Model Evaluation Discussion

Moderators:
Forrest Hoffman (forrest@climatemodeling.org)
Martial Mancip (Martial.Mancip@ipsl.jussieu.fr)

Marie Curies/iLEAPS Workshop, Hyères, France

November 17–20, 2008
Participants

Igor Oliveira  U. Cape Town  igor@csog.uct.ac.za
Juan Manuel Sachez  U. Valencia  Juan.M.Sanchez@uv.es
Victoria Wittig  U. Illinois  witting@illinois.edu
Matthieu Guimberteau  IPSL  Matthieu.Guimberteau@lmd.jussieu.fr
Anna Wramneby  Lund U.  anna.wramneby@nateko.iu.se
Stephen Sitch  Hadley Centre  stephen.sitch@metoffice.gov.uk
Pierre Gentine  MIT  gentine@mit.edu
Emmanuel Kpemlie  INRA Avignon  Emmanuel.Kpemlie@avignon.inra.fr
Carlos Jimenez  LERMA  carlos.jimenez@obspm.fr
Christine Delire  Météo France  cdelire@gmail.com
Ben Cook  NASA-GISS  bc9z@ldeo.columbia.edu
Rita Wania  U. Bristol, UK  rita@wania.net
Sam Levis  NCAR  slevis@ucar.edu
Martial Mancip  IPSL  Martial.Mancip@ipsl.jussieu.fr
Forrest Hoffman  ORNL  forrest@climatemodeling.org

Moderators: Forrest Hoffman (forrest@climatemodeling.org)
How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?
Topic of Group Discussion

*How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?*

Clarification

We took these questions to refer to land surface models in general, not just dynamic vegetation/biogeography models.
Limits of Land Surface Models/Schemes

- What are the spatial and time scales involved?
- Should we try to restrict or limit the use of models?
- Development and evaluation of models is driven primarily by the desired applications.
- What are the acceptable limits of model assumptions?
- Is peer review an acceptable method of establishing or maintaining these limits?
- It is dangerous to “over tune” models in offline mode, possibly breaking the coupled model.
- It is dangerous to tune for the wrong time or spatial scale or for a specific region or continent.
Steps for Model Evaluation

- **Protocol** - designed to elucidate performance under past, present, and future climate across all space and time scales
- **Metadata Standards** - for simplified manipulation and analysis, especially in preparation for AR5 and beyond
  - mapping PFTs to standard biome types?
  - mapping carbon pools to standard pool types?
- **Metrics** - based on comparison with best available satellite- and ground-based observational datasets
- **Diagnostics** - standard, open source package supporting all the metric comparisons
- **Scoring** - community-developed weighting of performance on metrics based on metric importance and data uncertainty
- **Distribution** - open distribution of model results to support related science by others, using the same Earth System Grid (ESG) system as IPCC

Moderators: Forrest Hoffman (forrest@climatemodeling.org)  Martial Mancip (Martial.Mancip@ipsl.jussieu.fr)
Prospects for Model Evaluation

- Many variables are needed to comprehensively evaluate processes in models.
- There are many ways to get the right answer for the wrong reason.
- It is important to 1) combine many datasets of similar observations for comparison with model results, and 2) these datasets must be processed in the same way for consistency.
- Fluxes are easier to validate than pools.
- We should frame our analysis in terms of processes (i.e., photosynthesis, phenology, etc.).
Forcing and Evaluation Datasets

- FluxNet - latent & sensible heat (Effort to use GEWEX)
- Model farm - Reto Stöckli’s system for running many models offline with FluxNet site data
- AmeriFlux and FACE sites
- River gauges - integrative (Trenberth & Dai 2002), GRACE
- NOAA GMD flasks for CO$_2$ seasonal cycle
- MODIS - pattern, phase ("modeled observations")
- Tree rings and other proxies

Forcing/Met:

- NLDAS Forcing: 1985–present ($\frac{1}{8}$°, US only)
- ISLSCP II: 1980s
- NCEP/NCAR reanalysis: 1948–2004 (Qian et al.)
- CRU (East Anglia): 2002 (more will come soon)
- NCC (NCEP Corrected with CRU): 1949–2000
- ERA interim: 1989–2008 (New version of ECMWF reanalysis)
Global FluxNet Sites
# Processes

