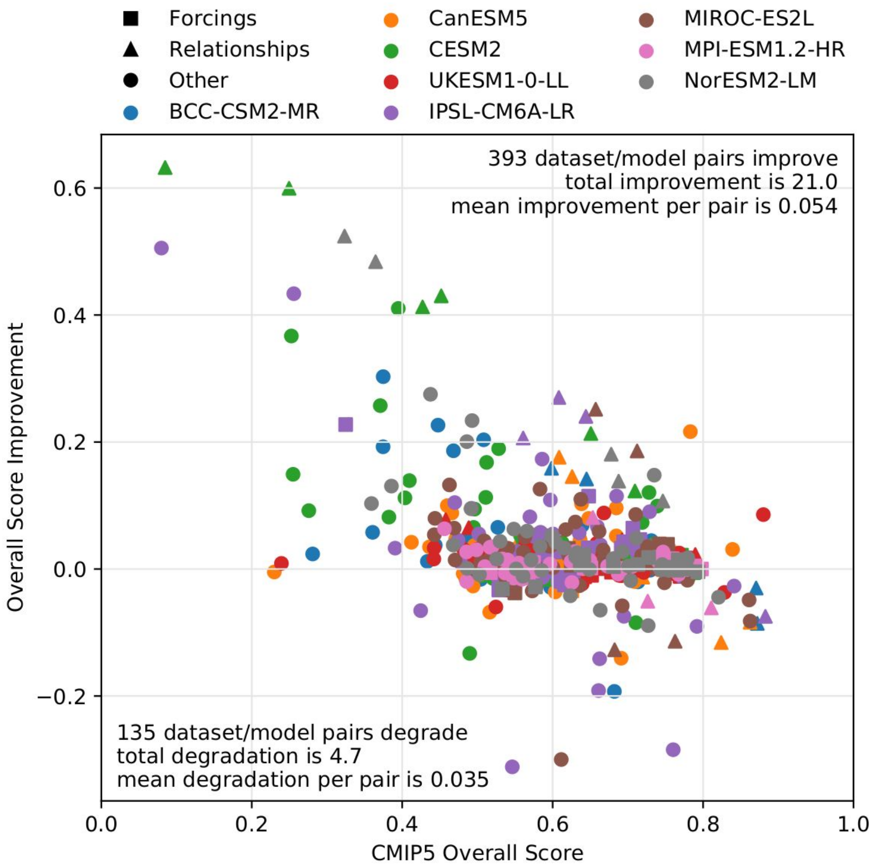
(Hoffman et al., in prep)



Across all land models, scores for most state and flux variables improved (216) or remained nearly the same (202), although some were degraded (74). While atmospheric forcings from CMIP6 ESMs were improved over those from CMIP5 ESMs, the largest improvements were in land model **variable-to-variable relationships**, suggesting that increased land model development was also partially responsible for higher CMIP6 land model scores.

Reasons for Land Model Improvements

While forcings got better, the largest improvements were in **variable-to-variable relationships**, suggesting that increased land model complexity was also partially responsible for higher CMIP6 model scores



ILAMB & IOMB CMIP5 vs 6 Evaluation

- (a) ILAMB and (b) IOMB have been used to evaluate how land and ocean model performance have changed from CMIP5 to CMIP6
- Model fidelity is assessed through comparison of historical simulations with a wide variety of contemporary observational datasets
- The UN's Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) from Working Group 1 (WG1) Chapter 5 contains the full ILAMB/IOMB evaluation as Figure 5.22

