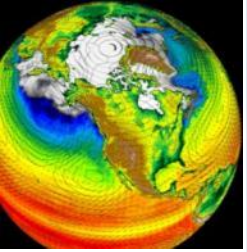
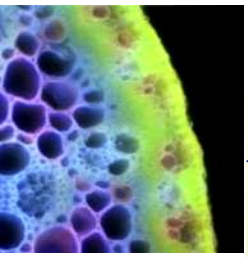
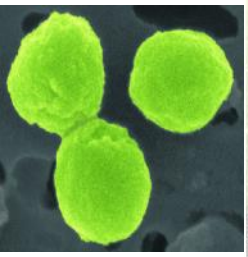


Systematic Evaluation of Land Surface Models Using the International Land Model Benchmarking (ILAMB) Package

Forrest M. Hoffman and Nathaniel O. Collier
Oak Ridge National Laboratory

May 18, 2017
4th International Conference of Hydrology delivers Earth System Sciences to Society (HESSS4)
May 16–19, 2017
University of Tokyo, Tokyo, Japan

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

Workshop Organizers: Renu Joseph and Dorothy Koch
Workshop Co-Chairs: Forrest M. Hoffman (ORNL), William J. Riley (LBNL), James T. Randerson (UCI), Gretchen Keppel-Aleks (UMich), and David M. Lawrence (NCAR)

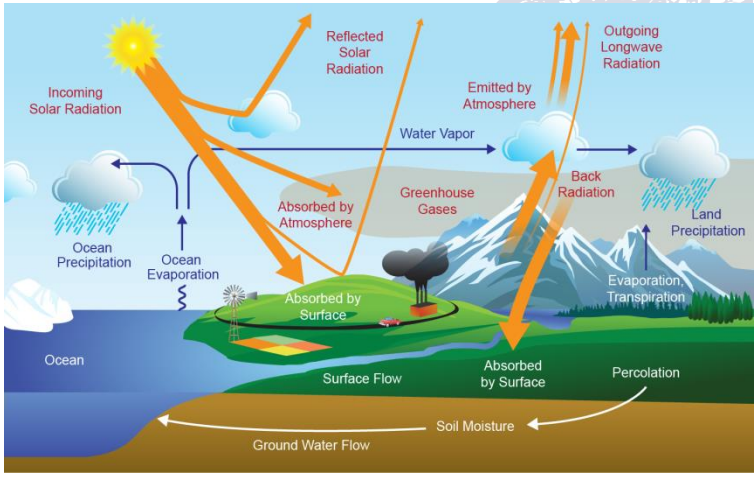
U.S. DEPARTMENT OF ENERGY
Office of Science
Office of Biological and Environmental Research

International Land Model Benchmarking (ILAMB)

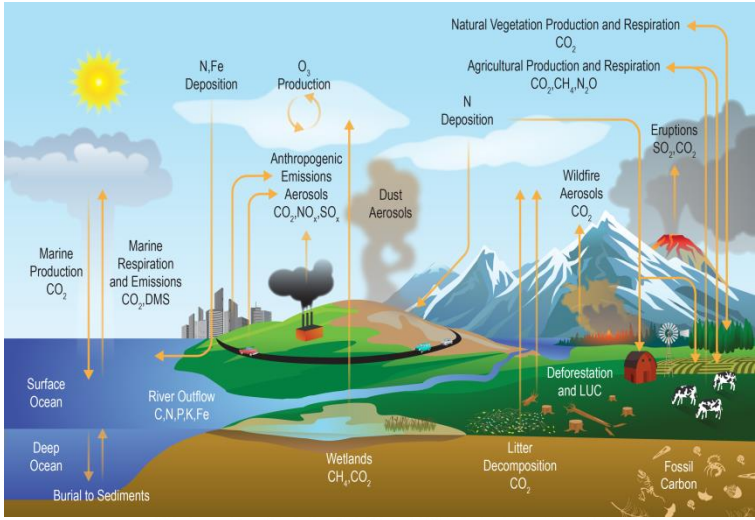
What is ILAMB?

A community coordination activity created to:

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

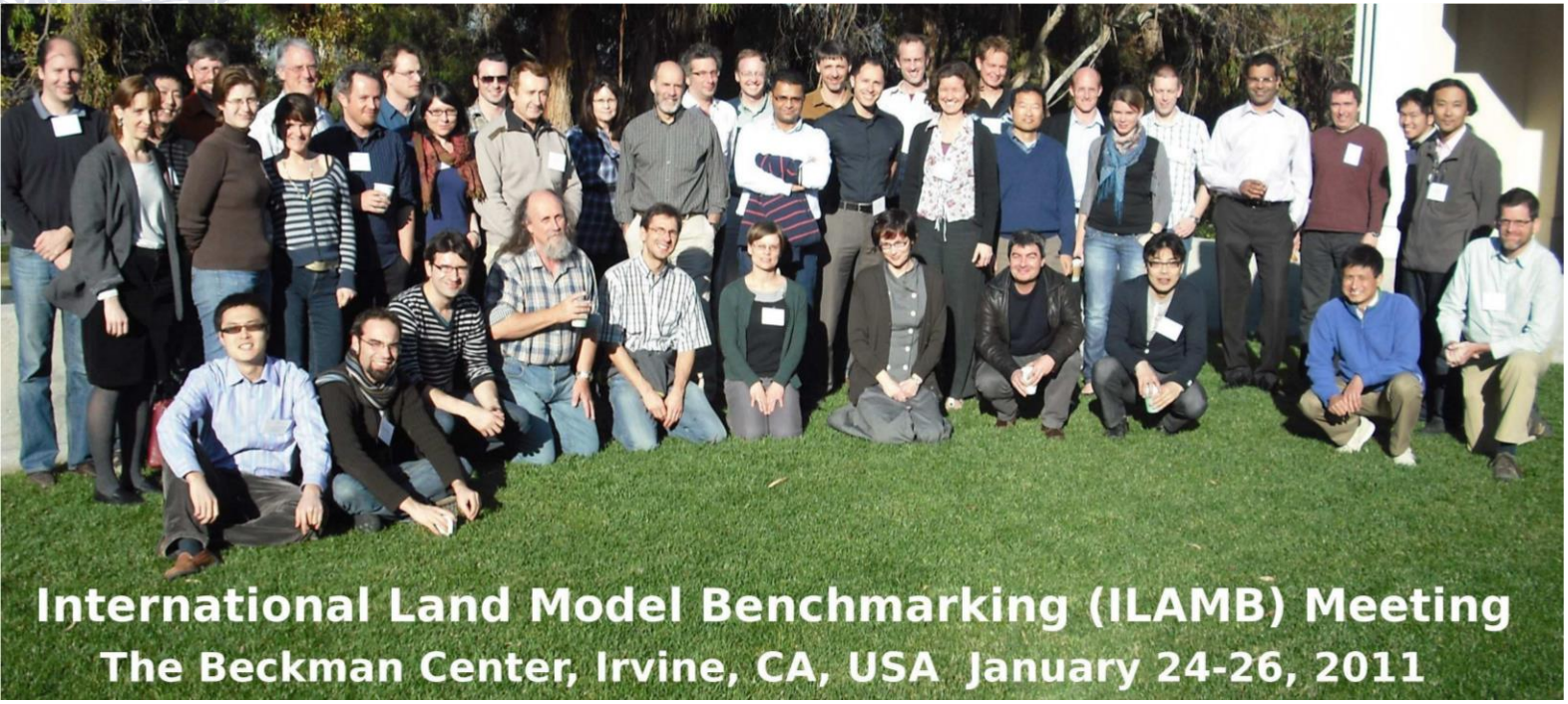


Energy and Water Cycles



Carbon and Biogeochemical Cycles

First US ILAMB Workshop, January 2011



- **BER co-sponsored** the first ILAMB Workshop in the US in 2011.
- ~45 researchers from the US, Canada, UK, Netherlands, France, Germany, Switzerland, China, Japan, and Australia participated.
- **Priority outcomes:** Develop internationally accepted benchmarks for model performance and design an open source software system (Luo et al., 2012).

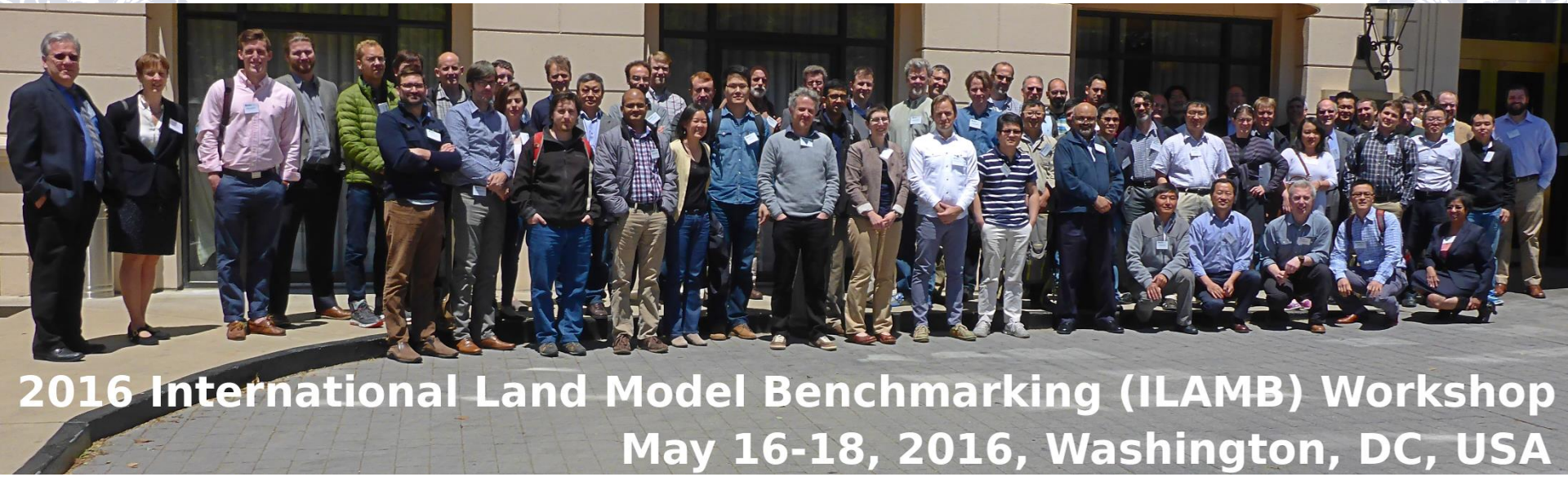
2015 AGU Fall Meeting ILAMB Town Hall Meeting

- ILAMBv1 package (Mu et al., in prep) demonstrated and released
- 70–80 attendees offered questions, ideas, and suggestions:
  - Site-level analysis (see new FLUXNET data release)
  - Consider multiple equally valid data sets
  - Need quantified uncertainties in observations
  - Include data producers in metrics development
  - Global synthesis evaluation from variety of MIPs
  - NASA Permafrost Benchmarking System (PBS) and other projects could leverage ILAMB framework
  - Important future development – perturbation experiments:
    - Ratios of related states and fluxes (e.g., NPP/precipitation)
    - Manipulative experiments (e.g., N, P fertilization, FACE, warming)
    - Natural "experiments" or extremes (e.g., drought, floods, heat waves)
  - Connect with uncertainty quantification frameworks (e.g., PEcAn)

