CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Harvard University - DASH

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Terrestrial Biosphere Model Performance for Inter-Annual Variability of Land-Atmosphere CO2 Exchange

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Keenan, T. F., Ian Baker, Alan Barr, Philippe Ciais, Ken Davis, Michael Dietze, Danillo Dragoni, et al. Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange. Global Change Biology 18(6): 1971-1987.
Published Version	doi:10.1111/j.1365-2486.2012.02678.x
Accessed	February 19, 2015 12:02:36 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:10621941
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)

Received Date : 17-Jan-2012
Revised Date : 17-Jan-2012
Accepted Date : 15-Feb-2012
Article type : Primary Research Articles

Title: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange

Running Title: Modeling interannual land CO_2 flux

T.F. Keenan[1], Ian Baker[2], Alan Barr[3], Philippe Ciais[4], Ken Davis[5], Michael Dietze[6], Danillo Dragoni[7], Christopher M Gough[8], Robert Grant[9], David Hollinger[10], Koen Hufkens[11], Ben Poulter[12], Harry McCaughey[13], Brett Rackza[14], Youngryel Ryu[15], Kevin Schaefer[16], Hanqin Tian[17], Hans Verbeeck[18], Maosheng Zhao[19], Andrew D. Richardson[1]

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge MA, USA

[2] Atmospheric Science Department, Colorado State University, Fort Collins, Colorado, USA.

[3] Climate Research Division, Atmospheric Science and Technology Directorate, Saskatoon, Saskatchewan, Canada.

[4] Laboratoire des Sciences du Climat et de l'Environnement, CE Orme des Merisiers, Gif sur Yvette, France

[5] Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.

[6] Department of Plant Biology, University of Illinois Urbana Champaign, Urbana, Illinois, USA

[7] Department of Geography, Indiana University, Bloomington, Indiana, USA

This is an Accepted Article that has been peer-reviewed and approved for publication in the *Global Change Biology*, but has yet to undergo copy-editing and proof correction. Please cite this article as an "Accepted Article"; doi: 10.1111/j.1365-2486.2012.02678.x

[8] Department of Biology, Virginia Commonwealth University, VA, USA

[9] Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada

[10] Northern Research Station, USDA Forest Service, Durham, New Hampshire, USA.

[11]Department of Geography and Environment, Boston University, MA USA

[12] Laboratoire des Sciences du Climat et de l'Environnement, CE Orme des Merisiers, Gif sur Yvette, France

[13] Department of Geography, Queen's University, Kingston, Ontario, Canada

[14] Department of Meteorology, Penn State University, Pennsyvannia NY, USA

[15] Department of Landscape Architecture and Rural System Engineering, Seoul National University, South Korea.

[16] National Snow and Ice Data Center, University of Colorado at Boulder, Boulder, Colorado, USA.

[17] School of Forestry and Wildlife Sciences, Auburn University, Auburn, Alabama, USA.

[18] Laboratory of Plant Ecology, Ghent University, Ghent, Belgium

[19] Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT, USA.

Corresponding author:

T.F. Keenan

Phone: +1 (617) 496-0825

Email: tkeenan@oeb.harvard.edu

Key Words: process-based models, interannual variability, carbon fluxes, biosphere-atmosphere interaction, carbon sequestration, remote sensing

Paper type: Primary research article

Abstract

Interannual variability in biosphere-atmosphere exchange of CO_2 is driven by a diverse range of biotic and abiotic factors. Replicating this variability thus represents the 'acid test' for terrestrial biosphere models. Although such models are commonly used to project responses to both normal and anomalous variability in climate, they are rarely tested explicitly against inter-annual variability in observations. Here, using standardized data from the North American Carbon Program, we assess the performance of 16 terrestrial biosphere models and 3 remote sensing products against long-term measurements of biosphere-atmosphere CO_2 exchange made with eddy-covariance flux towers at 11 forested sites in North America. Instead of focusing on model-data agreement we take a systematic, variability-oriented, approach and show that although the models tend to reproduce the mean magnitude of the observed annual flux variability, they fail to reproduce the timing. Large biases in modeled annual means are evident for all models. Observed interannual variability is found to commonly be on the order of magnitude of the mean fluxes. None of the models consistently reproduce observed interannual variability within measurement uncertainty. Underrepresentation of variability in spring phenology, soil thaw and snowpack melting, and difficulties in reproducing the lagged response to extreme climatic events are identified as systematic errors, common to all models included in this study.

Introduction

The terrestrial biosphere acts as a net sink for atmospheric CO_2, with global forests absorbing on average 4 Pg C yr^{-1} (Pan et al., 2011), which, excluding deforestation, offsets roughly half of all anthropogenic emissions from fossil fuel burning and cement production (Pan et al., 2011). Interannual variability in this sink is often on the order of magnitude of the mean (e.g., Zeng et al., 2005; Reichstein et al., 2007a; Pan et al., 2011), and drives interannual variability in the growth rate of atmospheric CO_2 (Bosquet et al., 2000; Knorr et al., 2007). Carbon fluxes in forest ecosystems are tightly coupled to climate (Richardson et al., 2007; Piao et al., 2008; Chen et al., 2009; Dragoni et al., 2011), and anomalous climatic signals generally drive the observed variability in their sink strength (Dunn et al., 2007; Desai, 2010; Le Maire et al., 2010). Such signals tend to affect photosynthesis and respiration (the two processes which determine net ecosystem carbon exchange) to different extents (Richardson et al., 2007; Luyssaert et al., 2007), and therefore provide an excellent test-bed to assess the skill of terrestrial biosphere models.

Terrestrial biosphere models are the primary tools used for predicting the impact of climate variability on terrestrial carbon fluxes. Built around the philosophy that a blend of mechanistic and semi-empirical descriptions can capture functional responses to environmental drivers, they have been used in conjunction with remote sensing products (Zhao and Running, 2010) and data mining tools (Papale and Valentini, 2003) to provide regional and global estimates of terrestrial carbon cycling (e.g., Friend et al., 2010; Beer et al., 2010). They are also commonly used to quantify terrestrial responses

to climatic variability, including anomalies and extreme events (Ciais et al., 2005; Reichstein et al., 2007; Vetter et al., 2008; Zhao and Running, 2010). Future model projections of the response of terrestrial carbon cycling to climate change (Heimann and Reichstein, 2008) are necessary to inform policy (IPCC, 2007), though current models show very divergent sensitivity to long-term changes in climate (Friedlingstein et al., 2006).

Biogeochemical models are often shown to capture diel and seasonal dynamics reasonably well (e.g., Braswell et al., 2005). This is not surprising, given the pronounced diurnal and seasonal cycles of climatic drivers. Over yearly and longer time scales, however, studies show poor model performance at reproducing gross fluxes and carbon budgets (e.g., Hanson et al., 2004; Braswell et al., 2005; Siqueira et al. 2006; Richardson et al., 2007; Urbanski et al. 2007). Such comparison studies are typically restricted to a limited number of models and sites, and a relatively short time series length. Nonetheless, the results suggest that although the response of terrestrial ecosystems to mean climatic drivers is relatively well captured, sensitivity to the impact of variability in climatic drivers may not be, leading to the accumulation of high frequency model error (e.g., Dietze et al., 2012) over longer time scales (Schwalm et al., 2010). No study, however, has yet identified systematic errors in model sensitivity to climatic variability.

In this analysis, we use 16 terrestrial biosphere models and 3 remote sensing products, along with eddy-covariance data from a range of sites included in the North American Carbon Program interim site synthesis, to assess model ability to reproduce observed variability in carbon fluxes. We examine the frequency distribution of temporal

anomalies in net ecosystem exchange (NEE), gross primary productivity (GPP), and ecosystem respiration (RE), for two plant functional types. We first assess individual model performance on an annual/interannual scale. As interannual variability can be driven by 'critical' periods within a year (Le Maire et al., 2010), we examine monthly systematic model errors (errors consistent across all models and sites). By using data from sites with a regionally coherent anomalous year, we then assess the possible role of extreme within-year climatic events and lagged effects on model performance for interannual variability in terrestrial carbon cycling.

