

Supporting Information

for

**Systematic assessment of terrestrial biogeochemistry in
coupled climate-carbon models**

James T. Randerson, Forrest M. Hoffman, Peter E. Thornton, Natalie M. Mahowald,
Keith Lindsay, Yen-Huei Lee, Cynthia D. Nevison, Scott C. Doney, Gordon Bonan,
Reto Stöckli, Curtis Covey, Steven W. Running, and Inez Y. Fung

Published in

Global Change Biology

2009

Corresponding author:
James T. Randerson
phone: (949) 824 - 9030
fax: (949) 824 - 3874
email: jranders@uci.edu

Supporting Information Table of Contents:

1. Supporting Information Figures S1 to S6.....	2
2. Description of Metrics and Scoring System.....	8
3. Supporting Information References.....	15

Figures S1-S6

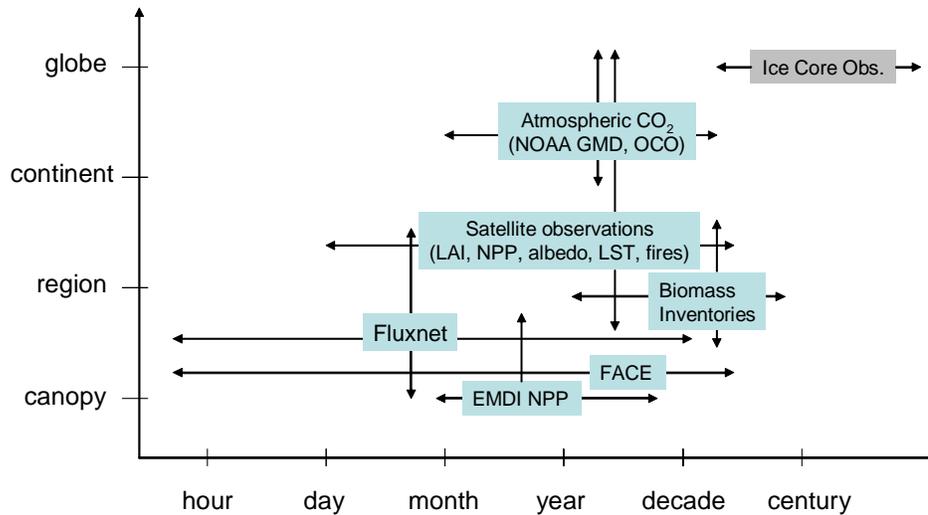


Figure S1. Conceptual diagram of observations available for testing carbon-climate models. Ice core measurements of the atmospheric CO₂ record provide constraints on the sum of ocean and land carbon fluxes when this information is combined with fossil fuel inventory time series. Isotope measurements from ice cores allow for similar constraints but including gross exchanges and reservoir turnover times. Contemporary atmospheric CO₂ observations from flask networks (NOAA GMD) and satellites (e.g., the Orbiting Carbon Observatory) provide information about the seasonal dynamics of net ecosystem exchange and continental-scale fluxes on timescales of years to decades. Biomass inventories are sparse but crucial for constraining allocation, tree mortality, and the mass of carbon vulnerable to deforestation. Satellite observations of leaf area index and other ecosystem variables provide global coverage at a high temporal resolution for a period of almost three decades, although cross-platform calibrations introduce considerable uncertainty. Free-Air Carbon dioxide Enrichment (FACE) experiments have quantified elevated CO₂ effects on ecosystem processes in temperate ecosystems, but less information exists for tropical forest and boreal biomes that account for most of terrestrial GPP and aboveground carbon storage.