AGU FALL MEETING

San Francisco | 14–18 December 2015

Second US ILAMB Workshop, May 16–18, 2016



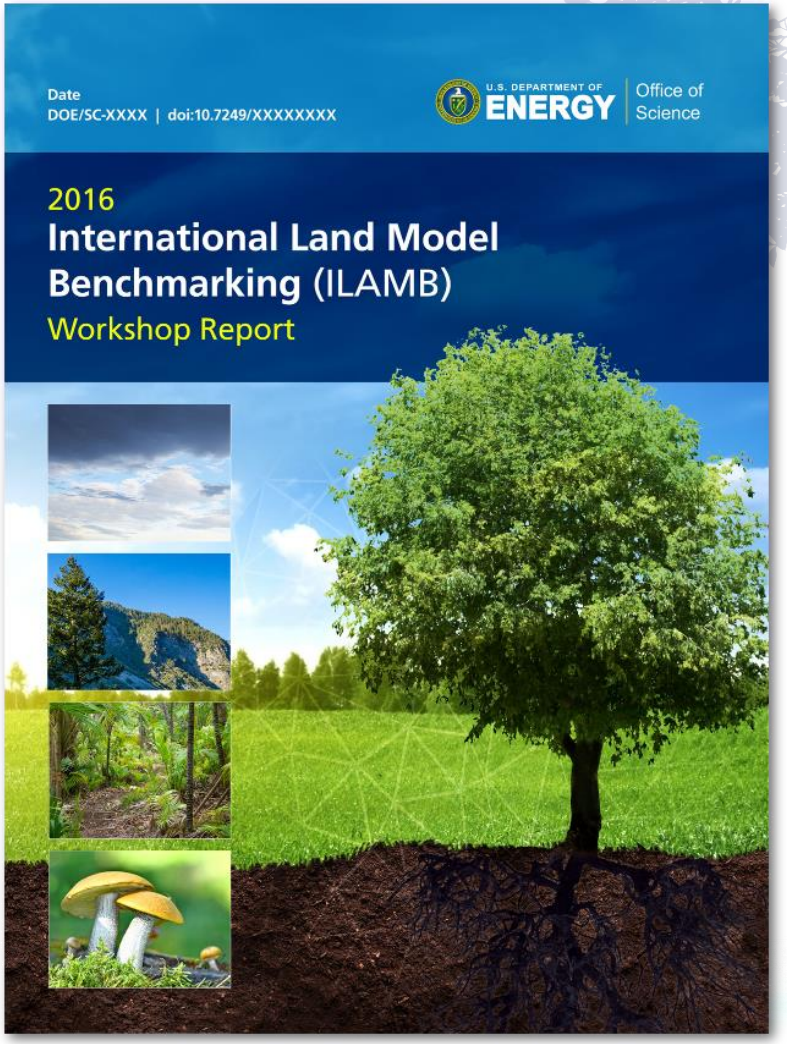
Overarching Workshop Goals

Engage the research community in defining scientific priorities for

- Design of new metrics for model benchmarking
- Model Intercomparison Project (MIP) evaluation needs
- Model development, testbeds, and workflow practices
- Observational data sets and needed measurements

Second US ILAMB Workshop, May 16–18, 2016

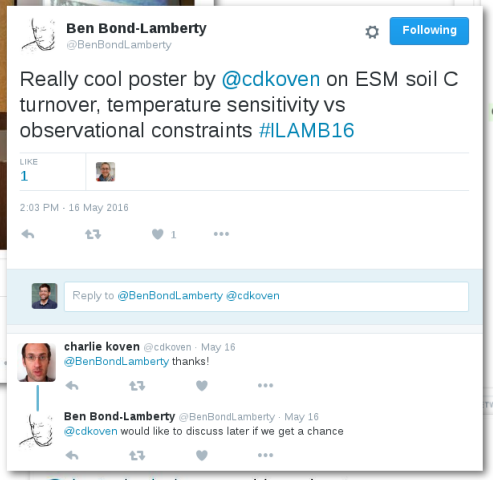
- **60+ participants** from Australia, Japan, China, Germany, Sweden, Netherlands, UK, and US
- **10 modeling centers** represented
- **~25 online attendees** at any time
- Report highlights
  - current state of the science
  - challenges and opportunities for benchmarking
  - model development needs
  - field and laboratory measurement priorities

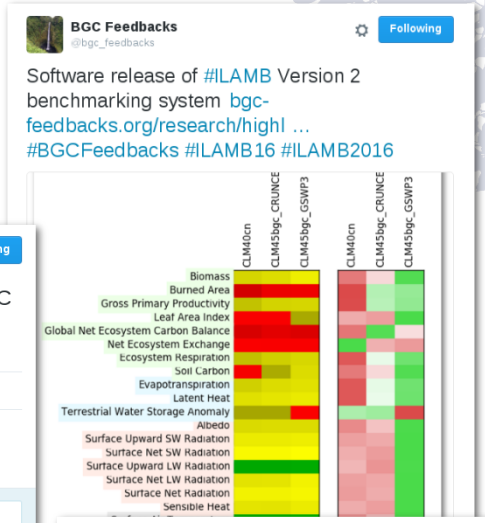


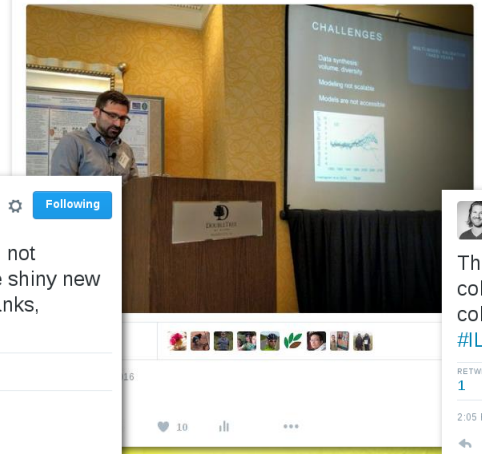
Hoffman, F. M., C. D. Koven, G. Keppel-Aleks, D. M. Lawrence, W. J. Riley, J. T. Randerson, A. Ahlström, G. Abramowitz, D. D. Baldocchi, M. J. Best, B. Bond-Lamberty, M. G. De Kauwe, A. S. Denning, A. Desai, V. Eyring, J. B. Fisher, R. A. Fisher, P. J. Gleckler, M. Huang, G. Hugelius, A. K. Jain, N. Y. Kiang, H. Kim, R. D. Koster, S. V. Kumar, H. Li, Y. Luo, J. Mao, N. G. McDowell, U. Mishra, P. R. Moorcroft, G. S. H. Pau, D. M. Ricciuto, K. Schaefer, C. R. Schwalm, S. P. Serbin, E. Shevliakova, A. G. Slater, J. Tang, M. Williams, J. Xia, C. Xu, R. Joseph, and D. Koch (2016), *International Land Model Benchmarking (ILAMB) 2016 Workshop Report*, DOE/SC-0186, U.S. Department of Energy, Office of Science, Germantown, Maryland, USA, doi:10.2172/1330803.