Methods

All models and data used were obtained through the North American Carbon Program (NACP) (http://www.nacarbon.org/nacp/). To allow for an ensemble approach and reduce the potential for spurious variability, we selected only sites with at least 5 years of data, from plant functional types that were represented by at least 3 such sites. This resulted in a total of 11 forested sites distributed through North America (Table 1). Of those, 6 were deciduous broadleaf, and 5 evergreen needleleaf. This gave a total of 91 site-years for the analysis.

Eddy-covariance flux data (produced by AmeriFlux and Fluxnet-Canada investigators) from the 11 selected sites was processed according to a common protocol from the NACP site level interim synthesis (http://www.nacarbon.org/nacp/). The observed NEE were corrected for storage, despiked (i.e., outlying values removed), and filtered to remove conditions of low turbulence (friction velocity filtered). Flux error estimates were calculated (Barr et al., 2009) by combining random uncertainty (calculated

following Richardson and Hollinger (2007)) and uncertainty due to the selection of the friction velocity threshold (Barr et al., 2009). Observed monthly and annual NEE values were then calculated using gap-filled data from each site (Barr et al., 2009). The gap-filled NEE values were also partitioned to gross ecosystem photosynthesis (GPP) and ecosystem respiration (RE). Multiple approaches were used in order to quantify additional uncertainty introduced by the partitioning (Desai et al., 2008; Barr et al. 2009).

Gaps in the meteorological forcing data occurred due to instrument failure or quality control. Such gaps were filled using the nearest available climate station in the National Climatic Data Center's Global Surface Summary of the Day (NCDC-GSOD) database. Gaps at sites where no such data was available were filled using DAYMET (Thornton et al., 1997). Daily NCDC-GSOD and DAYMET data were temporally downscaled to hourly or half-hourly values (see Ricciuto et al., 2009 for details) (http://nacp.ornl.gov/docs/Site_Synthesis_Protocol_v7.pdf).

Sixteen terrestrial biosphere models (Table 2) were run at the sites for the period of available measurements (Table 1). The terrestrial biosphere models simulated carbon cycling with process-based formulations of varying detail for the component carbon fluxes of photosynthesis and respiration. Simulated NEE was based on model specific runs using gap filled observed weather at each site and locally observed values of soil texture according to a standard protocol (Ricciuto et al., 2009). Each model used species or plant functional type specific parameterizations as defined by the individual model teams, with the exception of LoTEC where parameters were optimized using data assimilation (Ricciuto et al., 2008). Three remote sensing products of terrestrial gross

primary productivity (MODISc5 (Running et al., 2000), MODISc5.1 (Zhao et al., 2005), BESS (Ryu et al., 2012)), not included in the North American Carbon Program, were also used to provide annual estimates of GPP for each site.

In order to assess interannual variability, we normalized the measured/modeled values of NEE, GPP and RE by subtracting the long-term calendar year measured/modeled mean for each site from individual site-year flux values, giving F_{obs} and F_{sim} for each flux and year. By comparing the long-term calendar year mean of measured and modeled fluxes, we also identified biases in model estimates of NEE, GPP and RE. Model-data agreement for interannual variability in annual flux sums was assessed in terms of the normalized root mean squared error (NRMSE) and the χ^2 statistic.

The NRMSE is the root mean square error of model-data mismatch normalized by the magnitude of observed variability at each site:

$$NRMSE_{ijk} = \frac{\sqrt{\frac{1}{n}\sum_l (F_{obs} - F_{sim})^2}}{\sigma(F_{obs})} \quad (1)$$

where F represents the observed (*obs*) or modeled (*sim*) value of a particular flux, i (NEE, GPP, or RE), for a particular year, l. Note that each flux here is represented as the interannual variability (F_{obs} and F_{sim}), not the mean flux. $\sigma(F_{obs})$ is the standard deviation of observed interannual variability at site k. NRMSE values are calculated for each model j at site k. The NRMSE thus reports the mean difference between the simulated and observed flux, relative to the variability in the observed flux.

The χ^2 statistic complements the NRMSE by incorporating measurement error. Here it is calculated for each model and PFT as the squared residual between paired model and data points for each flux (after normalization to the long-term mean as described above), relative to the observational error:

$$\chi^2_{ijk} = \frac{1}{n}\sum_{ijkl}\left(\frac{F_{obs} - F_{sim}}{2\delta(F_{obs})}\right)^2 \quad (2)$$

where $\delta(F_{obs})$ is uncertainty related to the annual observed value of that flux, 2 normalizes the uncertainty in the observed flux to correspond to a 95% confidence interval. A χ^2 value of less than one indicates that the model agrees with the data relative to data uncertainty. I.e., interannual variability for a model with a χ^2 value of less than one will always fall within 1 standard deviation of data error. Above one, the χ^2 scales model error relative to observation uncertainty.

Interannual variability in observed fluxes commonly stems from specific, short-lived, periods of anomalous fluxes within the year (Krishnan et al., 2008, 2009; Chen et al., 2009; Le Maire et al., 2010). We therefore also assessed model performance for variability on a monthly scale. The variability of monthly fluxes between years was calculated in the same way as annual variability, as the difference between the observed or modeled monthly value and the associated long-term mean for the month in question. By differencing the observed and predicted monthly variability (here termed variance residuals) specific periods during the year at which the models under- or over-represent the observed monthly variability can be identified. We define periods of systematic model error (statistically common to all models) as times when all models show the same-signed bias in variance residuals with 95% confidence. We also assess persistent biases, which are mean biases of more than one month in duration that are not necessarily systematic.

Extreme climatic events, detectable as regionally coherent deviations outside the normal range of variability, provide a strong test of model performance. We identified one such event in our database. At three sites in mid-western Canada, mean spring monthly

temperatures in 2002 were between 8 and 10 °C below the long-term mean. We used this event to assess model skill and identify systematic model error.

Results

Biases and the magnitude of variability:

In order to quantify interannual variability, we normalizing all models and data by subtracting respective mean annual totals from individual annual totals. This process identified considerable biases in model estimates of all total annual fluxes at all sites (Fig. 1). In particular, biases in annual NEE were commonly of similar magnitude to mean observed annual NEE fluxes. The majority of models were biased towards underestimating ecosystem carbon uptake for both deciduous and evergreen sites (Fig. 1). Note that biases here are reported relative to the observed mean NEE for each site, and therefore have the potential to be particularly larger for sites with low mean annual NEE. See Table 1 for per-site mean annual NEE values.

The magnitude of modeled interannual variability in each annual flux was on average of the same order of that observed (Fig. 2). A large range in model performance was evident (Table 3), but in general the models proved 'flexible' enough to reproduce the observed range of variability. Observed interannual variability in NEE for deciduous broadleaved forests was twice that of evergreen needle-leaved forests, a distinction only reproduced by six of the included models. The magnitude of interannual variability in both GPP and RE was greater (55%, 23%) in deciduous broadleaved forests than in evergreen needle-leaved forests. The remote sensing products, however, consistently predicted higher GPP variability in evergreen than in deciduous forests.

Statistical performance of models on an interannual scale

Although the mean magnitude of model variability on the interannual scale was similar to the mean observed magnitude of variability (Fig. 2), all models fell outside of the data error of the observed for individual site-years (>1 χ^2, Fig. 3, S1). This means that the general magnitude of interannual variability was well reproduced, but not the timing. Interannual variability in the annual net ecosystem exchange of evergreen forests was better simulated on average than deciduous forests (Fig. 3). A larger range of model performance was observed for variability in annual GPP than that of RE. Our results suggest that on average the inability of models to match the timing of observed variability in GPP is the main cause of errors in the simulation of interannual variability in NEE, though this is very model dependent (Fig. 3). The MODISc5 remote sensing product performed worse than the average model (Fig. 3, Fig S2). The MODISc5.1 data product proved to be a large improvement over the MODISc5 GPP product. The BESS remote sensing product, a process based model interpretation of remotely sensed data (Ryu et al., 2012), performed better than either MODIS product for deciduous forests, though that was not the case for evergreens. Although process based models of different types were represented (e.g., light use efficiency vs enzyme kinetic model for GPP, Table 2) no model characteristic performed statistically better than any other (data not shown). This could be due to the limited number of models with particular characteristics.

Variability within the year

The models showed persistent systematic biases (see definition in Methods section) for monthly flux variability. In deciduous forests, models consistently underestimating monthly variability in NEE throughout spring (May and June) (Fig. 4). Model underestimation of variability in deciduous spring NEE fluxes was mostly due to underrepresentation of variability in spring GPP (Fig. 4). A smaller peak in the deciduous GPP variance residuals (predicted monthly variability – observed monthly variability) was also evident in September and October. Variability in deciduous RE showed no bias that was consistent across all models.