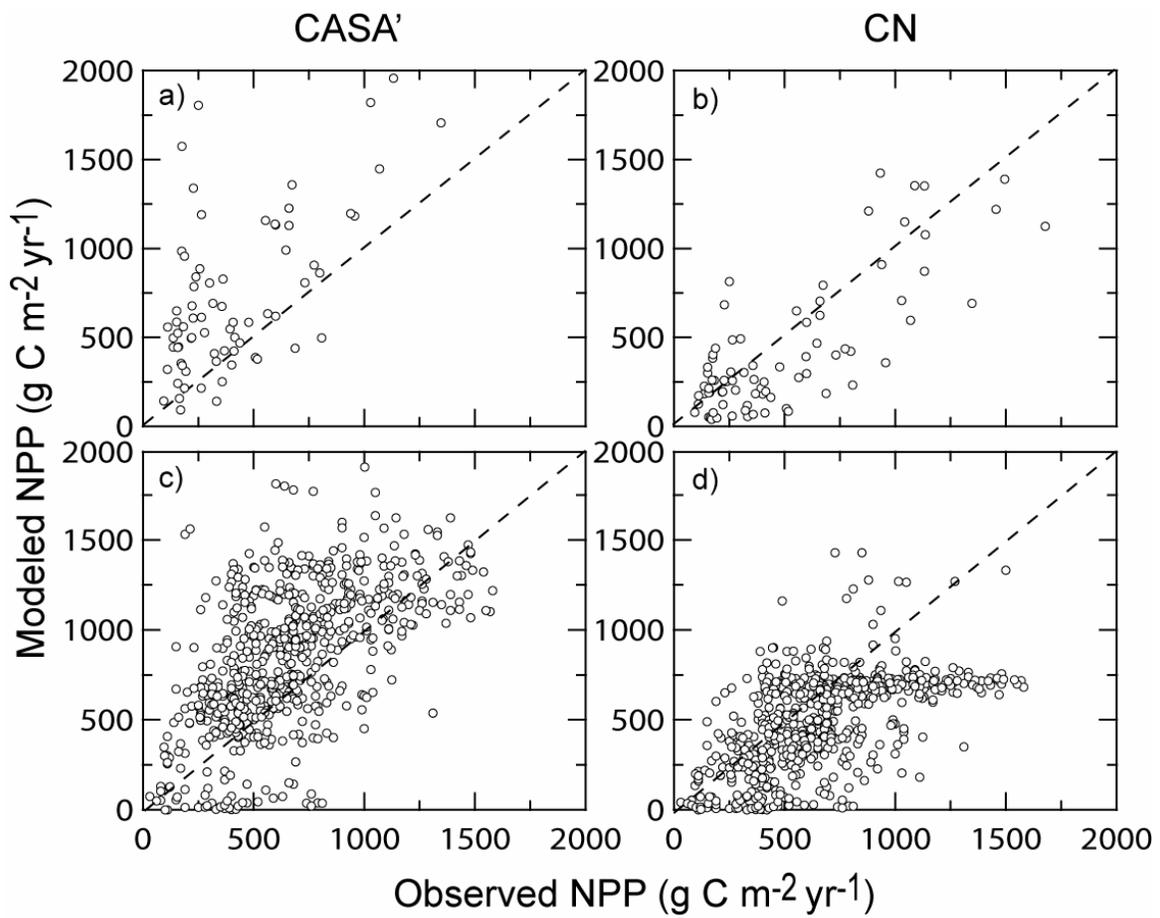


Figure S2. Comparison of net primary production for a) CASA' and b) CN models with class A observations from the Ecosystem Model Data Intercomparison Initiative (EMDI). The same comparison for class B observations is shown in c) and d).