	CMIP5 ESMs									CMIP6 ESMs										
	bcc-csm1-1	CanESM2	CESM1-BGC	GFDL-ESM2G	IPSL-CM5A-LR	MIROC-ESM	MPI-ESM-LR	NorESM1-ME	HadGEM2-ES	BCC-CSM2-MR	CanESM5	CESM2	GFDL-ESM4	IPSL-CM6A-LR	MIROC-ES2L	MPI-ESM1.2-LR	NorESM2-LM	UKESM1-0-LL	Mean CMIP5	Mean CMIP6
(a) Land Benchmarking Results																				
Land Ecosystem & Carbon Cycle	-0.72	-0.93	-1.55	-1.51	-0.13	0.60	-0.43	-1.31	0.19	-0.43	0.66	0.48	-1.09	0.22	0.60	-0.07	1.00	0.49	1.63	2.30
Biomass	0.20	-0.45	-1.52	-0.40	-1.26	-0.26	-1.07	-1.77	0.92	1.39	0.74	-0.20	-0.54	0.16	0.93	-0.96	-0.01	1.04	1.23	1.82
Burned Area			-0.87				0.10	-0.83				1.60								
Leaf Area Index	-0.20	-0.64	-1.30	-2.53	-0.01	0.30	0.01	-1.85	-0.16	0.27	0.08	0.34	-0.70	1.19	0.82	0.46	0.37	0.69	1.04	1.81
Soil Carbon	0.27	1.26	-1.46	0.07	0.75	0.47	-0.03	-1.14	0.07	0.23	1.35	-0.99	-2.04	-1.55	0.90	-0.75	-0.17	0.24	1.01	1.48
Gross Primary Productivity	0.59	-1.23	0.01	-1.81	-1.40	0.29	-0.53	-0.24	-1.04	0.77	0.04	0.59	-0.38	1.17	-1.02	-0.37	0.73	0.09	1.51	2.22
Net Ecosystem Exchange	-0.42	-1.81	-0.21	-0.65	1.10	-0.24	0.80	0.02	-1.03	-1.02	-1.19	0.59	1.69	-0.42	0.63	-0.21	1.08	-1.43	1.28	1.43
Ecosystem Respiration	0.90	-0.56	-0.86	-0.24	-1.35	0.99	-0.01	-0.94	-1.54	0.81	0.59	0.51	-0.79	0.90	-0.21	-1.24	0.43	-0.94	1.34	2.21
Carbon Dioxide		-1.54	-0.36	-2.92	-0.74	1.53	-0.00	0.37	0.85		0.42	0.26	0.39	0.59	1.10	-0.87	0.21	0.69	0.09	-0.07
Global Net Carbon Balance		-1.64	-0.88	-1.13	0.17	-0.31	-0.38	-0.50	0.24		-0.23	1.34	-1.70	0.17	-0.74	1.45	1.56	0.26	0.92	1.40
Land Hydrology Cycle	-2.65	-0.42	0.44	-0.18	-0.49	-0.52	-0.57	0.17	0.70	0.15	-0.47	1.51	-1.24	0.58	-0.72	-0.83	0.97	0.87	1.00	1.70
Evapotranspiration	-0.82	-0.99	-0.27	-1.02	0.64	-1.14	-0.62	-0.60	0.28	0.39	-1.08	1.09	0.65	0.43	-1.40	-1.01	0.82	1.05	1.41	2.20
Evaporative Fraction	-0.34	0.74	0.74	-0.14	-0.85	0.21	-1.98	0.22	-0.34	0.10	0.11	1.25	-0.88	1.29	-1.65	-1.81	1.11	-0.06	0.98	1.29
...																				
Terrestrial Water Storage Anomaly	-2.79	-0.45	0.47	0.50	-0.38	0.34	0.35	0.43	0.58	0.15	-0.08	0.95	-2.91	0.43	0.37	0.15	0.39	0.51	0.49	0.50
Permafrost	-0.88	-2.26	0.01	0.13	0.83	0.69	0.56	0.69	-0.56	-0.11	-3.02	0.83	0.74	-0.18	0.49	0.42	0.89	0.43	0.06	0.23
(b) Ocean Benchmarking Results																				
Ocean Ecosystems			2.18	0.20	-0.20		0.04		0.22		-0.37	0.83	-0.37	-0.26	-0.91	-0.67	-1.93	0.27	0.30	0.67
Chlorophyll		-1.50	2.15	0.44	1.02		0.49		0.56		-0.67	0.88	-0.21	0.10	-1.02	-0.41	-2.19	0.18	0.13	0.04
Oxygen, surface			0.73	-0.13	-1.98		-0.53	-1.53	-0.29		0.73	0.34	-0.09	-0.41	0.35	-0.30	0.40	0.49	0.64	1.57
Ocean Nutrients			-0.84	-0.10	0.91		-0.80	-1.25				-0.02	1.00	1.88		-0.90	-1.14	-0.17	-0.16	1.60
Nitrate, surface		0.21	-1.63	0.67	1.22		-0.18	-1.70	0.82		1.21	-0.90	0.29	1.21	1.02	0.39	-1.78	-0.56	-0.47	0.18
Phosphate, surface			-0.69	-0.04	0.04		-0.45	-0.43				0.39	-0.14	0.17	-0.41	-0.98	0.00	0.02	0.88	1.63
Silicate, surface			0.44	-0.71	0.24		-0.81	-0.20	-2.16			0.50	1.24	1.60		-1.21	-0.19	0.18	-0.29	1.37
Ocean Carbon											1.24	-0.23	-0.62	-0.69	-1.08	-1.12	1.31			1.19
TAlk, surface		-0.27	1.01	0.12	0.19		0.32	-2.31	-0.22		0.06	-0.36	0.85	-0.42	0.29	-2.40	1.27	0.06	1.27	0.54
...																				
Salinity, 700m	0.44	-0.35	-1.06	-0.54	0.70	0.46	-0.46	-0.80	0.32	0.36	0.25	-1.16	-0.47	0.54	0.33	-0.39	-0.87	-0.54	1.58	1.64
Ocean Relationships			-1.86	-0.36	-0.29		1.50	-0.43	0.68		-0.02	0.72	1.20	0.17	-1.86	0.02		-1.12	0.39	1.25
Oxygen, surface/WOA2018			0.27	0.23	-0.63		-0.26	-0.12	-0.38		0.29	-0.21	0.19	0.18	0.14	-0.07		0.03	-0.23	0.53
Nitrate, surface/WOA2018		-2.41	-1.38	-0.18	0.06		1.41	-0.16	0.78		0.09	0.79	1.07	0.26	-1.35	0.20		-0.74	0.52	1.04

