
U.S. DEPARTMENT OF ENERGY | Office of Science

# CESD Cyberinfrastructure Working Groups

## Environmental System Science (ESS) PI Meeting

Bolger Center, Potomac, Maryland, USA

April 30, 2018

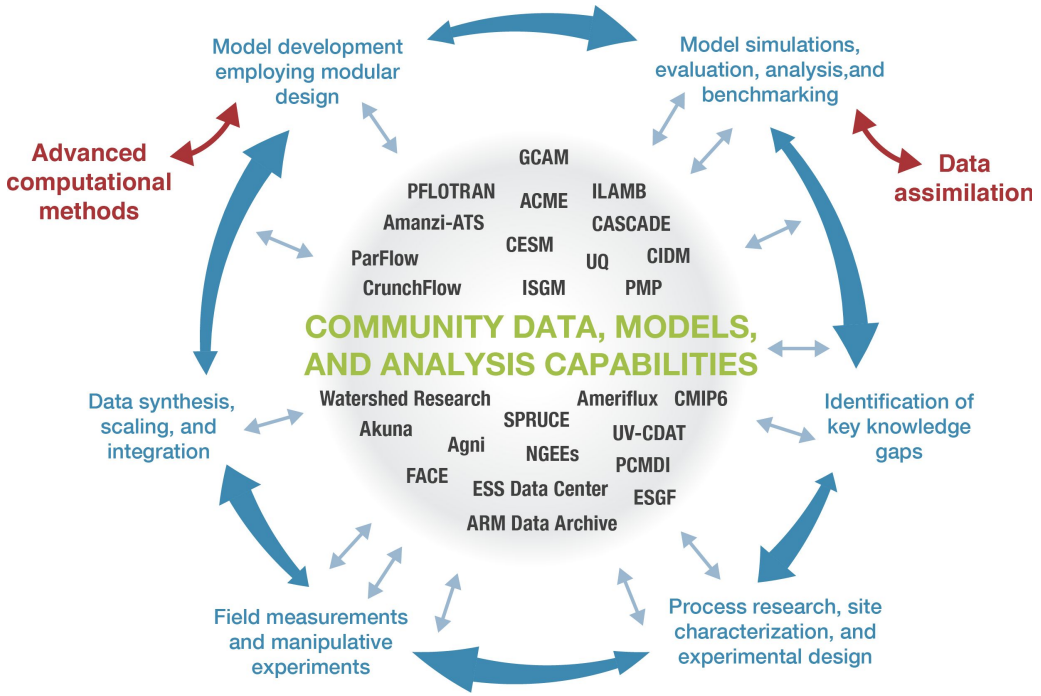
# *Model–Data Integration*

**Leads: Forrest M. Hoffman (ORNL) and Xingyuan Chen (PNNL)**

**Participating Team Members: Bhavna Arora, Ben Bond-Lamberty, Eoin Brodie, Laura Condon, Beth Drewniak, Moayi Huang, Colleen Iversen, Elchin Jafarov, Jitu Kumar, Umakant Mishra, Bill Riley, Joel Rowland, Tim Scheibe, Shawn Serbin, Xiaoying Shi, Peter Thornton, and Anthony Walker**

# Model–Data Integration Scope

- Model–data comparison
- Uncertainty quantification (UQ) & data assimilation (DA)
- Management of model results and observational data
- Geospatial and remote sensing data analysis
- Data analytics methods and techniques, e.g.,
  - –Data mining
  - –Neural networks
  - –Genetic algorithms
  - –Other machine learning techniques
  - –Visual analytics
- Model–data fusion



# Short-Term Goals (2016–2018)

- **Encourage archiving and versioning of publications, data, models, and software tools**
  - Document best practices jointly with other Working Groups
  - Versioning for synthesized & combined data sets (e.g., FLUXNET2015)
  - Digital Object Identifiers (DOIs) for pubs, data, models, and tools
- **Identify available scientific workflows, UQ frameworks, and model–data tools (e.g., ESGF, UV-CDAT, PEcAn, ILAMB)**
  - What workflows are people using and when does one assign a DOI?
  - Develop a user survey to capture initial information
- **Initiate subgroup on geospatial analysis and remote sensing**
  - Google Earth Engine and similar useful tools are rapidly evolving
  - Identify tools and resources for geospatial data analytics
  - Individual community projects have pockets of expertise (e.g., ARM)
- **Advocate for open and standard data formats & conventions**
  - Engage in groups to develop standards and educate users
  - Deploy tools/APIs to transform observational data into model formats
  - Foster API consistency across multi-agency/federated data centers

# Short-Term Goals (2016–2018) (continued)

- **Support community activities to make observational data quickly and easily available for model evaluation (e.g., ILAMB)**
  - Sponsor working groups focused on individual data sets and corresponding model metrics
  - Make AmeriFlux, NGEE Arctic, NGEE Tropics, SPRUCE, FACE, and similar data sets rapidly available to modelers by creating benchmarks
- **Organize disparate uncertainty quantification (UQ) activities to foster collaboration and establish best practices**
  - Standardize methods and approaches
  - Create workflows for common modeling frameworks

# Progress Since 2016

- **Geospatial analysis and remote sensing**
  - 2017 whitepaper : **Geospatial Science to Inform Land Surface Models** (Mishra, Serbin, Wainwright, Kumar, Huang, and Chen)
- **Model–data comparison and benchmarking**
  - International Land Model Benchmarking (ILAMB) Workshop and Tools (described by Hoffman later)
- **Archiving of publications, data, models, & software tools and open data standards & conventions**
  - Data management plan plus software productivity and sustainability requirements for CESD projects
  - Work with new ESS-DIVE
  - Draw on work of ESIP, ISMC, CSDMS, EarthCube
- **Uncertainty quantification (UQ) & data assimilation (DA)**
  - Akuna-CLM, DART-PFLOTRAN, PEcAn
- **Scientific workflows and model & data analysis tools**
  - Jupyter notebooks
- **Community outreach**
  - AGU Fall Meeting sessions on “Computational Methods and Tools for Model–Data Integration” and “Big Data in the Geosciences” and “ML”

# Path Forward

- **Community survey on workflows and model–data integration tools being conducted**
  - Please take this survey by the end of the ESS PI Meeting: https://goo.gl/forms/BdLCDpq1IZckhKPI3
- **Preliminary results with 27 responses:**
  - 56% Python
  - 33% MATLAB
  - 30% R
  - 30% Visit
  - 22% NCL
  - 19% C/C++
  - 11% IDL
  - 7% FORTAN
  - 48% Jupyter notebooks
  - 44% ILAMB
  - 15% PEcAn
  - 15% Akuna