DRAFT 2/21/07

Metrics for the CLAMP Intercomparison

Overview:

Table 1 on page 2 summarizes metrics used for evaluating terrestrial biogeochemical model performance. There is a corresponding section describing the approach used to compute each metric component and a justification for the relevance. All of the datasets assembled have their strengths and weaknesses. The idea is that for each new model run, we can run a script executing these metrics so users can rapidly assess model performance.

There are short-term and long-term goals of this analysis. The short-term goal this fall is to use these metrics for the CLAMP project to improve the land carbon model for the CCSM. The long term goal is to incorporate these diagnostics as a standard evaluation package for land biogeochemistry in the NCAR CCSM, as well as to propose standardized procedures for other modeling groups.

When possible, we will try to examine the sensitivity of biogeochemical fluxes and pools to changes in a driver variable, rather than the absolute value of these fluxes and pools, recognizing that biogeochemistry critically depends on the physical climate within the model. For example, we compare NPP observations normalized by precipitation to model results normalized in the same way, to adjust for possible model biases in precipitation.

We also plan to analyze a number of model properties that may not necessarily map onto observed datasets, but nevertheless will provide insight about model performance and behavior.

Dataset location:

We currently have a copy of the NPP and LAI datasets on Natalie’s website. We are working on the others.

Table 1. Score Sheet for Biogeochemical Model Evaluation

Metric	Metric components	Sub-Score Total	Total Possible Score
NPP	Matching EMDI Net Primary Production (NPP) observations	5	20
	EMDI comparison, *normalized by PPT*	5	
	Correlation with MODIS (r^2)	5	
	Latitudinal profile (r^2)	5	
LAI	Matching MODIS observations		
	Phase (derived separately for major biome classes)	5	20
	Mean (derived separately for major biome classes)	5	
	Maximum (derived separately for major biome classes)	5	
	Growing season length (derived separately for major biome classes)	5	
CO_2 Seasonal Cycle	Matching the phase and amplitude at NOAA observation stations		20
Carbon Stocks	Aboveground vegetation within the Amazon Basin from *Saatchi et al.* (2006)	5	10
	Global belowground carbon (top 30 cm) from *Batjes* (2005).	5	
Eddy covariance energy & CO_2 fluxes	Net radiation	5	20
	Latent heat (annual mean)	5	
	Sensible heat (annual mean)	5	
	CO_2 fluxes	5	
Transient dynamics	Beta factor for CO_2 fertilization – Norby mean for temperate ecosystems		10
	Rate constants for litter decomposition (LiDAT, Post?)		
	El Nino anomaly 1998 (NEE and fire components) – Van der Werf et al. (2004)		
	dNPP/dT, dNPP/PPT		
		Total:	100

1. Net Primary Production

Rational: NPP represents the carbon flow potentially available as an energy source for heterotrophs and to humans for use as food or fiber.

Site NPP Comparison
Data Source: EMDI Net Primary Production dataset from the Oak Ridge National Laboratory (http://www.daac.ornl.gov/NPP/npp_home.html). These observations are in the directory: NPPobservationEMDI/EMDI_ClassA_NPP_81.csv

Metric:

$$M = 1 - \frac{\sum_{i=0}^{nsites} \frac{|m_i - o_i|}{m_i + o_i}}{nsites} \qquad (1)$$

Where M is the evaluation metric, m_i is the model NPP at the grid cell from the control run corresponding to the observed data set (o_i), and *nsites* is equal to the 81 data points in the ORNL holding. To compute the number of points for this metric, M would be multiplied by the total number of points available for this class (in this case, 5).

Strengths: NPP is a key biogeochemical variable and this is a direct model-data comparison.

Weaknesses: There is a spatial mismatch between the scale of the observations and the size of model grid cells. This will contribute to scatter in equation 1, possible biases, and a reduction in M.

Supporting plots and tables: A scatter plot of NPP observed vs. NPP modeled (plot has 81 points).