Systematic underrepresentation of monthly variability during May was also evident for evergreen forests (Fig. 4). Here, however, model error for NEE was dominated by the lack of variability in modeled RE during spring. Although evergreen forests do not exhibit the marked phenological transitions observed in deciduous forests, all evergreen forests included in this study maintain a large snow-pack throughout winter. Persistent, non-systematic biases were evidenced throughout the year, in particular an overestimation of winter variability in evergreen NEE and GPP, and a persistent underestimation of variability in evergreen RE during the summer.

Response to anomalous climate forcing

Three sites (CA-Ojp, CA-Obs, CA-Oas; see site description Table 1) experienced a regionally coherent extreme climatic event during 2002, where monthly mean temperatures were between 8 (CA-Ojp, CA-Obs) and 10 °C (CA-Oas) below the long-term mean. The anomaly largely affected canopy GPP at all three sites, and to a lesser extent RE (Fig. 5). At CA-Ojp and CA-Obs, anomalously low temperatures during the month of April lowered observed GPP by more than twice the normal range of

variability for that month (Fig. 5). The models accurately captured the drop in productivity, with the mean of all model projections capturing both the sign and the magnitude of the April GPP anomaly at both sites. The temperature anomaly was observed at CA-Oas one month later, and again the models accurately reproduced the observed magnitude in anomalous GPP. During the following May, June and July, observed temperature remained colder than normal but returned to within the normal range of variability for the three sites. Observed GPP, however, remained anomalously low during those months and did not return to within the normal range of variability until July at each site. This lagged effect between anomalous climate forcing and resulting fluxes was not well reproduced by the models. At CA-Obs, all models returned to within the normal range of GPP variability in the month directly following the temperature anomaly. The same behavior was apparent at CA-Ojp and CA-Oas, though the average model GPP estimates remained just outside the normal range of variability due to persistent low temperatures. The extended period of low productivity in CA-Oas may be in part also due to consistently low precipitation during the year.

A similar, though smaller anomaly pattern was observable for RE (Fig. 5). Low spring temperatures caused a prolonged anomaly of low ecosystem respiration. The models tended to overestimate the reduction in RE as a result of the colder temperatures. After the initial anomaly, RE as estimated by eddy-covariance measurements took a few months to return to within the normal range of variability. Modeled RE quickly returned to 'normal' at CA-Ojp and CA-Obs. Temporal dynamics at CA-Oas differed from those of the other two sites due to the additional pressure of persistently low temperatures and precipitation during the year.

Discussion

This analysis has shown that, although capable of reproducing the magnitude of interannual variability, terrestrial biosphere models are not consistent with the timing of observations of interannual variability in surface-atmosphere exchanges of CO_2 at mid-latitude forests over North America. By examining interannual variability in measured and modeled monthly fluxes, we show that all the models used for the NACP interim site synthesis systematically fail to reproduce observed variability during spring. Underestimation of spring variability is largest for GPP in deciduous forests, and RE for evergreens, suggesting different processes may be responsible for plant functional type specific model error.

It has been shown that terrestrial biosphere models are typically unable to adequately explain the observed interannual variability in deciduous canopy phenology (Richardson et al., 2012), and that variability in spring GPP often drives observed interannual variability in net ecosystem exchange (Krishnan et al., 2008, 2009). Here we show that this is a systematic cause of the low agreement between modeled and observed interannual variability in terrestrial carbon fluxes.

In a similar fashion, it has been shown that the current available models of snow pack dynamics perform poorly for both spatial and interannual variability. Rutter et al. (2009) tested 33 models of snowpack dynamics across a range of sites, and found that although a model could perform well when tuned to a particular site-year, this did not transfer to good performance for other years at the same site, or other sites. Interactions between snowmelt, soil thaw and water table depth are known to directly affect interannual

variability in NEE (Goulden et al., 1998; Dunn et al., 2007; Hu et al., 2010). Results here suggest that this may be a direct systematic contributor to the low agreement between observed and modeled interannual variability in net ecosystem carbon exchange, in particular for evergreen sites. These results do not imply, however, that a lack of phenological variability in canopy or soil dynamics are necessarily the main culprits for the lack of agreement between the observations and output from any one model, as individual models showed large persistent biases at other times of the year (Fig. 3).

The remote sensing products performed comparably to the average process-based model when assessed against interannual variability in GPP. The MODISc5.1 data set is a post-processed version of the MODISc5 data set where corrections are made for poor quality driver data (Zhao et al., 2005). The remote sensing products, which are themselves models, are driven by a global daily meteorological reanalysis dataset not site-specific meteorology and the uncertainties in the meteorological reanalysis can introduce biases in GPP estimates (Zhao et al., 2006). Although estimates of GPP based on remote sensing have been used to evaluate process-based models (e.g., Poulter et al., 2009), results here suggest that estimates of interannual variability from both approaches are subject to similar magnitudes of error.

Although there was a general tendency for the models to persistently underestimate flux variability in summer, it should be noted that the flux data are subject to random error roughly in proportion to the size of the flux (Richardson & Hollinger, 2007; Richardson et al., 2008). Even if the model were perfect, modeled variability should be smaller than that observed. Carbon fluxes are typically higher in the summer, and subject to larger

uncertainty. The apparent higher variability in the data during summer could therefore be due to random errors in the flux measurements generating larger variability in monthly totals.

The analysis of model responses to the regionally coherent climatic anomaly of spring 2002 suggests that models have the potential to correctly reproduce the magnitude of instantaneous biological response to climate anomalies (Desai, 2010). Although the models accurately captured the direct effect of an isolated climate extreme, the models included here failed to accurately reproduce lagged effects of climate anomalies on both gross primary production and ecosystem respiration. Lagged effects of climate variability on ecosystem function have previously been reported (Gough et al., 2009), and our results suggest that such lagged effects are not well incorporated into models. The nature of such lagged effects depends on the type of climatic anomaly. Spring frosts (Gu et al., 2008; Marino et al., 2011), for example, are known to directly effect canopy structure, an aspect not currently accounted for in models. The affects of other disturbances, such as ice storms, strong winds and insect outbreaks are known to be poorly represented by models (Liu et al., 2011) and affect long-term carbon dynamics. . Lagged effects unrelated to disturbances can be caused by changes in nutrient cycling (Richardson et al., 2009) or changes in the size of carbon pools such as litter (Rocha et al., 2008), or non-structural carbohydrates (Gough et al., 2009) due to climatic conditions in previous years. Model aspects related to lagged and cumulative effects can be improved through direct comparisons with observations (e.g., Keenan et al., 2009), though many related issues remain (Liu et al., 2011). Although lagged effects are apparent at the three sites showing a coherent regional extreme event, we did not detect similar lagged events for other climatic anomalies in the database. This is likely due to

two confounding effects: that smaller anomalous climate signals do not produce lagged (on monthly scales) ecosystem effects, and that biotic effects could play a role in driving some of the interannual variability in observed fluxes (Richardson et al., 2007).

Open questions remain as to the proportion of interannual variability in land-atmosphere carbon exchange that is directly explainable by variability in climate (Hui et al., 2003; Polley et al., 2010; Richarsdon et al., 2007). Controls on interannual variability can also be manifest in the form of functional changes in the ecosystem, or lagged effects on pool sizes and dynamics. By contrasting the interannual performance of a simple empirical model with fixed parameters against the same model with interannually varying parameters, Richardson et al. (2007) reported that forest functional change at a spruce forest was responsible for 55% of interannual variations in land-atmosphere CO_2 exchange. i.e., 45% of the observed variability was not explainable by the direct impacts of climate. Polley et al. (2010) used a similar approach to determine a significant contribution of ecosystem functional change to interannual variability in grasslands. Using an optimized process-based model, however, Desai (2010) found that 81% of interannual variability in annual CO_2 exchange could be explained by variability in climate for five mature hardwood forests, a value that likely underestimates model performance given that it does not account for observational error. This result supports multi-site synthesis efforts that show that ~79% of interannual variability for mid-altitude deciduous broadleaved forests can be explained by variability in temperature (Yuan et al., 2009). Clearly a detailed assessment of the relative roles of climate and functional change on the interannual variability of CO_2 flux across a wide range of sites and climate zones is needed.