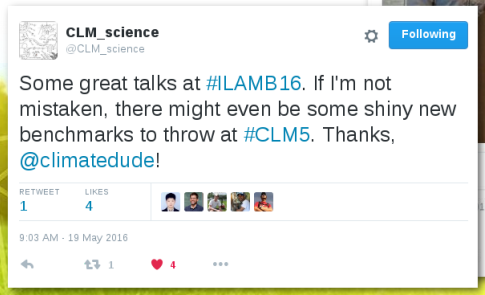
Crowdsourcing and Social Media at the Workshop

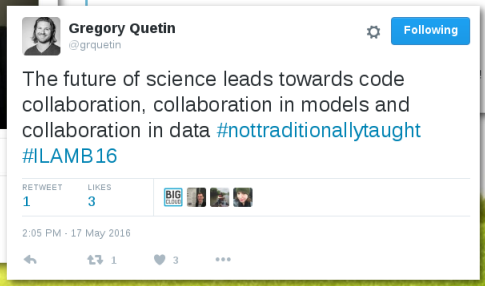
- Videoconferencing used for all plenary sessions
- All slides and meeting notes crowdsourced in Google Slides and Google Docs
- Twitter used for ideas, comments & questions
- White papers written and reviewed through crowdsourcing
- Technology reduced gender, racial, and cultural imbalances and travel costs and emissions











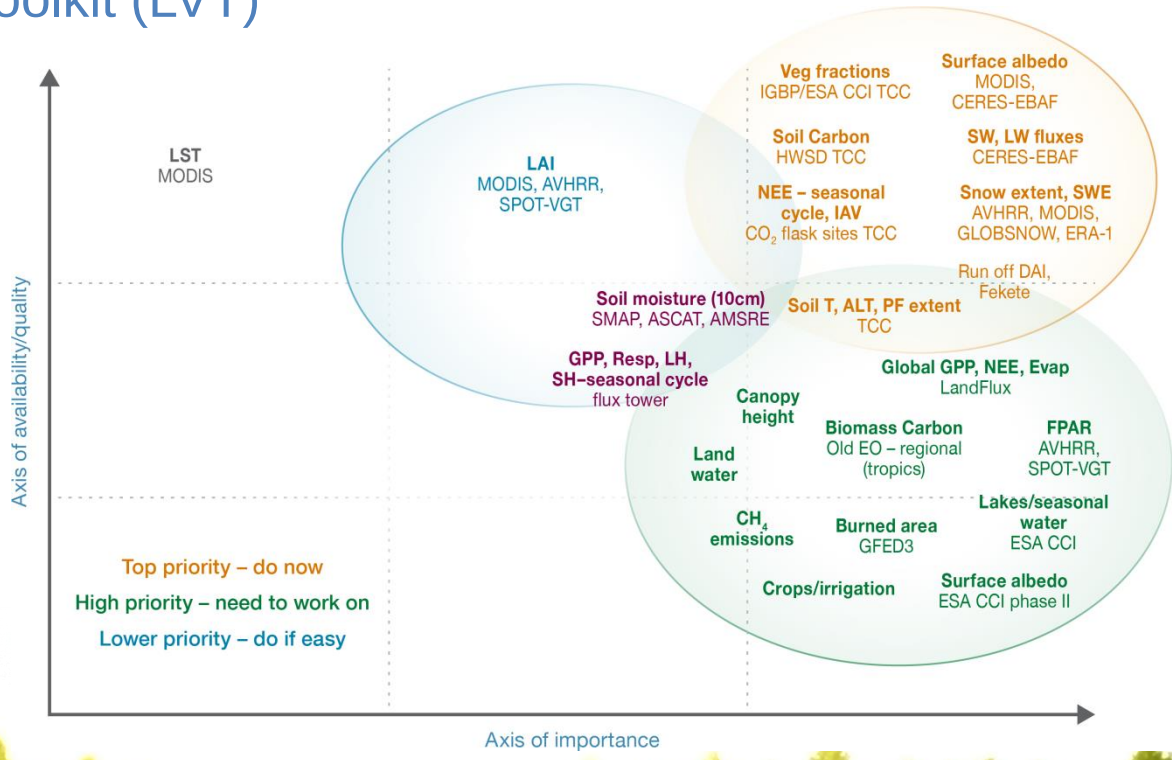
White Paper Synthesis and Workshop Outcomes

Benchmarking Tools

- **Existing model evaluation and benchmarking tools** assessed:
  - Protocol for the Analysis for Land Surface models (PALS)
  - Program for Climate Model Diagnosis and Intercomparison (PCMDI) Metrics Package (PMP)
  - Earth System Model Evaluation Tool (ESMValTool)
  - Land surface Verification Toolkit (LVT)