Site NPP Comparison Normalized by Precipitation
Name: Oak Ridge National Laboratory EMDI Net Primary Production dataset:
Source: http://www.daac.ornl.gov/NPP/npp_home.html

Metric: Same as equation 1, but now NPP has been normalized by annual mean PPT in discrete PPT histogram intervals, and the metric is summed over the number of histogram bins instead of the number of individual sites.

Strengths: By normalizing the observations by precipitation, we can make adjustments for inadequacies in the GCM’s hydrological cycle.

Weaknesses: In some areas, NPP is limited by the length of the growing season (temperature) and not precipitation.

Supporting plots and tables: A histogram of NPP observed and modeled output, of the following form:

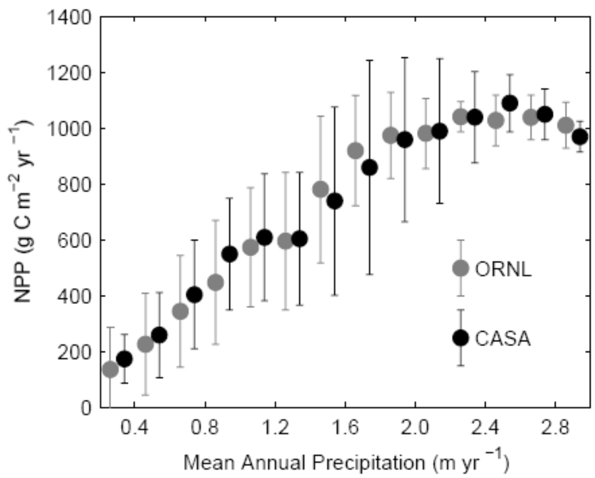

Figure 1. Subplot example for NPP model- data comparison. Original figure from *Van der Werf et al.* [2006].

Satellite NPP observations
Source: MODIS MOD 17 annual net primary production from Steve Running and Maosheng Zhao [*Zhao, et al.*, 2005].

Metric: Compute Pearson's r correlation coefficient between observations and model distribution using all land grid cells. To compute the number of points for this metric, the r^2 value would is multiplied by the total number of points available for this class.

Strengths: Satellites do an excellent job at capturing spatial variability. This metric tests the model's ability to capture this variability.

Supporting plots and tables: Global maps of annual NPP, including the observations, the model, and the difference.

Latitudinal NPP distribution
Source: MODIS MOD 17 annual net primary production from Steve Running and Maosheng Zhao [*Zhao, et al.*, 2005].

Metric: Compute Pearson's r correlation coefficient between satellite NPP observations and model distribution using latitudinal profile (g C m^{-2} per zonal mean of land area). To compute the number of points for this metric, the r^2 value would is multiplied by the total number of points available for this class. A separate plot of the latitudinal zonal mean distribution of NPP from MODIS and the model would help with diagnosis of model biases.

Strengths: This would help identify extra-tropical/tropical biases in model performance.

Supporting plots and tables: Latitudinal zonal mean plot showing model and MODIS values.

2. Leaf area

Data Source: MODIS MOD 17 Leaf Area Index from Steve Running and Maosheng Zao

Rational: Leaf area has important consequences for the surface energy budget, the hydrological cycle, rates of photosynthesis and thus carbon inputs to ecosystems.

Metrics:

We will use satellite observations to identify four leaf area index (LAI) characteristics for each biome: 1) the time of year (month) of the observed maximum LAI, 2) the mean annual LAI, 3) the maximum observed LAI, and 4) the growing season length. Following *Churkina et al.* [2005] [*Churkina, et al.*, 2005], we will define the growing season length as the number of days that LAI is above an arbitrary threshold – here set equal to an LAI of 1.0.

Phase:

$$M = \frac{6 - \left| T_{LAIMAX}(mod) - T_{LAIMAX}(obs) \right|}{6} \quad (2)$$

Where T_{LAIMAX} is the month of maximum leaf area either for the model (*mod*) or the observations (*obs*), averaged over all the grid cells within each biome. If they agree perfectly, then the metric *M* is 1.0. If the model and observations are exactly 6 months out of phase, then the metric *M* is 0.0.