We could not distinguish any model structure or characteristic (see Table 2) that tended to give a better model performance. All models are subject to errors resulting from both parameter choice (parameter mis-specification) and model structure (process mis-representation) (Keenan et al., 2011). The fact that no model structure proved consistently better suggests that parameter error was excessively large. In future efforts, model-data fusion techniques (Wang et al., 2009; Keenan et al., 2011) could aid in reducing the relative magnitude of parameter errors, thus allowing for a more rigorous assessment of model structural differences.

Our estimates of the magnitude of observed interannual variability in land-atmosphere CO_2 exchange (DBF: ~85 gC m^{-2}; EVG: 44 gC m^{-2}, Table. 3) are roughly 50% and 33% of the mean flux respectively. Given that this represents one standard deviation about the mean, variability in ecosystem carbon uptake is commonly on the order of magnitude of the mean. This supports previous results from single sites (Reichstein et al., 2007a), and modeling studies (Zeng et al., 2005), across the range of sites included here. Variability in GPP has been found to be the main contributor to variability in NEE for a variety of terrestrial ecosystems (Luyssaert et al., 2007). Here, we show that for deciduous forests, the interannual variability in GPP is on average 26% greater than that of RE (Table 3). All though on average both GPP and RE show a similar magnitude of variability at the evergreen needle-leaf forest sites, four of the six evergreen sites had higher variability in GPP. This suggests that variability in GPP dominates variability in NEE in deciduous mid-latitude forests, though this rule is not applicable to all sites included here.

Using 91 site-years at 11 long-term eddy-covariance forest sites, we show that terrestrial biosphere models have difficulty in simulating land-atmosphere CO_2 exchange at annual and interannual time scales, with the potential for large biases on the interannual scale, and incorrect simulation of the timing of interannual variability. Instead of focusing on model-data agreement, we present a variability-oriented approach of diagnosing systematic and persistent model-data disagreement. Given that studies of the impact of climate variability on terrestrial fluxes are likely to reveal a more informative picture of biosphere-atmosphere interactions (Le Maire et al., 2010), such a variability orientated approach should greatly aid modeling teams in future model assessment and development. Our results highlight three potential mechanisms - spring canopy phenology, soil thaw and the melting of the snow pack, and lagged effects - common to all models included in the study, which contribute to the low agreement between the models and the observed interannual variability in land-atmosphere CO_2 exchange. Addressing these issues in future model efforts will be the first step towards improving the sensitivity of models to climatic variability on interannual time scales.

Acknowledgments

We thank all those involved in the NACP Site Synthesis, in particular the modeling teams, and the AmeriFlux and Fluxnet-Canada Research Network/Canadian Carbon Program PIs who provided the data on which this analysis is based. We also thank the funding agencies that have supported model development and long-term flux measurements. TFK and ADR acknowledge support from Office of Science (BER), US Department of Energy, through the Northeastern Regional Center of the National Institute for Climatic Change Research. TFK acknowledges Philippe Peylin for useful comments on an earlier draft.

References

Arain, M. A., F. Yuan, and T. Andrew Black (2006), Soil–plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada, *Agricultural and Forest Meteorology*, *140*(1-4), 171-192, doi:10.1016/j.agrformet.2006.03.021.

Baker, I. T., L. Prihodko, A. S. Denning, M. Goulden, S. Miller, and H. R. da Rocha (2008), Seasonal drought stress in the Amazon: Reconciling models and observations, *Journal of Geophysical Research*, *113*, G00B01, doi:10.1029/2007JG000644.

Barr, A. G., T. a Black, E. H. Hogg, N. Kljun, K. Morgenstern, and Z. Nesic (2004), Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production, *Agricultural and Forest Meteorology*, *126*(3-4), 237-255, doi:10.1016/j.agrformet.2004.06.011.

Barr, A., D. Hollinger, and A. D. Richardson (2009), CO2 flux measurement uncertainty estimates for NACP, *Eos Trans. AGU Fall Meet. Suppl.*, *90*(52), Abstract B54A□04.

Beer, C. et al. (2010), Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate., *Science (New York, N.Y.)*, *329*(5993), 834-8, doi:10.1126/science.1184984.

Bousquet, P., P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980., *Science (New York, N.Y.)*, *290*(5495), 1342-7.

Bradford, J. B., R. A. Birdsey, L. A. Joyce, and M. G. Ryan (2008), Tree age, disturbance history, and carbon stocks and fluxes in subalpine Rocky Mountain forests, *Global Change Biology*, *14*(12), 2882-2897, doi:10.1111/j.1365-2486.2008.01686.x.

Braswell, B. H., W. J. Sacks, E. Linder, and D. S. Schimel (2005), Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem

exchange observations, *Global Change Biology*, *11*(2), 335-355, doi:10.1111/j.1365-2486.2005.00897.x

Chen, B., T. A. Black, N. C. Coops, P. Krishnan, R. Jassal, C. Brümmer, and Z. Nesic (2009), Seasonal controls on interannual variability in carbon dioxide exchange of a near-end-of rotation Douglas-fir stand in the Pacific Northwest, 1997-2006, *Global Change Biology*, *15*(8), 1962-1981, doi:10.1111/j.1365-2486.2008.01832.x.

Ciais, P. et al. (2005), Europe-wide reduction in primary productivity caused by the heat and drought in 2003., *Nature*, *437*(7058), 529-33, doi:10.1038/nature03972.

Davis, K. J., P. S. Bakwin, C. Yi, B. W. Berger, C. Zhao, R. M. Teclaw, and J. G. Isebrands (2003), The annual cycles of CO2 and H2O exchange over a northern mixed forest as observed from a very tall tower, *Global Change Biology*, *9*, 1278-1293, doi:10.1046/j.1365-2486.2003.00672.x.

Desai, A. R. (2010), Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape, *Journal of Geophysical Research*, *115*, 1-13, doi:10.1029/2010JG001423.

Dragoni, D., H. P. Schmid, C. A. Wayson, H. Potter, C. S. B. Grimmond, and J. C. Randolph (2011), Evidence of increased net ecosystem productivity associated with a longer vegetated season in a deciduous forest in south-central Indiana , USA, *Global Change Biology*, *17*, 886-897, doi:10.1111/j.1365-2486.2010.02281.x.

Dunn, A. L., C. C. Barford, S. C. Wofsy, M. L. Goulden, and B. C. Daube (2007), A long-term record of carbon exchange in a boreal black spruce forest: means, responses to interannual variability, and decadal trends, *Global Change Biology*, *13*(3), 577-590, doi:10.1111/j.1365-2486.2006.01221.x.

Friedlingstein, P. et al. (2006), Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison, *Journal of Climate*, *19*(14), 3337–3353.

Friend, A. D. (2010), Terrestrial plant production and climate change., *Journal of experimental botany*, *61*(5), 1293-309, doi:10.1093/jxb/erq019.

Gough, C. M., C. E. Flower, C. S. Vogel, D. Dragoni, and P. S. Curtis (2009), Whole-ecosystem labile carbon production in a north temperate deciduous forest, *Agricultural and Forest Meteorology*, *149*, 1531-1540, doi:10.1016/j.agrformet.2009.04.006.

Goulden, M. L. et al. (1998), Sensitivity of Boreal Forest Carbon Balance to Soil Thaw, *Science*, *279*(5348), 214-217, doi:10.1126/science.279.5348.214.

Grant, R. F., A. Arain, V. Arora, A. Barr, T. A. Black, J. Chen, S. Wang, F. Yuan, and Y. Zhang (2005), Intercomparison of techniques to model high temperature effects on CO2 and energy exchange in temperate and boreal coniferous forests, *Ecological Modelling*, *188*, 217-252, doi:10.1016/j.ecolmodel.2005.01.060.

Griffis, T., T. A. Black, K. Morgenstern, A. G. Barr, Z. Nesic, G. B. Drewitt, D. Gaumont-Guay, and H. McCaughey (2003), Ecophysiological controls on the carbon balances of three southern boreal forests, *Agricultural and Forest Meteorology*, *117*, 53-71, doi:10.1016/S0168-1923(03)00023-6.