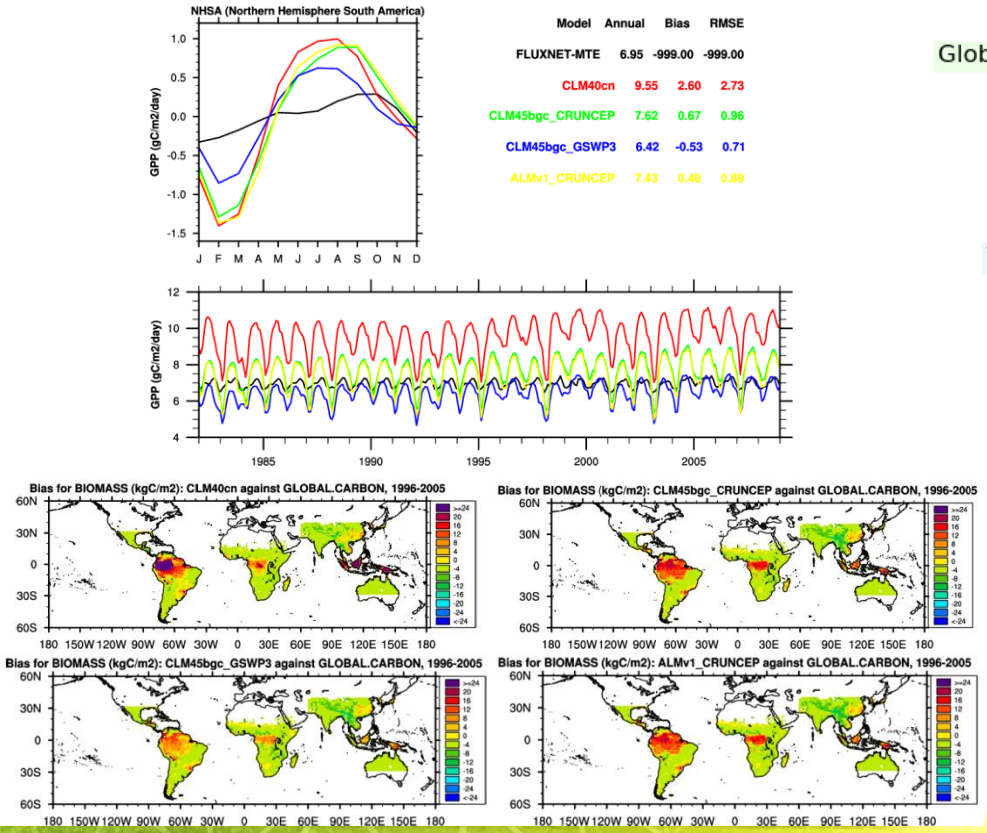
Benchmarking systems should

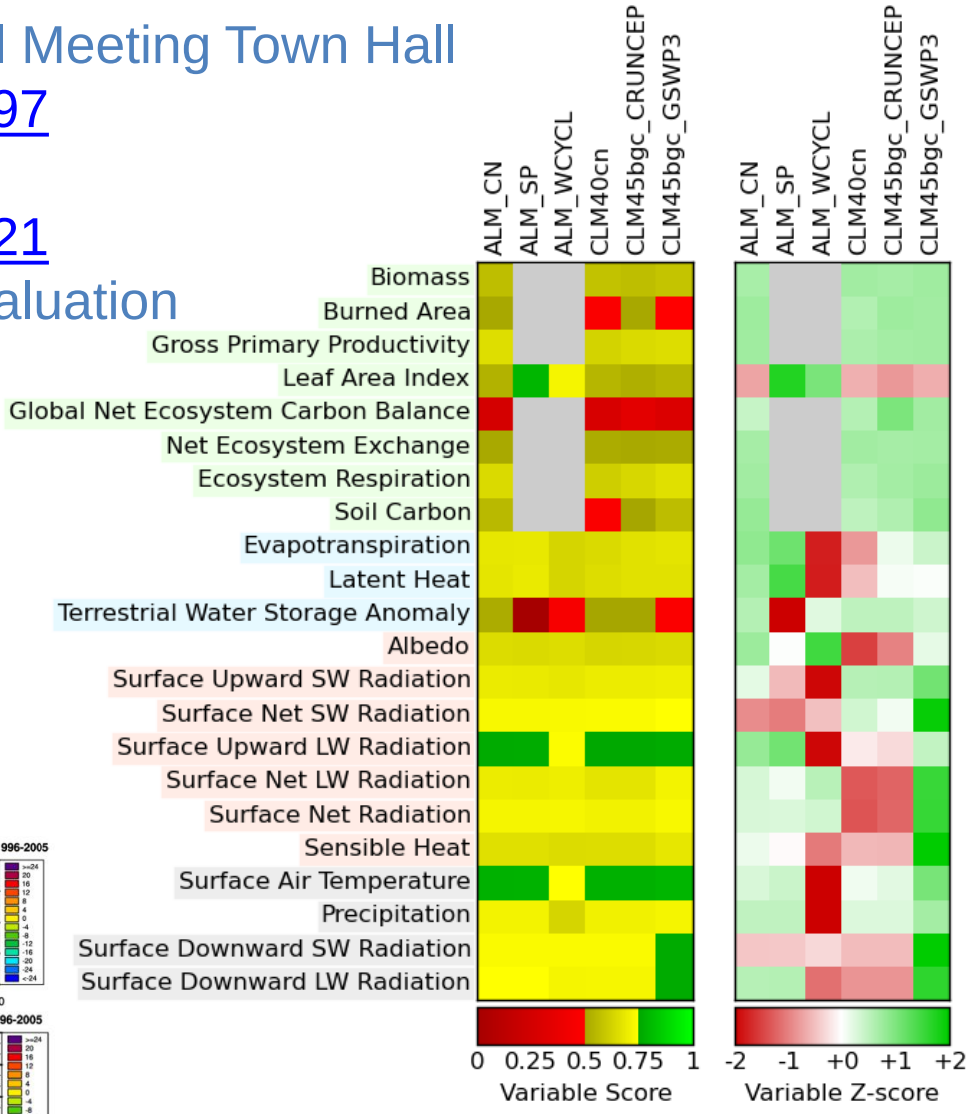
- Test predictive power of models under changing climate
- Span a wide range of spatial and temporal scales & extents
- Be open source to leverage work of many teams and minimize redundancy
- Be integrated with data repositories and archives



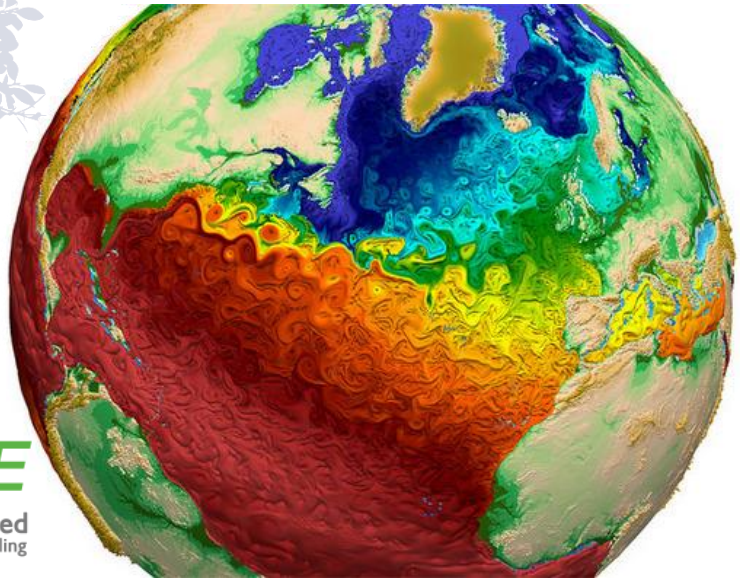
Current Status of the ILAMB Packages

- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall doi:10.18139/ILAMB.v001.00/1251597
- **ILAMBv2** released at this workshop doi:10.18139/ILAMB.v002.00/1251621
- Being used for ACME and CESM evaluation