Mean:

For the mean LAI of each biome, we will use an equation similar to equation 1 operating on all the grid cells within each biome.

Maximum:

For the maximum LAI of each biome, we will use an equation similar to equation 1 and operate on all the grid cells within the biome. We will first compute the mean monthly maximum from both the models and the observations.

Growing season length:

We will compute the growing season length using a simple threshold approach on leaf area (when LAI > 1.0). We will use an equation similar to equation 1 and operate on all the grid cells within the biome.

For each of the 4 LAI metrics described above, we will average them across all the biomes to obtain a single model score.

Strengths: LAI is a key variable and one that is controlled prognostically by model biogeochemistry.

Weaknesses: Particularly for the mean, maximum, and growing season length, there maybe biases in the satellite-derived estimates of LAI. That is, the satellite LAI estimates depend on atmospheric and canopy radiative transfer models that require validation. The metric of the phase should be less sensitive to these types of potential bias.

Supporting plots and tables:

A. Global maps for each of the 4 LAI metrics (mean, maximum, phase, and growing season length). For each metric, the observations, the model, and the difference are plotted on the same page.
B. Monthly mean plots of LAI from the model and the observations for each biome.
C. A table with the following form: The rows are the different biomes. There are 4 sets of columns for each of the 4 LAI metrics. Each set consists of the biome mean from the observations, the model, and the computed metric described above. At the bottom of the Table in an additional row these quantities are averaged over each biome.

3. Seasonal cycle of atmospheric carbon dioxide

Rational: This metric tests the combined effect of the seasonal timing and magnitude of NPP and heterotrophic respiration in northern hemisphere biomes. As such, it provides some confidence in the temperature sensitivity of respiration in these biomes and that prognostic leaf area (and thus GPP and NPP) is working properly.

Data source: The observations of the seasonal cycle from Globalview have a .seas extension and can be found at the following site: ftp.cmdl.noaa.gov/ccg/co2/GLOBALVIEW/gv/.

Metric: Model performance in each of 4 latitudinal zones will be evaluated for a maximum of 5 points per region. The four regions are: 90°N - 60°N, 30°N - 60°N, EQ - 30°N, and 90°S - EQ. The northern hemisphere is weighted more than the southern hemisphere because the signal to noise of the observations are better there; the seasonal cycle is so small in the Southern Hemisphere it is difficult to separate from interannual variability and the long-term growth rate. Within each region, both the correlation coefficient (r) and the magnitude of the simulated seasonal amplitude relative to the observations will be weighted equally. This information will allow us to evaluate both the phase and amplitude of model estimates.

Within each region, model grid cells will be extracted for each surface station in the Globalview station list. A monthly mean seasonal cycle will be constructed by equally weighting all station locations within each latitude band. The same will be done for the observations. Using these two time series, the Pearson's correlation coefficient (r) will be computed. For the amplitude metric, we will report the ratio of monthly mean of the model amplitude (A_m) relative to the observations (A_o). This could be represented by:

$$M = 1 - \left| \frac{A_M}{A_O} - 1 \right| \quad (1)$$

Where M is the amplitude metric, A_m is the monthly peak to trough amplitude of the model, and A_o is the monthly peak to trough amplitude of the observations. These two metrics could be combined as $(M + r^2)/2.0$ and then in a final step this combination could be multiplied by the number of points assigned to each latitudinal zone.
Here is a possible schematic of the output for a model scorecard for the seasonal cycle metric:

Table 2. Seasonal cycle of atmospheric CO_2 metrics (filled with example values)

Region	Pearson's Correlation Coefficient (r)	Amplitude ratio (model/observations)	Combined Score	Total Model Score
60°N- 90°N	0.87	1.32	3.59	
30°N- 60°N	0.90	1.09	4.30	
EQ - 30°N	1.00	1.00	5.00	
90°S- EQ	0.91	0.78	4.02	
Total:				16.91

Strengths: Observations of this metric are robust. Ocean and fossil fuels contribute only weakly to the CO_2 seasonal cycle, making it a good tracer of terrestrial biosphere-atmosphere exchange in northern ecosystems.