Relative Scale

Worse Value **Better Value**

Missing Data or Error

ILAMBv3: Ongoing Development

- Continuously updated documentation as new modules are being developed: https://ilamb3.readthedocs.io/

```
In [2]: import matplotlib.pyplot as plt
        import ilamb3
        from ilamb3.analysis import bias_analysis
```

Initialize an analysis, specifying the variable name to be compared.

```
In [3]: analysis = bias_analysis("biomass")
```

Load two ILAMB data products using a built-in catalog

```
In [4]: cat = ilamb3.ilamb_catalog()
        ds_esacci = cat["biomass | ESACCI"].read()
        ds_xu = cat["biomass | XuSaatchi2021"].read()
```

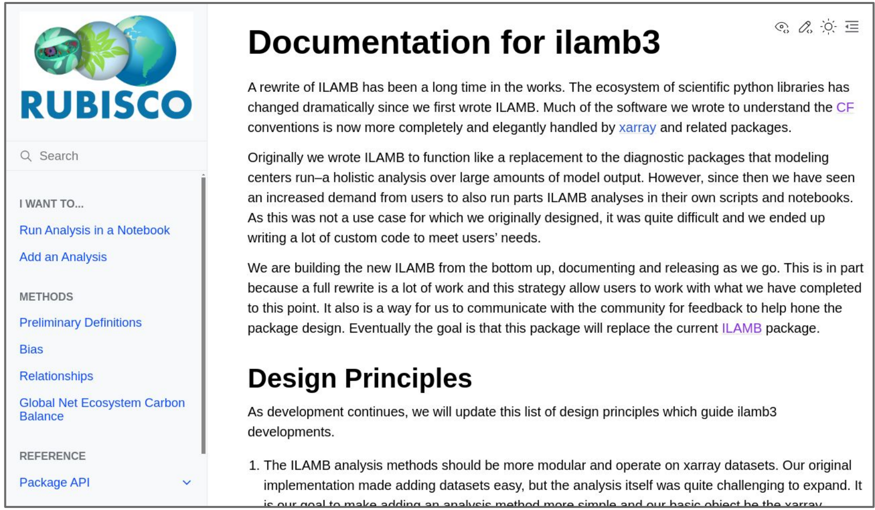
Apply the analysis using the ESACCI product as a reference.

```
In [5]: df,ds_esacci,ds_xu = analysis(ds_esacci,ds_xu)

In [6]: df
```

Out[6]:

	source	region	analysis	name	type	units	value
0	Reference	None	Bias	Period Mean	scalar	Mg ha-1	24.019928
1	Comparison	None	Bias	Period Mean	scalar	Mg ha-1	25.676543
2	Comparison	None	Bias	Bias	scalar	Mg ha-1	-16.670597
3	Comparison	None	Bias	Bias Score	score	1	0.398491


RUBISCO

Search

I WANT TO...

Run Analysis in a Notebook

Add an Analysis

METHODS

Preliminary Definitions

Bias

Relationships

Global Net Ecosystem Carbon Balance

REFERENCE

Package API

Documentation for ilamb3

A rewrite of ILAMB has been a long time in the works. The ecosystem of scientific python libraries has changed dramatically since we first wrote ILAMB. Much of the software we wrote to understand the CF conventions is now more completely and elegantly handled by xarray and related packages.

Originally we wrote ILAMB to function like a replacement to the diagnostic packages that modeling centers run–a holistic analysis over large amounts of model output. However, since then we have seen an increased demand from users to also run parts ILAMB analyses in their own scripts and notebooks. As this was not a use case for which we originally designed, it was quite difficult and we ended up writing a lot of custom code to meet users' needs.

We are building the new ILAMB from the bottom up, documenting and releasing as we go. This is in part because a full rewrite is a lot of work and this strategy allow users to work with what we have completed to this point. It also is a way for us to communicate with the community for feedback to help hone the package design. Eventually the goal is that this package will replace the current ILAMB package.

Design Principles

As development continues, we will update this list of design principles which guide ilamb3 developments.

1. The ILAMB analysis methods should be more modular and operate on xarray datasets. Our original implementation made adding datasets easy, but the analysis itself was quite challenging to expand. It

- ILAMBv3 allows for analysis methods to be imported into Python scripts and Jupyter notebooks and used to produce the scalars and plots synthesized in the full analysis
- At left, the ILAMB bias analysis is applied to two reference data products and show a table of scalar values

Thank You!