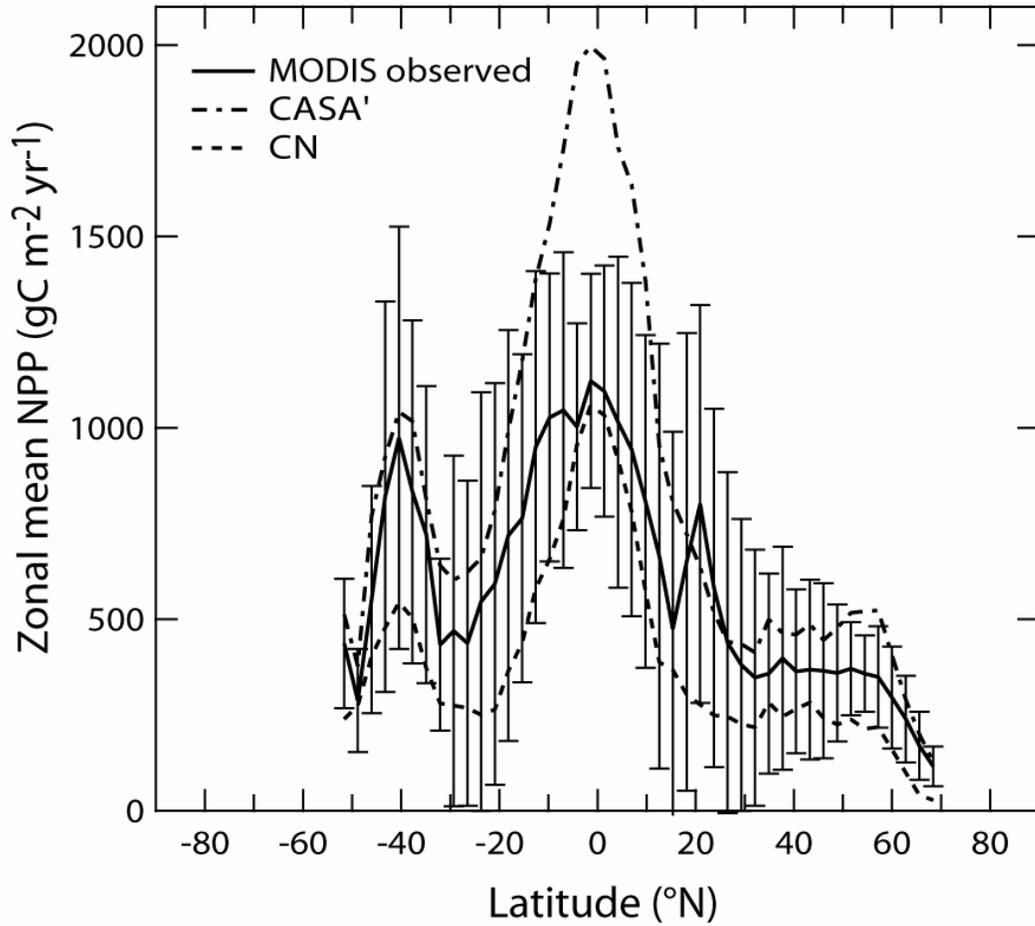


Figure S3. Zonal mean net primary production from MODIS satellite-based estimates compared with the models. We used the MOD17A3 collection 4.5 product from MODIS for this comparison (Heinsch *et al.*, 2003). We show the 200-2004 zonal mean and compare this model experiment 1.4 during the same period.