Gu, L., P. J. Hanson, W. M. Post, D. P. Kaiser, B. Yang, R. Nemani, S. G. Pallardy, and T. Meyers (2008), The 2007 Eastern US Spring Freeze: Increased Cold Damage in a Warming World, BioScience, 58(3), 253, doi:10.1641/B580311.

Hanson, P. J. et al. (2004), Oak Forest Carbon and Water Simulations: Model Intercomparisons and Evaluations Against Independent Data, *Ecological Monographs*, *74*(3), 443-489, doi:10.1890/03-4049.

Heimann, M., and M. Reichstein (2008), Terrestrial ecosystem carbon dynamics and climate feedbacks, *Nature*, *451*(7176), 289-92, doi:10.1038/nature06591

Hu, J., D. J. P. Moore, S. P. Burns, and R. K. Monson (2010), Longer growing seasons lead to less carbon sequestration by a subalpine forest, *Global Change Biology*, *16*(2), 771-783, doi:10.1111/j.1365-2486.2009.01967.x.

Hui, D., Y. Luo, and G. Katul (2003), Partitioning interannual variability in net ecosystem exchange between climatic variability and functional change., Tree physiology, 23(7), 433-42.

Keenan, T., R. Garcia, A.D. Friend, S. Zaehle, C. Gracia, and S. Sabate (2009), Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO2 and water fluxes through combined in situ measurements and ecosystem modelling, Biogeosciences, 6(8), 1423-1444, doi:10.5194/bg-6-1423-2009.

Keenan, T. F., M. S. Carbone, M. Reichstein, and A. D. Richardson (2011), The model-data fusion pitfall: assuming certainty in an uncertain world, Oecologia, (167), 587-597, doi:10.1007/s00442-011-2106-x.

Knorr, W., N. Gobron, M. Scholze, T. Kaminski, R. Schnur, and B. Pinty (2007), Impact of terrestrial biosphere carbon exchanges on the anomalous CO2 increase in 2002–2003, *Geophysical Research Letters*, *34*(9), 1-6, doi:10.1029/2006GL029019.

Krinner, G., N. Viovy, N. de Noblet-Ducoudre, J. Ogee, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Krishnan, P., T. A. Black, A. G. Barr, N. J. Grant, D. Gaumont-Guay, and Z. Nesic (2008), Factors controlling the interannual variability in the carbon balance of a southern boreal black spruce forest, *Journal of Geophysical Research*, *113*(D9), 1-16, doi:10.1029/2007JD008965.

Krishnan, P., T. A. Black, R. S. Jassal, B. Chen, and Z. Nesic (2009), Interannual variability of the carbon balance of three different-aged Douglas-fir stands in the Pacific Northwest, *Journal of Geophysical Research*, *114*(G4), 1-18, doi:10.1029/2008JG000912

Liu, J., J. Chen, J. Cihlar, and W. Chen (1999), Net primary productivity distribution in the BOREAS region from a process model using satellite and surface data, *Journal of Geophysical Research*, *104*(D22), 27375-27754.

Liu, S., N. Bliss, E. Sundquist, and T. G. Hungtington (2003), Modeling carbon dynamics in vegetation and soil under the impact of soil erosion and deposition, *Global Biogeochemical Cycles*, *17*(2), 1074, doi:10.1029/2002GB002010.

Liu, S. et al. (2011), Simulating the impacts of disturbances on forest carbon cycling in North America: Processes, data, models, and challenges, Journal of Geophysical Research, 116, 1-22, doi:10.1029/2010JG001585.

Luyssaert, S. et al. (2007), Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes, *Global Change Biology*, *13*(10), 2110-2127, doi:10.1111/j.1365-2486.2007.01432.x.

Marino, G. P., D. P. Kaiser, L. Gu, and D. M. Ricciuto (2011), Reconstruction of false spring occurrences over the southeastern United States, 1901-2007: an increasing risk of spring freeze damage?, Environmental Research Letters, 6(2), 024015, doi:10.1088/1748-9326/6/2/024015.

le Maire, G. et al. (2010), Detecting the critical periods that underpin interannual fluctuations in the carbon balance of European forests, *Journal of Geophysical Research*, *115*, 1-16, doi:10.1029/2009JG001244

Medvigy, D., S. C. Wofsy, J. W. Munger, D. Y. Hollinger, and P. R. Moorcroft (2009), Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2, *Journal of Geophysical Research*, *114*(G1), 1-21, doi:10.1029/2008JG000812.

Pan, Y. et al. (2011), A Large and Persistent Carbon Sink in the World's Forests, *Science*, *14*(July), 1-11, doi:10.1126/science.1201609.

Papale, D., and R. Valentini (2003), A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, *Global Change Biology*, *9*(4), 525-535, doi:10.1046/j.1365-2486.2003.00609.x

Peichl, M., and M. A. Arain (2007), Allometry and partitioning of above- and belowground tree biomass in an age-sequence of white pine forests, *Forest Ecology and Management*, *253*(1-3), 68-80, doi:10.1016/j.foreco.2007.07.003

Piao, S. et al. (2008), Net carbon dioxide losses of northern ecosystems in response to autumn warming., *Nature*, *451*(7174), 49-52, doi:10.1038/nature06444

Polley, H. W. et al. (2010), Physiological and environmental regulation of interannual variability in CO2 exchange on rangelands in the western United States, Global Change Biology, 16(3), 990-1002, doi:10.1111/j.1365-2486.2009.01966.x.

Reichstein, M. et al. (2007), Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites, *Geophysical Research Letters*, *34*(1), doi:10.1029/2006GL027880.

Riccuito, D. M., K. A.W., L. Gu, and W. M. Post (2008), Estimates of terrestrial carbon cycle model parameters by assimilation of FLUXNET data: Do parameter variations cause bias in regional flux estimates?, *Eos Trans. AGU,Fall Meet. Suppl.,*, *89*(53), Abstract B54A□03.

Riccuito, D. M., P. E. Thornton, K. Schaefer, R. B. Cook, and K. J. Davis (2009), How uncertainty in gap□filled meteorological input forcing at eddy covariance sites impacts modeled carbon and energy flux, *Eos Trans. AGU, Fall Meet. Suppl.,*, *90*(52), Abstract B54A□03.

Richardson, A. D., D. Y. Hollinger, J. D. Aber, S. V. Ollinger, and B. H. Braswell (2007), Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biology*, *13*(4), 788-803, doi:10.1111/j.1365-2486.2007.01330.x.

Richardson, A. D., D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, J. O. Keefe, and J. O'keefe (2009), Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, *Tree physiology*, *29*(3), 321-31, doi:10.1093/treephys/tpn040.

Richardson, A. D. et al. (2008), Statistical properties of random CO2 flux measurement uncertainty inferred from model residuals, *Agricultural and Forest Meteorology*, *148*(1), 38-50, doi:10.1016/j.agrformet.2007.09.001.

Richardson, A. D., and D. Y. Hollinger (2007), A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO2 flux record, *Agricultural and Forest Meteorology*, *147*(3-4), 199-208, doi:10.1016/j.agrformet.2007.06.004.

Richardson, A. D. et al. (in press), Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program Site Synthesis, Global Change Biology, n/a-n/a, doi:10.1111/j.1365-2486.2011.02562.x.

Rocha, a V., D. L. Potts, and M. L. Goulden (2008), Standing litter as a driver of interannual CO_2 exchange variability in a freshwater marsh, Journal of Geophysical Research, 113(G4), 1-10, doi:10.1029/2008JG000713.

Running, S., R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto (2004), A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production, *BioScience*, *54*(6), 547-560, doi:10.1641/0006-3568(2004)054.

Rutter, N. et al. (2009), Evaluation of forest snow processes models (SnowMIP2), *Journal of Geophysical Research*, *114*(D6), D06111, doi:10.1029/2008JD011063.

Ryu, Y. et al. (2012), Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. Global Biogeochemical Cycles, doi:10.1029/2011GB004053.

Schaefer, K., T. Zhang, A. G. Slater, L. Lu, A. Etringer, and I. Baker (2009), Improving simulated soil temperatures and soil freeze/thaw at high-latitude regions in the Simple Biosphere/Carnegie-Ames-Stanford Approach model, *Journal of Geophysical Research*, *114*(F2), 1-18, doi:10.1029/2008JF001125.