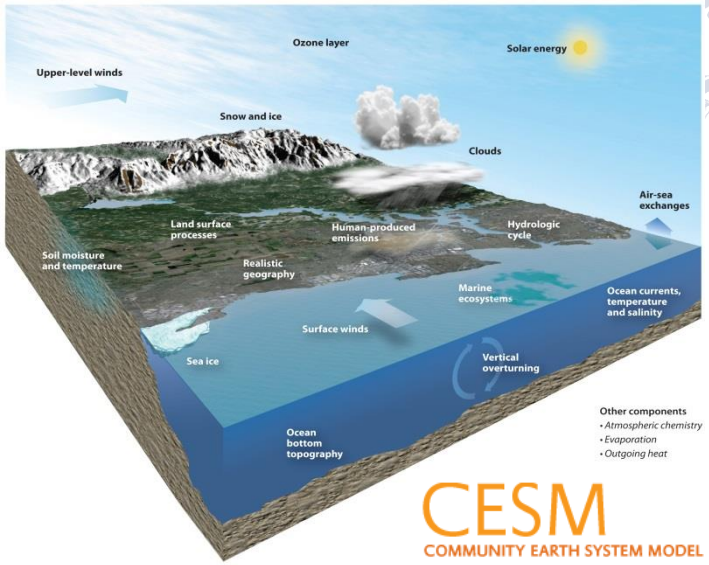




Advancing Model Benchmarking



ACME
Accelerated Climate Modeling for Energy

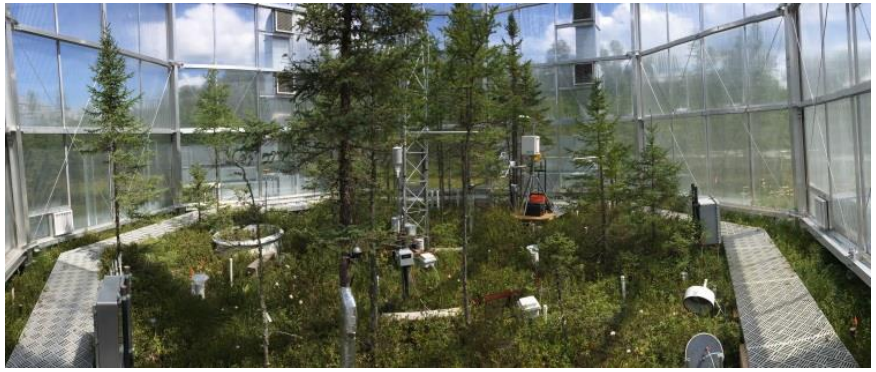


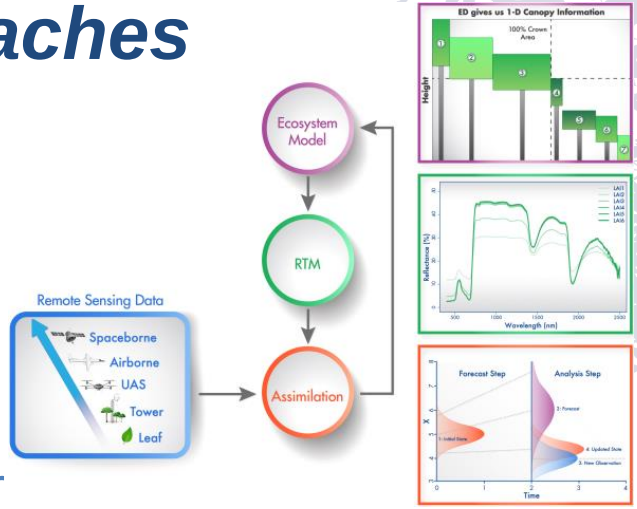
CESM
COMMUNITY EARTH SYSTEM MODEL

Developing metrics that make appropriate use of observational data remains **a scientific challenge** because of

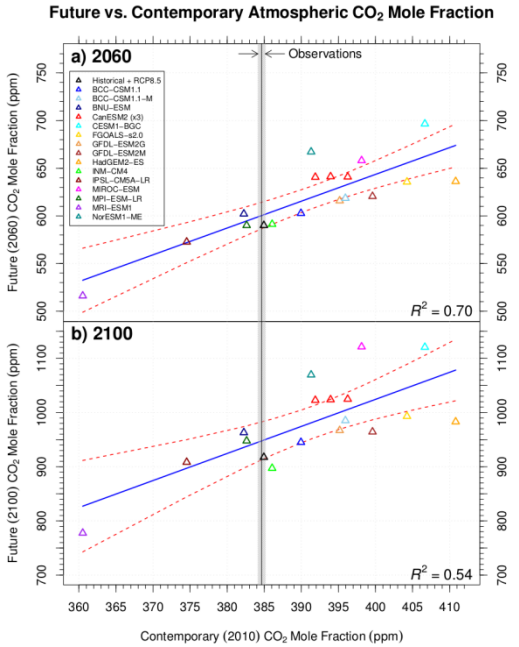
- **spatial and temporal mismatch** between models and measurements,
- **poorly characterized uncertainties** in observational data products,
- **biases** in reanalysis and forcing data,
- **model simplifications**, and
- structural and parametric **uncertainties**.