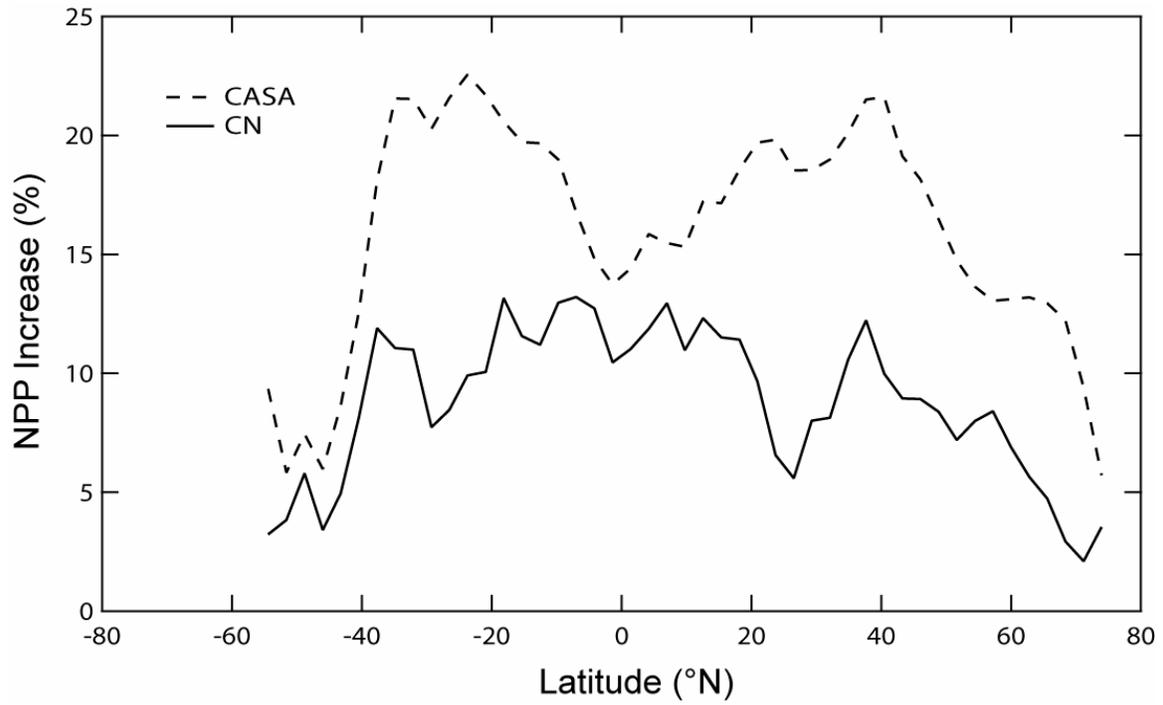


Figure S4. The zonal mean response of NPP to a step change in atmospheric CO₂ following the FACE experimental protocol. The model NPP response was averaged over the first 5 years after enrichment.