Schmid, H. P., H.-B. Su, C. S. Vogel, and P. S. Curtis (2003), Ecosystem-atmosphere exchange of carbon dioxide over a mixed hardwood forest in northern lower Michigan, *Journal of Geophysical Research*, *108*(D14), 4417, doi:10.1029/2002JD003011.

Schmid, H., C. S. B. Grimmond, F. Cropley, B. Offerle, and H.-B. Su (2000), Measurements of CO2 and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agricultural and Forest Meteorology*, *103*(4), 357-374, doi:10.1016/S0168-1923(00)00140-4.

Schwalm, C. R., T. A. Black, K. Morgenstern, and E. R. Humphreys (2007), A method for deriving net primary productivity and component respiratory fluxes from tower-based eddy covariance data: a case study using a 17-year data record from a Douglas-fir chronosequence, *Global Change Biology*, *13*(2), 370-385, doi:10.1111/j.1365-2486.2006.01298.x

Schwalm, C. R. et al. (2010), A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis, *Journal of Geophysical Research*, *115*, 22, doi:10.1029/2009JG001229.

Siqueira, M. B., G. G. Katul, D. a Sampson, P. C. Stoy, J.-Y. Juang, H. R. Mccarthy, and R. Oren (2006), Multiscale model intercomparisons of CO2 and H2O exchange rates in a maturing southeastern US pine forest, *Global Change Biology*, *12*(7), 1189-1207, doi:10.1111/j.1365-2486.2006.01158.x.

Sitch, S. et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Global Change Biology*, *9*(2), 161-185, doi:10.1046/j.1365-2486.2003.00569.x.

Thornton, P. E., S. W. Running, and E. R. Hunt (2005), *Biome□BGC: Terrestrial Ecosystem Process Model, Version 4.1.1. model product, Oak Ridge Natl. Lab. Distrib. Act. Arch. Cent., Oak Ridge, Tenn. (Available at http://daac.ornl.gov)*.

Thornton, P. E., S. W. Running, and M. a White (1997), Generating surfaces of daily meteorological variables over large regions of complex terrain, *Journal of Hydrology*, *190*(3-4), 214-251, doi:10.1016/S0022-1694(96)03128-9.

Tian, H., G. Chen, M. Liu, C. Zhang, G. Sun, C. Lu, X. Xu, W. Ren, S. Pan, and A. Chappelka (2010), Model estimates of net primary productivity, evapotranspiration, and water use efficiency in the terrestrial ecosystems of the southern United States during 1895–2007, *Forest Ecology and Management*, *259*(7), 1311-1327, doi:10.1016/j.foreco.2009.10.009.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, and J. W. Munger (2007), Factors controlling CO 2 exchange on timescales from hourly to decadal at Harvard Forest, *Journal of Geophysical Research*, *112*(G2), 1-25, doi:10.1029/2006JG000293.

Vetter, M. et al. (2008), Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models, *Biogeosciences*, *5*(2), 561-583, doi:10.5194/bg-5-561-2008.

Wang, Y.-P. P., C. M. Trudinger, and I. G. Enting (2009), A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales, *Agricultural and Forest Meteorology*, 149(11), 1829-1842, doi:10.1016/j.agrformet.2009.07.009.

Weng, E., and Y. Luo (2008), Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: A modeling analysis, *Journal of Geophysical Research*, *113*(G3), 1-16, doi:10.1029/2007JG000539

Williamson, T. B., D. T. Price, J. L. Beverly, P. M. Bothwell, B. Frenkel, J. Park, and Patriquin, M.N. (2008), Assessing potential biophysical and socioeconomic impacts of climate change on forest□based communities: A methodological case study, *Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB. Inf. Rep. NOR□X□415E.*

Yang, X., V. Wittig, A. K. Jain, and W. Post (2009), Integration of nitrogen cycle dynamics into the Integrated Science Assessment Model for the study of terrestrial ecosystem responses to global change, *Global Biogeochemical Cycles*, *23*(4), 1-18, doi:10.1029/2009GB003474.

Yuan, W., Luo, Y., Richardson, A.D., et al. (2009), Latitudinal patterns of magnitude and interannual variability in net ecosystem exchange regulated by biological and environmental variables, Global Change Biology, 15(12), 2905-2920, doi:10.1111/j.1365-2486.2009.01870.x.

Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO 2 variability, *Global Biogeochemical Cycles*, *19*(1), GB1061, doi:10.1029/2004GB002273.

Zhan, X., Y. Xue, and G. J. Collatz (2003), An analytical approach for estimating CO2 and heat fluxes over the Amazonian region, *Ecological Modelling*, *162*(1-2), 97-117, doi:10.1016/S0304-3800(02)00405-2.

Zhao, M., F. A. Heinsch, R. R. Nemani, and S. W. Running (2005), Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sensing of Environment*, *95*(2), 164-176, doi:10.1016/j.rse.2004.12.011.

Zhao, M., S. W. Running, and R. R. Nemani (2006), Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses, *Journal of Geophysical Research*, *111*(G1), 1-13, doi:10.1029/2004JG000004.

Zhao, M., and S. W. Running (2010), Drought-induced reduction in global terrestrial net primary production from 2000 through 2009., *Science (New York, N.Y.)*, *329*(5994), 940-3, doi:10.1126/science.1192666.

Fig. 1. Mean model bias (Modeled-Measured, gC m^{-2} yr^{-1}) over all years when compared to annual gap-filled observations of net ecosystem exchange (NEE), gross primary productivity (GPP) and ecosystem respiration (RE). Results are grouped by two plant functional types (EVG: Evergreen needle-leaf forest; DBF: Deciduous broadleaf forest). The right panel is the normalized frequency distribution of model biases grouped by plant functional type, showing the distribution of values in the bar charts on the left, reported as a total bias for NEE, and a percentage of the annual total [(Modeled-Measured)/Measured] for GPP and RE. X-axis ranges are truncated to represent only observed range of biases.

Fig. 2. The distribution of the magnitude of interannual variability (IAV) in annual totals of net ecosystem exchange (NEE), gross primary production (GPP) and ecosystem respiration (RE), over all sites for each model (grey lines) and the data (black line). See Table 3 for individual model values.

Fig. 3. Statistical comparison (on a log-log scale) of model performance (normalized root mean square error vs χ^2 statistic) for interannual variability (IAV) in annual totals of net ecosystem exchange (NEE), gross primary productivity (GPP) and ecosystem respiration (RE) for the two plant functional types (DBF (green): Deciduous

broadleaved forests; EVG (blue): Evergreen needleleaf forests). See supplementary material for graphs with error bars (Fig. S1, S2).

Fig. 4. Residuals (predicted-observed) of monthly variability in net ecosystem exchange (NEE), gross primary productivity (GPP) and ecosystem respiration (RE). Positive values indicate a higher variability in the observations than in a model. The mean model-data residuals are presented as a dashed line. The grey area represents the standard deviation about the mean. Values represent averages over all sites for each plant functional type, and all sites taken together. Model codes – A: BEPS, B: Biome-BGC, C: CanIBIS, D: CNCLASS, E: DLEM, F: EDLUEEDCM, G: ECOSYS, H: ED2, I: ISAM, J: LoTEC-DA, K: LPJml, L: ORCHIDEE, M: SiB, N: SiBCASA, O: SSiB2, P: TECO, Q: BESS, R: MODISc5, S: MODISc5.1, X: Mean

Fig. 5. Monthly modelled and observed anomalies (mean model: green dashed line; standard deviation of models: green shaded area; observed: solid black line) in gross primary production (GPP), ecosystem respiration (RE) and net ecosystem exchange (NEE) for the year 2002 at three sites (Ca-Ojp, Ca-Obs, Ca-Oas) which exhibit a regionally coherent anomaly in that year. The standard deviation of normal observed monthly variability is presented as the grey area with dark grey columns. The duration of the anomalous temperature event is shown in the crosshatched grey area for each site.