Benchmarking Approaches



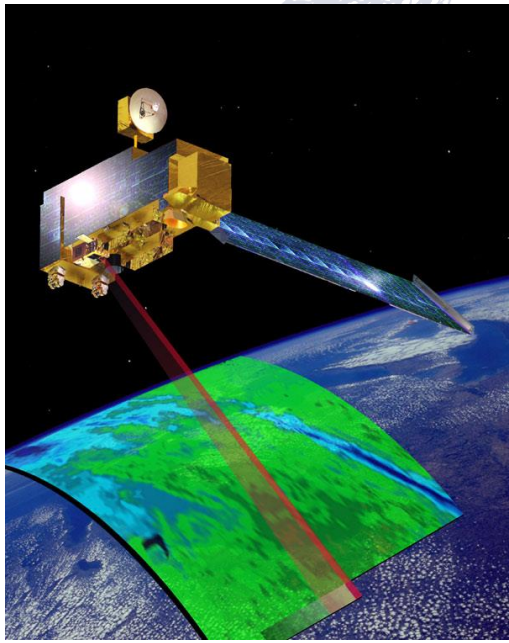


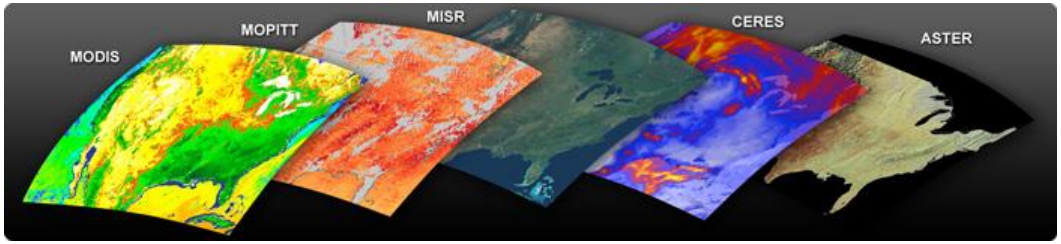
- Statistical comparisons (bias, root-mean-square er (RMSE), phase, amplitude, spatial distribution, Taylor diagrams and scores)
- Functional response metrics or variable-to-variable comparisons
- Emergent constraints
- Reduced complexity models and traceability analysis
- Formal uncertainty quantification methods
- Meta-analyses of perturbation experiments



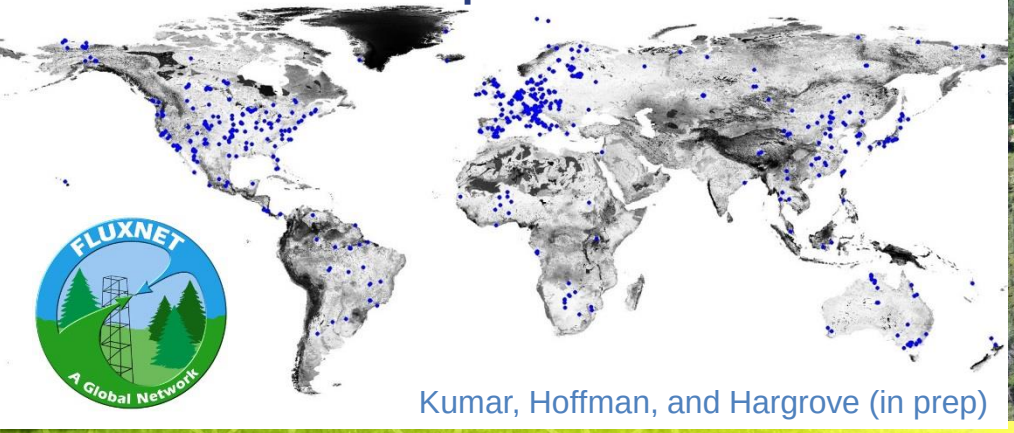
Benchmarking Challenges and Priorities

- **Super site benchmarks** for AmeriFlux and FLUXNET
- **Benchmarks for soil carbon** turnover, distribution, transport
- **Metrics for extreme events** & response of ecosystems
- **Data for vegetation** recruitment, growth, mortality, phenology, canopy structure
- Benchmarks for critical **high latitude & tropical ecosystems**
- Leverage **field projects & remote sensing methods**