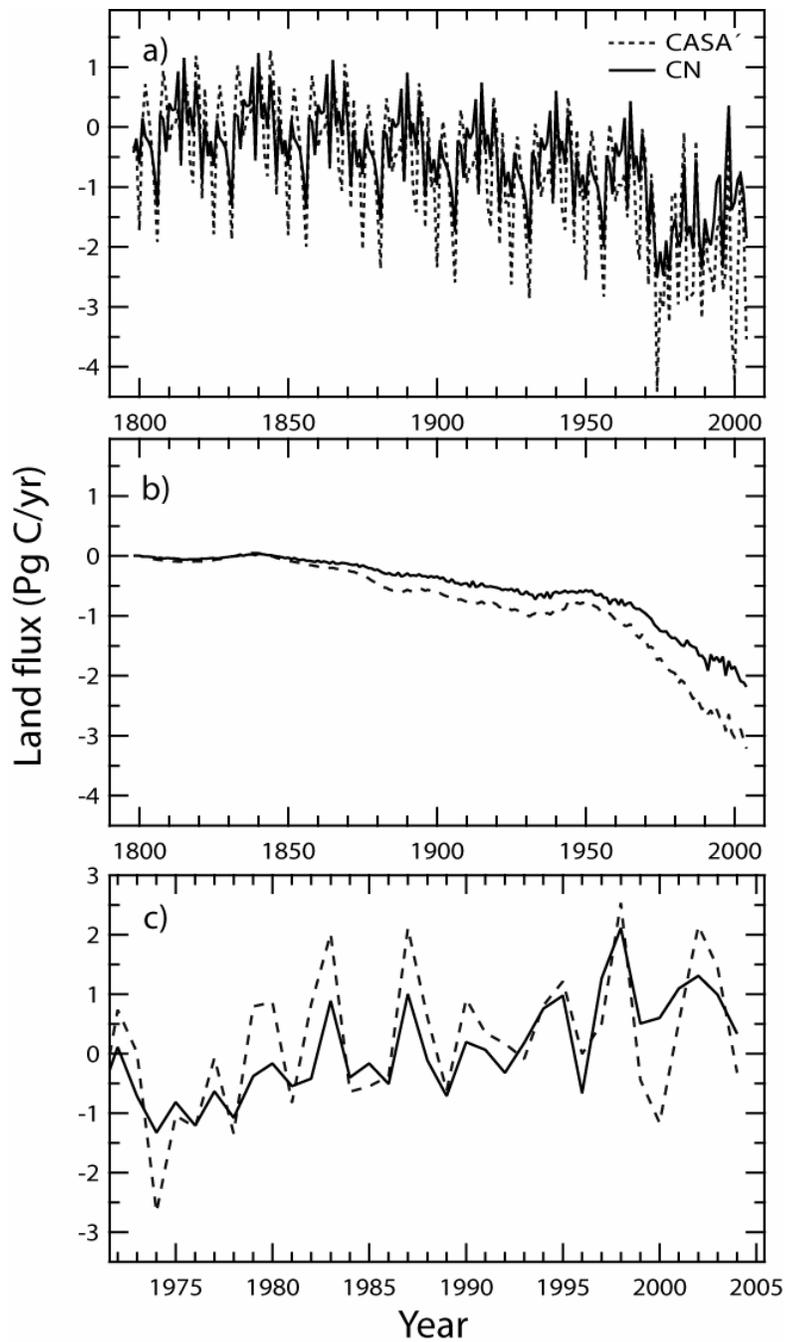


Figure S5. a) The global net land flux from experiment 1.4. This simulation includes climate variability and time-varying atmospheric CO₂ and nitrogen deposition. Climate for a 25-year span (1948-1972) was cycled until 1948, the beginning of the NCAR/NCEP reanalysis period. b) The difference in flux between experiments 1.4 and the climate only simulation (experiment 1.3). This panel shows the fluxes caused solely from the atmospheric CO₂ and nitrogen deposition forcing. c) The land flux driven solely by climate (experiment 1.3) during 1973-2004.

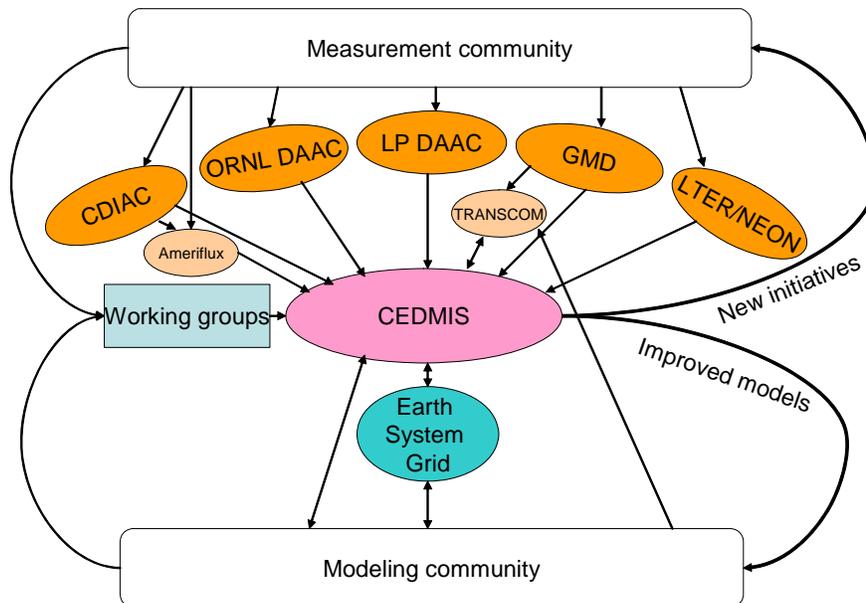


Figure S6. Conceptual diagram showing how a climate ecosystem data-model intercomparison system (CEDMIS) might function in the context of existing data centers and model archiving capabilities. CEDMIS would extract information from archived data sets and models to generate intercomparison diagnostics, using a series of scoring, visualization, and data extraction software tools. A key goal would be make the intercomparison diagnostics into modules that could be reused in multiple model-intercomparison projects (MIPs) in an open source format. This system could be used in a stand alone mode for individual model development or as the basis for community wide MIPs. Key data sources would include the Carbon Dioxide Information and Analysis Center (CDIAC), NASA’s Oak Ridge National Lab (ORNL) and Land Processes (LP) Distributed Active Archiving Centers (DAACs), NOAA’s Global Monitoring Division trace gas archives (including retrieved fluxes by means of atmospheric inversions such as TRANSCOM and CarbonTracker), and NSF’s Long Term Ecological Research (LTER).

Description of Scoring System

1. Leaf area index

Data Source: MOD15A2 collection 4 (Myneni *et al.*, 2002) with additional adjustments to interpolate across periods of cloud contamination as described by Zhao *et al.* (2005).

Scoring Metric:

1A. Phase

$$M = \frac{6 - |T_{DIFF}(model - obs)|}{6} \quad (s1)$$

where $T_{DIFF}(model - obs)$ is the difference (in months) between the month of maximum leaf area for the model and the month of maximum leaf area for the satellite observations at each grid cell. This quantity was averaged over all the grid cells within each biome to obtain a biome-level mean. If the models and observations were to agree perfectly within a biome (there was no phase offset), then M would be 1.0. If all the model and observations were exactly 6 months out of phase, then the metric M would be 0.0.

A global mean was created by averaging M across different biome types.