Country-site	Name	Lat.	Long.	Elevation (m.a.s.l.)	Years	Biome	Mean NEE	δ NEE	NEE error	Mean GPP	δ GPP	GPP error	Mean RE	δ RE	RE error	Reference
CA-Ca1	BC, Campbell River □ Mature Forest Site	49.87	-125.33	300	1998-2005	EVG	-244	65	32	2310	112	44	2065	130	72	Schwalm et al., 2007
US-UMB	MI □ University of Michigan	45.56	-84.71	234	1999-2005	DBF	-132	65	22	1189	43	32	1057	87	42	Schmid et al., 2003
CA-Oas	Sask. □ SSA Old Aspen Biological Station	53.63	-106.20	530	1997-2005	DBF	-158	100	15	1090	150	24	932	86	35	Barr et al., 2004
CA-Obs	Sask. □ SSA Old Black Spruce	53.99	-105.12	629	2000-2005	EVG	-56	20	8	795	55	14	738	41	19	Griffis et al., 2003
CA-Ojp	Sask. □ SSA Old Jack Pine	53.92	-104.69	579	2000-2005	EVG	-30	26	9	612	48	16	582	30	22	Griffis et al., 2003
CA-TP4	Ontario □ Turkey Point Mature White Pine	42.71	-80.36	219	2001-2005	EVG	-133	72	16	1391	84	20	1258	75	30	Peichl and Arain, 2007
US-Ha1	MA □ Harvard Forest EMS Tower (HFR1)	42.54	-72.17	303	1992-2005	DBF	-217	117	36	1409	156	87	1192	125	78	Urbanski et al., 2007
US-Ho1	ME □ Howland Forest (Main Tower)	45.20	-68.74	60	1996-2004	EVG	-223	53	18	1518	90	22	1295	82	33	Richardson et al., 2009
US-MMS	IN □ Morgan Monroe State Forest	39.32	-86.41	275	1999-2005	DBF	-348	37	23	1331	56	26	983	68	31	Schmid et al., 2000
US-NR1	CO □ Niwot Ridge Forest (LTER NWT1)	40.03	-105.55	3050	1998-2005	EVG	-37	25	15	804	54	25	767	52	38	Bradford et al., 2008
US-Pfa	WI □ Park Falls/WLEF	45.95	-90.27	485	1995-2005	DBF	45	26	21	1005	49	35	1050	35	45	Davis et al., 2003

Table 2. Summary of model characteristics.

	Model									
Attribute	BEPS	BIOME-BGC	Can-IBIS	CNCLASS	DLEM	ECOSYS	ED2	EDCM	ISAM	LoTEC-DA
Temporal Resolution	Daily	Daily	Half-hourly	Half-houly	Daily	Hourly	Half-hourly	Monthly	Weekly	Half-hourly
Vegetation Pools	4	7	3	4	6	9	6	8	5	4
Soil Pools	9	4	7	3	3	9	4	5	8	5
Canopy Phenology	Semi-prognostic	Prognostic	Prognostic	Prognostic	Semi- Prognostic	Prognostic	Prognostic	Prognostic		Prognostic
Gross Primary Photosynthesis (GPP)	Enzyme Kinetic model	Stomatal Conductance Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Stomatal Conductance Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Light Use Efficiency	Enzyme Kinetic Model	Enzyme Kinetic Model
Heterotrophic Respiration (HR)	Temperature (Air + Soil) Precipitation Soil Evaporation Soil Carbon + Nitrogen	Soil Temperature Soil Moisture Soil Carbon	First or Greater Order Model	First or Greater Order Model	Air Temperature Soil Temperature Litter and Soil Carbon Soil Nitrogen Soil Moisture	Dissolved Carbon Loss Soil Temperature Soil Moisture Shortwave & Longwave Radiation Soil Carbon Vegetation Carbon	Soil Temperature Soil Moisture Soil Carbon Soil Nitrogen	Soil Temperature Soil Moisture Soil Carbon Dissolved Carbon Loss Vegetation Carbon Soil Nitrogen	First or Greater Order Model	Soil Temperature Soil Moisture Soil Carbon

						Soil Nitrogen Leaf Nitrogen				
Autotrophic Respiration (AR)	Air Temperature GPP	Air Temperature Vegetation Carbon Leaf Nitrogen	Air Temperature Soil Temperature Precipitation Soil Moisture Incident Shortwave + Longwave Radiation Vegetation Carbon	Fraction of Instantaneous GPP	Air Temperature Vegetation Carbon Leaf Nitrogen GPP	Air Temperature Soil Temperature Vegetation Carbon Leaf Nitrogen	Air Temperature Soil Temperature Vegetation Carbon Leaf Nitrogen GPP	Proportional to Growth	Proportional to Growth	Air Temperature Soil Temperature Soil Moisture Vegetation Carbon GPP
Ecosystem Respiration	AR+HR	Air Temperature Soil Temperature Soil Moisture Soil Carbon Vegetation Carbon	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR	AR + HR
Net Primary Productivity (NPP)	GPP-AR	Shortwave Radiation Vapor Pressure Deficit CO2 Vegetation Carbon	GPP □ AR	Fraction of Instantaneous GPP	GPP □ AR	GPP □ AR	GPP □ AR	Air Temperature Precipitation Soil Carbon Soil Nitrogen Soil Moisture	GPP □ AR	GPP □ AR

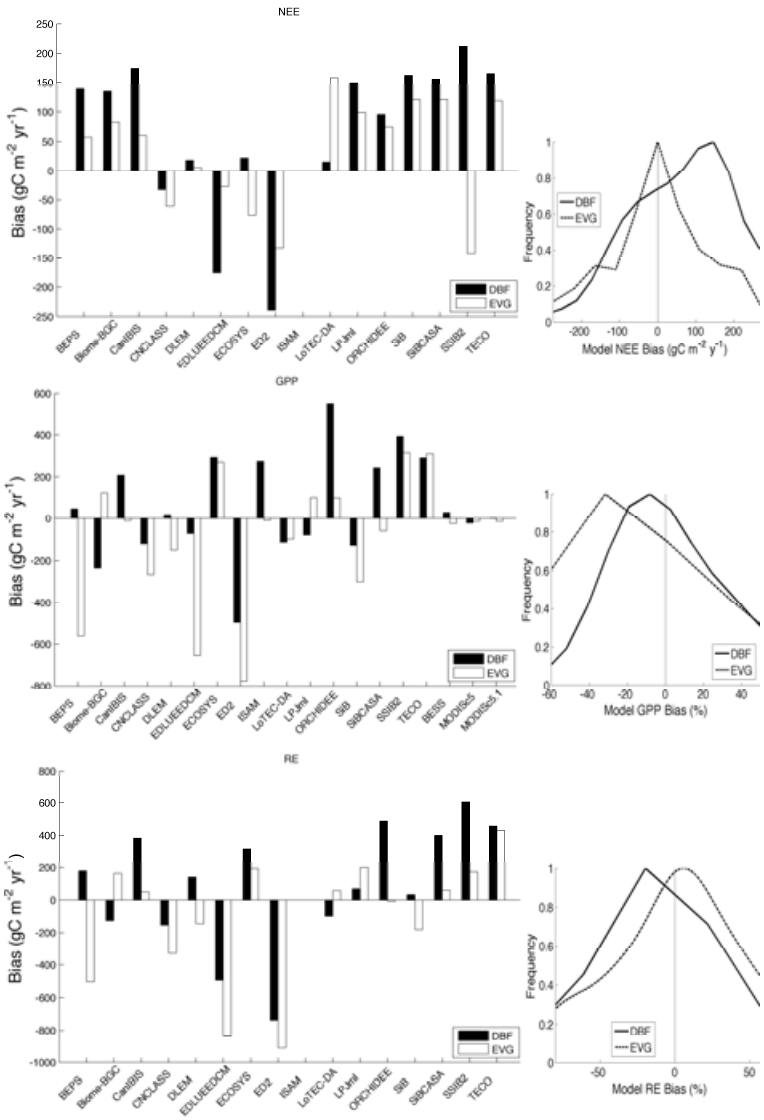
		Leaf Nitrogen						Vegetation Carbon Leaf Nitrogen		
Net Ecosystem Exchange	NPP-HR	Soil Temperature Soil Moisture Shortwave Radiation Vapor Pressure Deficit	NPP □ HR	GPP □ R	NPP □ HR	GPP □ R	NPP □ HR	NPP □ HR	NPP □ HR	
Reference	Liu et al., 1999	Thornton et al., 2005	Williamson et al., 2008	Arain et al., 2006	Tian et al., 2010	Grant et al., 2005	Medvigy et al., 2009	Liu et al., 2003	Yang et al., 2009	Hanson et al., 2004