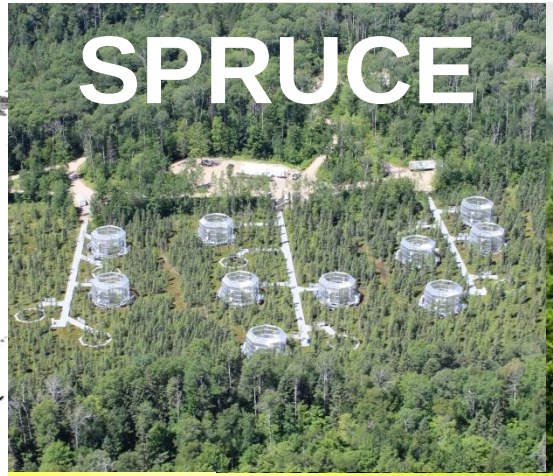




FLUXNET Representativeness

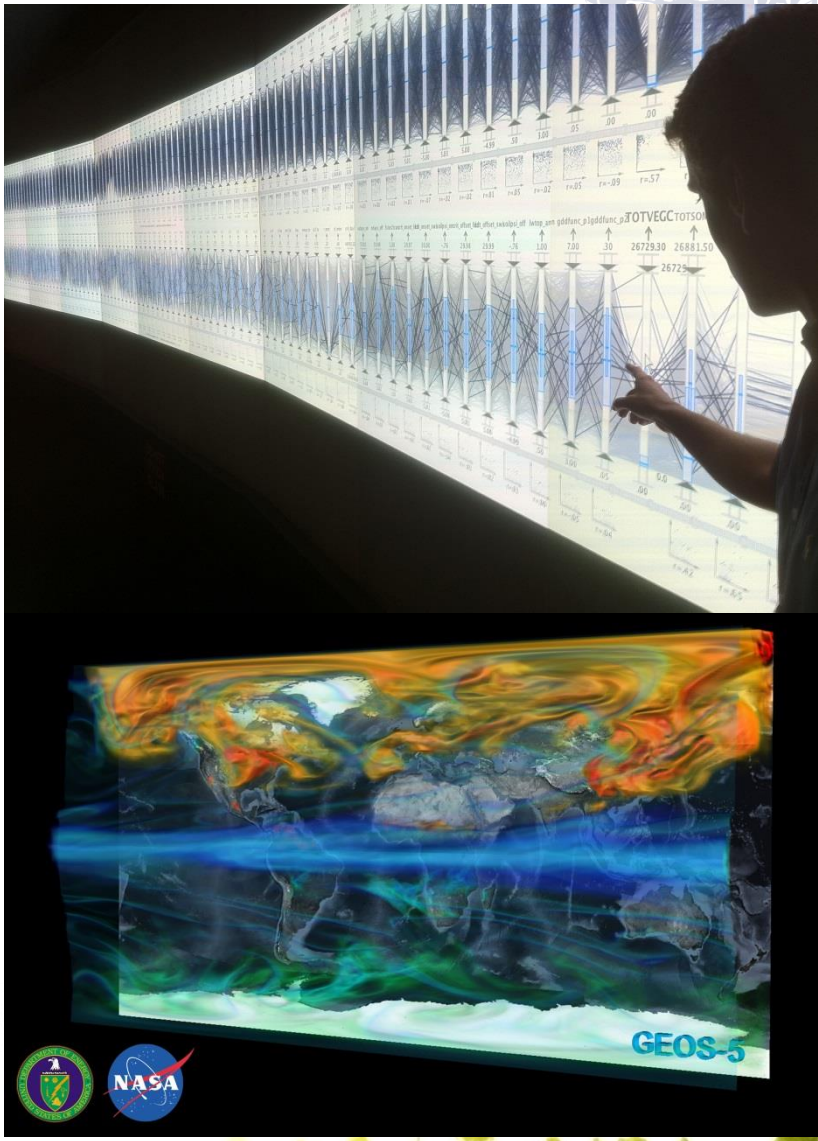


Kumar, Hoffman, and Hargrove (in prep)



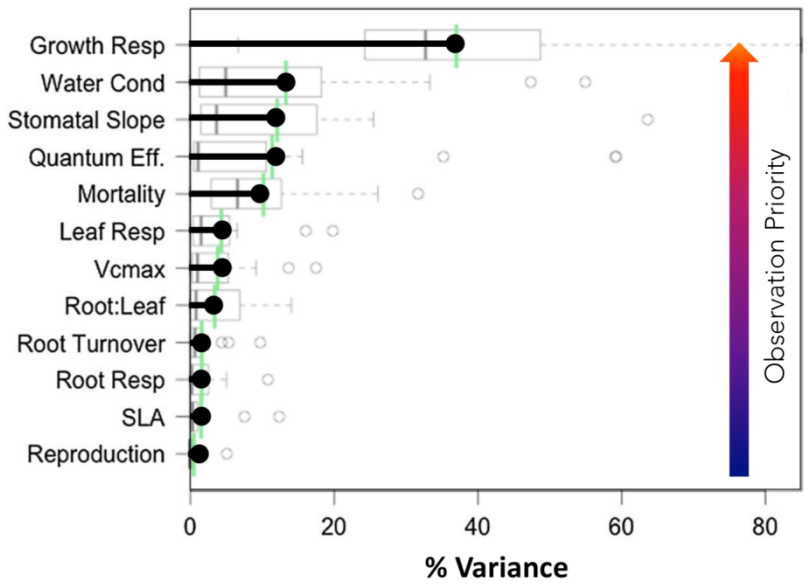
Model Development and Evaluation Testbeds

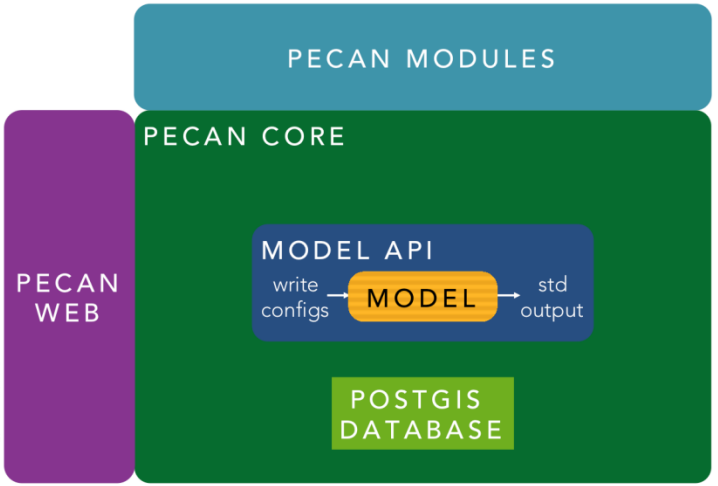
- New development required for improved process representation, additional model outputs, and in situ diagnostics
- Land model testbed (LMT) needed for execution, calibration, and evaluation of alternative model formulations
- Initial LMT should implement AmeriFlux and FLUXNET “super sites” for offline point simulations
- LMT capability should be incorporated into routine model development testing (e.g., nightly/weekly automated testing)



Uncertainty Quantification (UQ) Frameworks

- Integrate and report carbon cycle model diagnostics as a matrix of flows and turnover times to attribute responses to specific ecosystem components
- Apply Bayesian UQ approaches that utilize leadership class computing facilities to identify model uncertainties
- Use UQ results to guide data collection activities and target process representation improvements
- Standardize collection, processing, archival, and distribution of observational data
- Investigate integration of UQ frameworks with ILAMB





PEcAn
pecanproject.org

BETYdb.org

Computational Needs and Requirements

- Scalable algorithms and machine learning methods should be developed
- Research organizations need cyber infrastructure to support large scale data, including model–data comparisons and online data assimilation
- Scientific computing facilities should strike a balance between resources for compute-intensive vs. data-intensive applications
- New development for ILAMB should include improved support for remote retrieval and version tracking for observational data (e.g., Obs4MIPs)

Conclusion and Next Steps

- **2016 ILAMB Workshop** successfully brought together the international community to identify scientific challenges and priorities for future research
- To address *Major Processes* and *Integrating and Cross-Cutting Themes*, **small targeted working groups** should be formed to research and publish
- A top priority is supporting **CMIP6 activities** with additional ILAMB development for automated analysis and model–data intercomparison
- We want to engage more of the **modeling and MIP, observational, and remote sensing communities** in the process of identifying model weaknesses and informing future measurement campaigns
- We envision ILAMB as a core capability for a **research institute** to provide:
  - Home for synthesis working groups
  - Host MIP-related activities
  - Support expanded use of Earth system models

Thank you!