1B and 1C. Maximum monthly and annual mean LAI

For maximum monthly and annual mean LAI comparisons, we used equation s2 to obtain a scoring metric:

$$M = 1 - \frac{\sum_{i=1}^{ncells} |m_i - o_i|}{\sum_{i=1}^{ncells} m_i + o_i} \quad (s2)$$

where m_i is the model LAI at the grid cell corresponding to the satellite observation (o_i) and $ncells$ is the number model grid cells in each biome. When m_i and o_i were both 0, we set the quantity inside the summation equal to 1 for that grid cell.

A global mean was created by averaging M from s2 across different biome types. We averaged all available MODIS observations during 2000-2004 to construct an annual mean cycle of monthly observations at each grid cell. We sampled the models in the same way using output from experiment 1.4 during 2000-2004.

Component scoring

To obtain an overall model score for LAI, we multiplied the M for phase by 6 points, the M for maximum LAI by 5 points, and the M for mean LAI by 4 points. The timing of maximum LAI was believed to be the most robust aspect of the observations and thus

given the largest weighting, followed by maximum monthly LAI (likely to occur during snow-free periods and with smaller errors introduced by variations in soil reflectance), and then the annual mean.

2. Net Primary Production

2A. Site-level NPP Comparison

Data Source: EMDI Net Primary Production dataset from the Oak Ridge National Laboratory (http://www.daac.ornl.gov/NPP/npp_home.html). The observations we used were in files EMDI_ClassA_NPP_81_R1.csv and EMDI_ClassB_NPP_933_v2.csv.

Scoring Metric:

For the NPP site-level comparison we used equation s2, where m_i was the model NPP at the grid cell from experiment 1.4 corresponding to the observed NPP data point (o_i), and $nsites$ was equal to the total number of data points in either the class A (81) or class B (933) data set.

We separately computed M for the Class A and Class B observations and averaged them together to obtain a single metric for each model. To compute the number of points for this metric, M was multiplied by the total number of points available for this measurement class (in this case, 2).

2B. Site-level NPP Comparison Normalized by Precipitation

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 file EMDI_ClassB_NPP_933_v2.csv.

Metric: We used equation s2 with NPP from both the model and the observations averaged in discrete (400 mm/yr) precipitation bins. For the model, we used annual PPT estimates during 1975-2000 from our climate data set described in the main text (Qian *et al.*, 2006). For the observations, we used the annual PPT for each site provided in the EMDI dataset. We separately computed M for the Class A and Class B histograms and averaged them together to obtain a single metric for each model.

2C. Satellite NPP observations – spatial correlation

Data Source: MODIS MOD 17 annual net primary production [Zhao *et al.*, 2005].

Metric: We computed the Pearson's correlation coefficient r between the observations and the model distribution using all land grid cells. Our scoring metric was the square of the correlation coefficient:

$$M = r^2 \tag{s3}$$

In computing r^2 , for the model we used a map of annual mean NPP averaged over 2000-2004 from experiment 1.4. For the observations, we used a map of mean annual NPP from MODIS for the same time period.

The fractional absorbed photosynthetically active radiation (fAPAR) estimates obtained from MODIS satellite observations provide a partial test of the spatial pattern of NPP in the models. We used a correlation coefficient as a scoring metric to recognize the outstanding ability of satellite observations to quantify spatial gradients in fAPAR across continents (e.g., variations introduced by mountain ranges and lakes and gradients from low to high precipitation). The conversion of fAPAR to NPP by means of a light use efficiency adds additional uncertainty and is more difficult to directly constrain using satellite observations. This is why we excluded information on the specific magnitude of NPP variability for our scoring metric for the MODIS NPP observations.

2D. Satellite NPP observations - latitudinal correlation

Data Source: MODIS MOD 17 annual net primary production [Zhao et al., 2005].

Metric: Similar to the scoring metric for section 2C, but here estimating r^2 using a zonal mean vector from the observations and the model. This score metric evaluates the model's ability to capture the latitudinal distribution of NPP, including relative differences between high latitude and tropical ecosystems. This metric was estimated using the same dataset and model output as described in section 2C.

Component scoring

To obtain an overall model score for NPP, we multiplied the M for the site-level comparison by 2 points, the M for the NPP comparison normalized by PPT by 4 points, and the M values for the two satellite comparisons each by 2 points. We assigned a higher weight to the NPP comparison normalized by PPT because this comparison may be less sensitive to sub-grid scale climate variability than the other types of comparisons.