<table>
<thead>
<tr>
<th>Processes - FluxNet, satellites, tree rings, NPN</th>
<th>short $t$</th>
<th>long $t$</th>
<th>short and long $t$</th>
<th>short and long $t$</th>
<th>small</th>
<th>small and large</th>
</tr>
</thead>
<tbody>
<tr>
<td>photosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>small</td>
<td></td>
</tr>
<tr>
<td>phenology</td>
<td>short $t$</td>
<td></td>
<td></td>
<td></td>
<td>small</td>
<td></td>
</tr>
<tr>
<td>land cover</td>
<td>long $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>snow</td>
<td>short and long $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>fire</td>
<td>short $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>other disturbances</td>
<td>short and long $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>climate response</td>
<td>short and long $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>response to extreme events</td>
<td>short and long $t$</td>
<td></td>
<td></td>
<td></td>
<td>small and large</td>
<td></td>
</tr>
</tbody>
</table>

Moderators: Forrest Hoffman (forrest@climatemodeling.org)
### Variables

<table>
<thead>
<tr>
<th>Surface energy fluxes - FluxNet, GEWEX, MODIS, others</th>
<th>temperature (2m–80m)</th>
<th>short t</th>
<th>small</th>
</tr>
</thead>
<tbody>
<tr>
<td>evapotranspiration</td>
<td>short t</td>
<td>small</td>
<td></td>
</tr>
<tr>
<td>albedo</td>
<td>short and long t</td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>Hydrology - FluxNet, river gauges, GRACE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>surface water fluxes</td>
<td>short t</td>
<td>small</td>
<td></td>
</tr>
<tr>
<td>soil moisture (+ deep soil)</td>
<td>short and long t</td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>snow</td>
<td>short and long t</td>
<td>small and large</td>
<td></td>
</tr>
<tr>
<td>permafrost</td>
<td>long t</td>
<td>large and large</td>
<td></td>
</tr>
</tbody>
</table>
### Variables

| Biogeochemistry - FluxNet, flasks, MODIS, OCO | 
|------------------------------------------------|---------------------------------------------------|
| gross primary productivity                    | short $t$                                          |
| net primary productivity                       | short $t$                                          |
| respiration                                    | short $t$                                          |
| net ecosystem exchange                         | short $t$                                          |
| CO$_2$ & CH$_4$ emissions                      | short $t$                                          |
| CO$_2$ seasonal cycle                          | long $t$                                           |
| C pools (+ slow pools)                         | short and long $t$                                 |
| **Land cover - AVHRR, MODIS, tree rings, DesDynI (future)** | 
| leaf area index                                | short $t$                                          |
| NDVI                                           | short $t$                                          |
| biomass/yield                                  | short and long $t$                                 |
| vegetation distribution                        | long $t$                                           |
| land use                                       | long $t$                                           |
Recommendations

- Write a review paper on the current state of best available datasets for model evaluation?
- Compare with what is available (considering scales of space/time). Can community develop “best” datasets?
- Better document model processes (useful for understanding analyses of model results).
- Offline improvements may not improve the coupled model; they may make it worse! (e.g., Sam’s talk)
- **Closer collaboration between measurement and modeling communities!**
- **Closer collaboration between modeling groups!**
- We will establish a mailing list to continue discussions and invite others to participate.
Did We Answer the Charge?

Topic of Group Discussion

_How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?_
Did We Answer the Charge?

**Topic of Group Discussion**

*How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?*

- We evaluate models by confronting them with best-available observational datasets.
Did We Answer the Charge?

Topic of Group Discussion

*How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?*

- We evaluate models by confronting them with best-available observational datasets.
- Models must be tested and evaluated in offline, partially coupled, and fully coupled modes over short and long time scales over small and large spatial scales.
Did We Answer the Charge?

Topic of Group Discussion

*How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?*

- We evaluate models by confronting them with best-available observational datasets.
- Models must be tested and evaluated in offline, partially coupled, and fully coupled modes over short and long time scales over small and large spatial scales.
- Experiments should include historical, present-day, and future time periods.
Topic of Group Discussion

*How can we properly evaluate our land-surface models (DGVMs) when they are embedded within climate models? By what means? What are the proper experiments to design?*

- We evaluate models by confronting them with best-available observational datasets.
- Models must be tested and evaluated in offline, partially coupled, and fully coupled modes over short and long time scales over small and large spatial scales.
- Experiments should include historical, present-day, and future time periods.
- **This is hard!** But we could take advantage of each others’ work.
Thank you!
Questions?
More Discussion?

Contact: Forrest Hoffman (forrest@climatemodeling.org) and Martial Mancip (Martial.Mancip@ipsl.jussieu.fr)