U.S. DEPARTMENT OF ENERGY | Office of Science

CESD Cyberinfrastructure Working Groups

Environmental System Science (ESS) PI Meeting

Bolger Center, Potomac, Maryland, USA

April 24, 2017

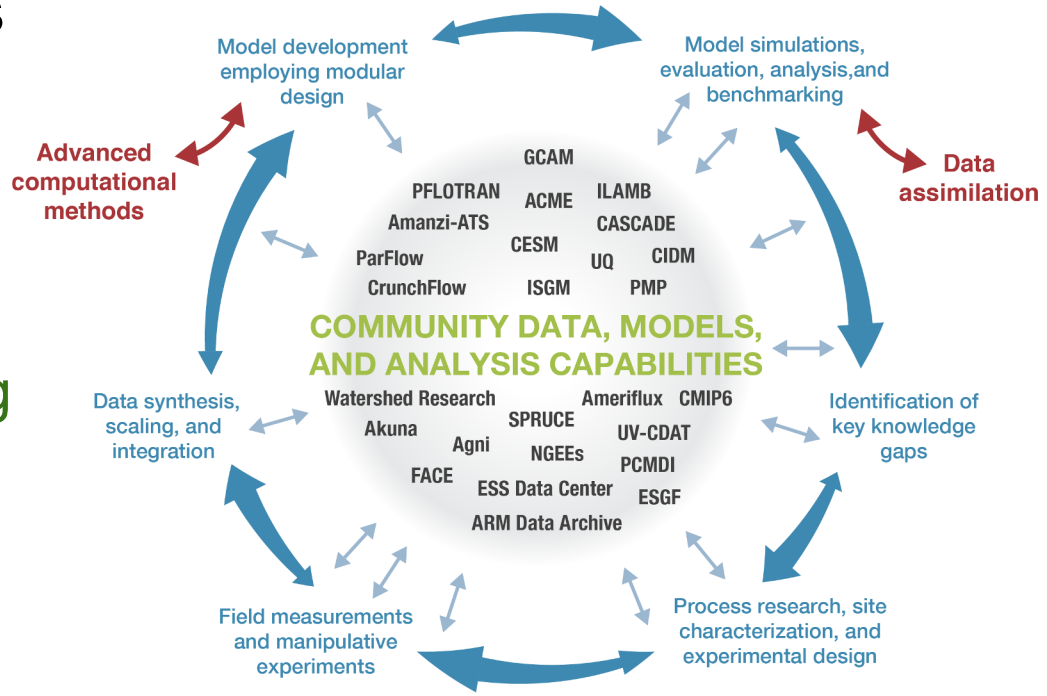
Model–Data Integration

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

Team Members: Tom Boden, Maoyi Huang, Lara Kueppers, Jitu Kumar, Nate McDowell, Umakant Mishra, Scott Painter, Bill Riley, Joel Rowland, Shawn Serbin, Peter Thornton, Charu Varadharajan, Roelof Versteeg, Haruko Wainwright, Cathy Wilson, Stan Wullschleger, Chonggang Xu

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 from 2016

- **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 from 2016 (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**
  - New whitepaper : **Geospatial Science to Inform Land Surface Models** (Mishra, Serbin, Wainwright, Kumar, Huang, and Chen)
  - Proposal: **Virtual Laboratory for Remote Sensing**
- **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 Data Archive
  - Draw on work of ESIP, ISMC, CSDMS, EarthCube (talks later)
- **Uncertainty quantification (UQ) & data assimilation (DA)**
  - Akuna-CLM, DART-PFLOTRAN, PEcAn
- **Scientific workflows and model & data analysis tools**
  - Jupyter notebooks

Path Forward in 2017

- **Community survey on workflows and model–data integration tools just started**
  - Please take this survey by the end of the ESS PI Meeting: https://goo.gl/forms/BdLCDpq1IZckhKPI3
- **Model–Data Integration Breakout Sessions at 3:00 p.m**
  - Geospatial/remote sensing and model diagnosis/benchmarking
  - Data assimilation, UQ, workflows, Jupyter notebooks
- **White papers and community survey results will be collected into a CESD Model–Data Integration Report by December**
- **We expect to have two or more proposals for Kickstarter-style voting for funding**