Weaknesses: Comparing biogeochemical model fluxes with observations requires a model of atmospheric transport. Biases in horizontal or vertical mixing within the atmospheric model can influence model-data comparisons. Also, this is only a very weak constraint on the seasonal cycle of fluxes from savanna and tropical forest ecosystems because of strong equatorward flow and vertical convection in tropical regions.

Supporting plots and tables: Plots of monthly mean observations vs. model output for each latitude zone and also for each individual station. Also, a table like Table 3.

4. Carbon Stocks

Amazon Forest Aboveground Biomass

Source: An LBA gridded dataset developed by Sasan Saatchi.

Metric:

Same metric as for NPP – a comparison of model and observed measurements for each grid cell.

Strengths: The carbon stocks within the Amazon represent a large and vulnerable pool.

Supporting plots and tables: A plot of the model aboveground biomass and the observed aboveground biomass for the Amazon region.

Soil carbon
Name: ISRIC-WISE global dataset of derived soil properties on a 0.5 by 0.5 degree grid (ver. 3.0) [*Batjes*, 1996].
Source:
http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE+spatial+ver.+3.0.htm

Metric:

We will use a metric similar to equation 1 for NPP, but comparing the gridded soil carbon inventory from *Batje*s (2004) with integrated carbon stocks in the first 30 cm from the models:

Where M is the evaluation metric, m_i is the model soil carbon to 30 cm at the grid cell corresponding to the observed data set (o_i). In a final step, M is multiplied by the total number of points available for this class.

Strengths: This gridded dataset draws upon over 4000 measurement profiles.

Weaknesses: The dataset excludes the litter layer. This will require sampling the models in a similar way (excluding litter, and including carbon pools that represent the top 30 cm).

5. Energy and Surface CO_2 Fluxes

Data Source: FLUXNET eddy covariance measurements observations from Reto and Forrest Hoffman

Rational: Partitioning of incoming radiation into shortwave and longwave radiation outgoing components and the flow of energy into sensible and latent heat fluxes is a key element of a coupled land model. Land surface and biogeochemical model components must interact to accurately provide these terms. For example, prognostic leaf area from the biogeochemical model plays an important role in setting surface albedo and the partitioning of sensible and latent heat fluxes. (This information will be extracted from 1.4 and 2.4).

Metric:

We will compare modeled and observed monthly mean fluxes of latent heat, sensible heat, net radiation, and CO_2 fluxes from a set of available eddy covariance tower sites. An average annual cycle (comprised of monthly values) will be constructed by sampling the model during the months that eddy flux observations are available. For the comparison of model and observations, we will use a metric similar to that in equation 1. Averages from the observations should exclude periods of missing data each month.

Initially we will focus on the 8 sites listed on Forrest's website http://climate.ornl.gov/bgcmip/fluxnet/

Table 3. Eddy covariance tower site metrics

Site Name	Latitude	Time period of obs.	Latent Heat	Sensible Heat	Net Radiation	CO_2 Flux	Mean metric at each site
Boreas …	… …		M1	M2	M3	M4	Mave1 Mave2 …
Tapajos Forest							Mave n
Mean Score for each Variable			Out of 5 points S1	Out of 5 points S2	Out of 5 points S3	Out of 5 points S4	Total Score (20 points total) S Final

Supporting plots and tables: A plot for each tower site and variable comparing the modeled and observed monthly mean fluxes averaged over the time interval that the flux observations are available. Each metric number in the above table should link to the appropriate plot for that station.

6. Transient Metrics (next step)

References

Batjes, N. H. (1996), Total carbon and nitrogen in the soils of the world, *European Journal of Soil Science, 47*, 151-163.

Churkina, G., et al. (2005), Spatial analysis of growing season length control over net ecosystem exchange, *Global Change Biology*, *11*, 1777-1787.

Saatchi, S. S., et al. (2006), Distribution of aboveground live biomass in the Amazon Basin, *Global Change Biol*.

Zhao, M. S., et al. (2005), Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sensing of Environment*, *95*, 164-176.