3. Annual Cycle of Atmospheric CO₂

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 is evaluated separately in each of three latitudinal zones. The three regions are: 90°N - 60°N, 30°N - 60°N, and EQ - 30°N. Within each zone, model grid cells are extracted for each surface station in the Globalview station list. A monthly mean seasonal cycle is constructed by equally weighting all station locations within each latitude band. The same is done for the observations.

We then compute two comparison metrics within each zone, one for the phase and the other for the amplitude. For the phase, we calculate the square of the Pearson's

correlation coefficient (r^2) using the 12 monthly mean values from the observations and the model. For the amplitude metric, we report the ratio of monthly mean of the model amplitude (A_m) relative to the observations (A_o). This is represented by:

$$M = 1 - \left| \frac{A_m}{A_o} - 1 \right| \quad (\text{s4})$$

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.

Component scoring

These two metrics were combined as $(M + r^2)/2.0$ where M is from s4 and then in a final step this value was multiplied by the number of scoring points assigned to each latitude zone. We assigned a higher number of possible points to the 90°N - 60°N and 30°N - 60°N zones (6 points each) than to the EQ - 30°N zone (3 points) for the following reasons. First, the amplitude of the seasonal cycle is larger in mid and high latitudes, making it easier to detrend the observations in order to construct a robust mean annual cycle. Second, the relative contribution to the mean annual cycle from ocean fluxes and fossil fuel emissions become increasingly important for stations near the equator, making it more challenging to make a direct comparison between the model simulations and the observations. It should be noted that for all of latitude bands described above, most of the signal comes from temperate and boreal ecosystems. The annual cycle of CO₂ does not provide a substantial constraint on the seasonal cycle of fluxes from tropical ecosystems.

4. Energy and CO₂ Fluxes from Ameriflux

Data source: We used L4 observations from Ameriflux that are located at: <ftp://cdiac.ornl.gov/pub/ameriflux/data/Level4/>

Scoring Metric:

We sampled the models during each year that the observations were available to build a multi-year set of mean monthly fluxes through 2004. We then created an annual mean cycle of fluxes from periods when the observations were available. We estimated model-data agreement using equation s2 at each site using the annual cycle of monthly means. Each flux tower site was weighted evenly in constructing our overall score. We separately constructed scores for net ecosystem exchange (NEE), sensible heat, latent heat, and gross primary production (GPP) fluxes.

Component scoring

To obtain an overall model score for energy and CO₂ fluxes section, we multiplied the M for the sensible heat and latent heat fluxes by 9 points, and the M for GPP and NEE by 6 points. GPP was weighted by a smaller amount than the energy fluxes because of the increased uncertainty associated with separating NEE into GPP and ecosystem respiration

components using models of ecosystem respiration. NEE is easier to directly measure, however, this flux is highly sensitive to the disturbance history of a site. This makes NEE difficult to directly compare with output from a model grid cell that, in our analysis, did not capture a site's disturbance history (i.e., this comparison has a high scaling mismatch).

5. Transient Fluxes

5A. Aboveground live biomass in the Amazon Basin

Data Source:

Saatchi, S.S., R.A. Houghton, D. Alves, B. Nelson. 2009. LBA-ECO LC-15 Amazon Basin Aboveground Live Biomass Distribution Map: 1990-2000. Data set. Available on-line [<http://daac.ornl.gov>] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAAC/908.

Available at:

http://daac.ornl.gov/LBA/guides/LC15_AGLB.html

Metric:

We used equation s2 to estimate M where $ncells$ was equal to the number of model grid cells within the spatial domain of the LC-15 dataset. We used model output from experiment 1.4 during year 2000 in this comparison. For CASA the wood pool was multiplied by 0.7 to get to the aboveground carbon component. For CN, aboveground live (sapwood) and aboveground dead (heartwood) tree pools (both in live trees) were combined for this comparison.

5B. Sensitivity of net primary production to elevated CO_2

Data Source:

The observations for the four FACE experiments we used are summarized by: Norby *et al.* (2005). Forest response to elevated CO_2 is conserved across a broad range of productivity. PNAS. 102: 18052–18056.

