Update on the International Land Model Benchmarking (ILAMB) Package and IOMB

Forrest M. Hoffman
Oak Ridge National Laboratory

ESS Cyberinfrastructure Working Group Meeting

May 11, 2020
The International Land Model Benchmarking (ILAMB) community coordination activity was designed to

- Develop internationally accepted benchmarks
- Promote the use of these benchmarks
- Strengthen linkages between experimental, remote sensing, and modeling communities
- Support the design and development of open source benchmarking tools (Luo et al., 2012), like the **ILAMB Package** (Collier et al., 2018)
Development of ILAMB Packages

- **ILAMBv1** released at 2015 AGU Fall Meeting Town Hall, doi: [10.18139/ILAMB.v001.00/1251597](10.18139/ILAMB.v001.00/1251597)
- **ILAMBv2** released at 2016 ILAMB Workshop, doi: [10.18139/ILAMB.v002.00/1251621](10.18139/ILAMB.v002.00/1251621)
- Open Source software freely distributed
- Routinely used for E3SMv1 and CESM2 evaluation during development
- Employed to evaluate CMIP5 models
- Models are scored based on statistical comparisons (bias, RMS error, phase, amplitude, spatial distribution, Taylor scores) and functional response metrics
ILAMBv2.5 Package Current Variables

- **Biogeochemistry:** Biomass (Contiguous US, Pan Tropical Forest), Burned area (GFED3), $\text{CO}_2$ (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 Several Generations of CLM

- CLM saw improvements in mechanistic treatment of hydrology, ecology, and land use with many more moving parts
- Simulations improved even with enhanced complexity
- Observational datasets not always self-consistent
- Forcing uncertainty confounds assessment of model development (not shown)

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

(Lawrence et al., 2019)
ILAMB Graphical Diagnostics
New PEcAn-ILAMB site-level diagnostics
Variable-to-Variable Comparisons
Land Model Performance Depends Strongly on Forcing

- Depending on the forcing used and the metric selected, different models may perform equally well.
- ILAMB scores for CLM4, CLM4.5, and CLM5 forced with GSWP3 vs. CRUNCEP (left) and the cumulative land carbon sink for CMIP5 vs. CLM offline models (right).

Bonan et al. (2019)
Evaluation of the DGVMs using the International Land Model Benchmarking system (ILAMB; Collier et al., 2018) (left) absolute skill scores and (right) skill scores relative to other models for a subset of ILAMB variables.

Le Quéré et al. (2018)
Addressing Observational Uncertainty

- Few observational datasets provide complete uncertainties
- ILAMB uses multiple datasets for most variables and allows users to weight them according to a rubric of uncertainty, scale mismatch, etc.
- ILAMB can also use:
  - Full spatial/temporal uncertainties provided with the data
  - Fixed, expert-derived uncertainty for a dataset
  - Uncertainties derived from combining multiple datasets
- Experiments with CLASS self-consistent data (Hobeichi et al., 2020) demonstrates that while scores shift, including uncertainty rarely alters the rank ordering of models (figure)
International Ocean Model Benchmarking (IOMB) Package

- Evaluates ocean biogeochemistry results compared with observations (global, point, ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base, also runs in parallel
- Built on python and open standards
- Is also open source and will be released soon

Chlorophyll / SeaWIFS

Bias

Spatial Distribution

Annual & Seasonal Cycles
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 for CMIP5 outperforms any single CMIP5 model
- The multi-model mean for CMIP6 outperforms any single CMIP6 model
- The multi-model mean CMIP6 land model is the “best model” overall

(Hoffman et al., in prep)
CMIP5 and CMIP6 Land Model Global Gross Primary Productivity

- Most models of the same lineage improved in various characteristics between CMIP6 and CMIP6.
- The mean CMIP6 and CMIP5 models perform best.

(Hoffman et al., in prep.)
For more information...

- International Land Model Benchmarking (ILAMB) Package
  https://www.ilamb.org/

- Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area
  https://www.bgc-feedbacks.org/

- Forrest M. Hoffman
  Oak Ridge National Laboratory
  forrest@climatemodeling.org
For more information...