Model results and the observed site means from Norby *et al.* (2005) we used can be found in a Table at: <http://www.climate modeling.org/c-lamp/results/face/>

Metric:

We used equation s2 to estimate M where $ncells$ was equal to 4 – the number of FACE experimental sites. At each site we estimated the mean NPP increase (in %) for the first 5 years of the FACE experiment. These were the o_i 's in equation s2. We then compared this to the NPP increase in % from the model sampled at grid cells corresponding to these

4 sites. These were the m_i 's in s2. To estimate the NPP increases in the model, we used the difference between model experiments 1.6 and 1.5.

5C. Interannual variability of terrestrial fluxes

Data Source: Baker *et al.* (2006) with updates through 2004 provided by D. Baker (personal communication) using the same TRANSCOM methodology.

Metric:

Following a similar approach to the model-data comparison for the annual cycle of CO₂, we developed one metric for phase and a second for the magnitude of the flux variability and weighted the information from the two metrics evenly. For the phase, we calculated the square of the Pearson's correlation coefficient (r^2) using the annual global total terrestrial flux time series from the observations and from the model during 1988-2004 ($n = 17$).

For the magnitude metric, we relied upon equation s4, using the standard deviation of the model (A_m) and the standard deviation of the observations (A_o) computed over the 17 year period of the observations.

A long-term mean was removed from the observations and the model prior to our analysis to allow for a direct comparison of the interannual variability. This was done since the model simulations did not include land use change – which would influence the magnitude of the net land carbon flux.

5D. Fire emissions

Data Source: Global Fire Emissions Database version 2 (GFEDv2) from Van der Werf *et al.* (2006). These data are online at: <http://ess1.ess.uci.edu/%7Ejranders/data/GFED2/>

Metric:

Following a similar approach to the model-data comparison for the annual cycle of CO₂, we developed one metric for phase and a second for the magnitude of the flux variability and weighted the information from the two metrics evenly. For the phase, we calculated the square of the Pearson's correlation coefficient (r^2) using the monthly global total fire emissions time series from the observations and from the model during 1997-2004 ($n=96$ months).

For the magnitude metric, we relied upon equation s4, using the mean emissions from the model (A_m) and the mean emissions from the observations (A_o) computed over the 8 year period of the observations.

Since CASA' did not have a fire emissions module, its score in this comparison section was zero.

Component scoring

We assigned 10 points to the Amazon aboveground biomass comparison because the amount of tropical forest carbon in wood influences 1) the temperature sensitivity of terrestrial ecosystem fluxes, 2) the magnitude of carbon losses associated with land use change, and 3) the potential of this biome to accumulate carbon in response to elevated levels of CO₂. The FACE comparisons were assigned the same number of subscore points because the sensitivity of NPP to elevated levels of CO₂ plays a central role in setting the strength of the climate-carbon feedback within a model.

The TRANSCOM interannual variability and fire emissions comparisons were assigned a subscore of 5 points each. Their lower subscore in this section reflects a subjective assessment by the authors that both these data products have substantial uncertainties that arise, in part, from their use of models in transforming the raw observations into fluxes.

Supporting Information References

Baker DF, Law RM, Gurney KR, *et al.* (2006) TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988-2003. *Global Biogeochemical Cycles*, **20**, GB1002, doi:10.1029/2004GB002439.

Heinsch, FA, *et al.* (2003), User's Guide GPP and NPP (MOD17A2/A3) Products NASA MODIS Land Algorithm. Version 2.0., 57 pp.

Myneni RB, Hoffman S, Knyazikhin Y, *et al.* (2002) Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, **83**, 214-231.

Norby RJ, DeLucia EH, Gielen B, *et al.* (2005) Forest response to elevated CO₂ is conserved across a broad range of productivity. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18052-18056.

Qian TT, Dai AG, Trenberth KE, Oleson KW (2006) Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations. *Journal of Hydrometeorology*, **7**, 953-975.

Saatchi SS, Houghton RA, Alvala R, Soares JV, Yu Y (2007) Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, **13**, 816-837.

van der Werf GR, Randerson JT, Giglio L, Collatz GJ, Kasibhatla PS, Arellano AF (2006) Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, **6**, 3423-3441.

Zhao MS, Heinsch FA, Nemani RR, Running SW (2005) Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, **95**, 164-176.