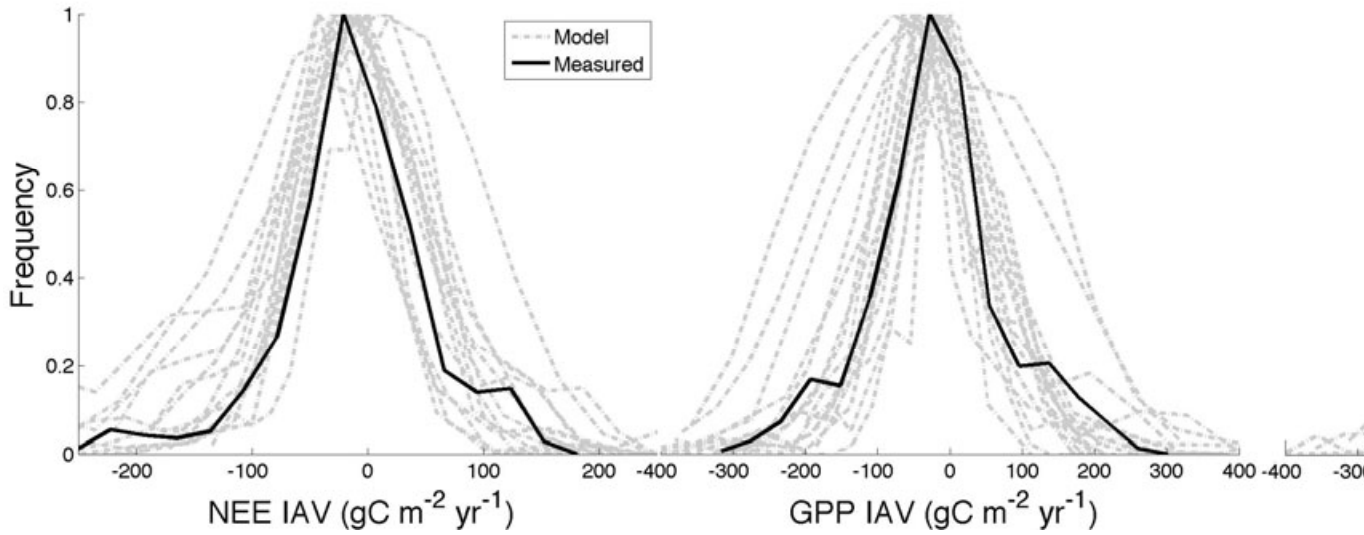
	Model						Remote Sensing		
Attribute	LPJml	ORCHIDEE	SiB	SiB-CASA	SSiB2	TECO	BESS	MODIScc5	MODIScc5.1
Temporal Resolution	Daily	Half-hourly	Half-hourly	10 min	Half-hourly	Hourly	Daily	Daily	Daily
Vegetation Pools	3	8	0	8	0	3	n/a	n/a	n/a
Soil Pools	2	8	0	5	0	5	n/a	n/a	n/a
Canopy Phenology	Prognostic	Prognostic	Prescribed	Prescribed	Prescribed	Prognostic	Prescribed	Prescribed	Prescribed
Gross Primary Photosynthesis (GPP)	Stomatal Conductance Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Enzyme Kinetic Model	Stomatal Conductance Model	Stomatal Conductance Model	Enzyme Kinetic Model	Light-use-efficiency model	Light-use-efficiency model
Heterotrophic Respiration (HR)	Soil Temperature Soil Moisture Soil Carbon	Soil Temperature Soil Moisture Soil Carbon	Zero□order Model	Soil Temperature Soil Moisture Soil Carbon	Zero□order Model	First or greater order model	n/a	n/a	n/a
Autotrophic Respiration (AR)	Air Temperature Soil Moisture Vegetation Carbon	Air Temperature Vegetation Carbon	Fraction of Instantaneous GPP	Air Temperature Soil Moisture Vegetation Carbon	Air Temperature Soil Moisture Surface Incident Shortwave Radiation	Air Temperature Vegetation Carbon	n/a	n/a	n/a

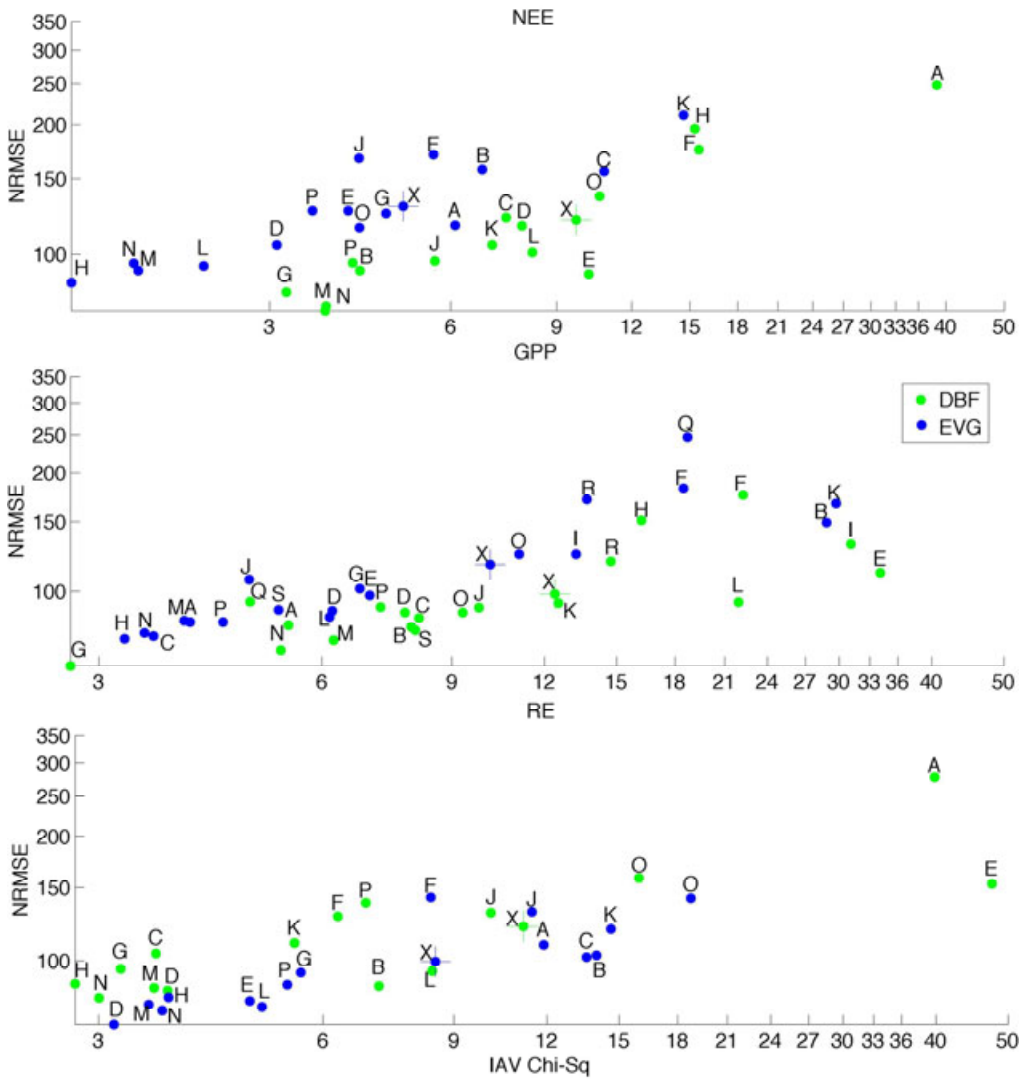
					Relative Humidity LAI fPAR CO2 Forced Annual Balance				
Ecosystem Respiration	AR + HR	AR + HR	Forced Annual Balance	AR + HR	GPP □ AR	AR + HR	n/a	n/a	n/a
Net Primary Productivity (NPP)	GPP □ AR	GPP □ AR	GPP □ AR	Air Temperature Soil Moisture CO2 Relative Humidity	GPP □ R	GPP □ AR	GPP	GPP	GPP
Net Ecosystem Exchange	NPP □ HR	GPP □ R	GPP □ R	GPP □ R	Zero□order Model	GPP □ R	n/a	n/a	n/a
Reference	Sitch et al., 2003	Krinner et al., 2005	Baker et al., 2008	Schaefer et al., 2009	Zhan et al., 2003	Weng and Luo, 2008	Ryu et al. 2012	Running et al. 2004	Zhao et al., 2005

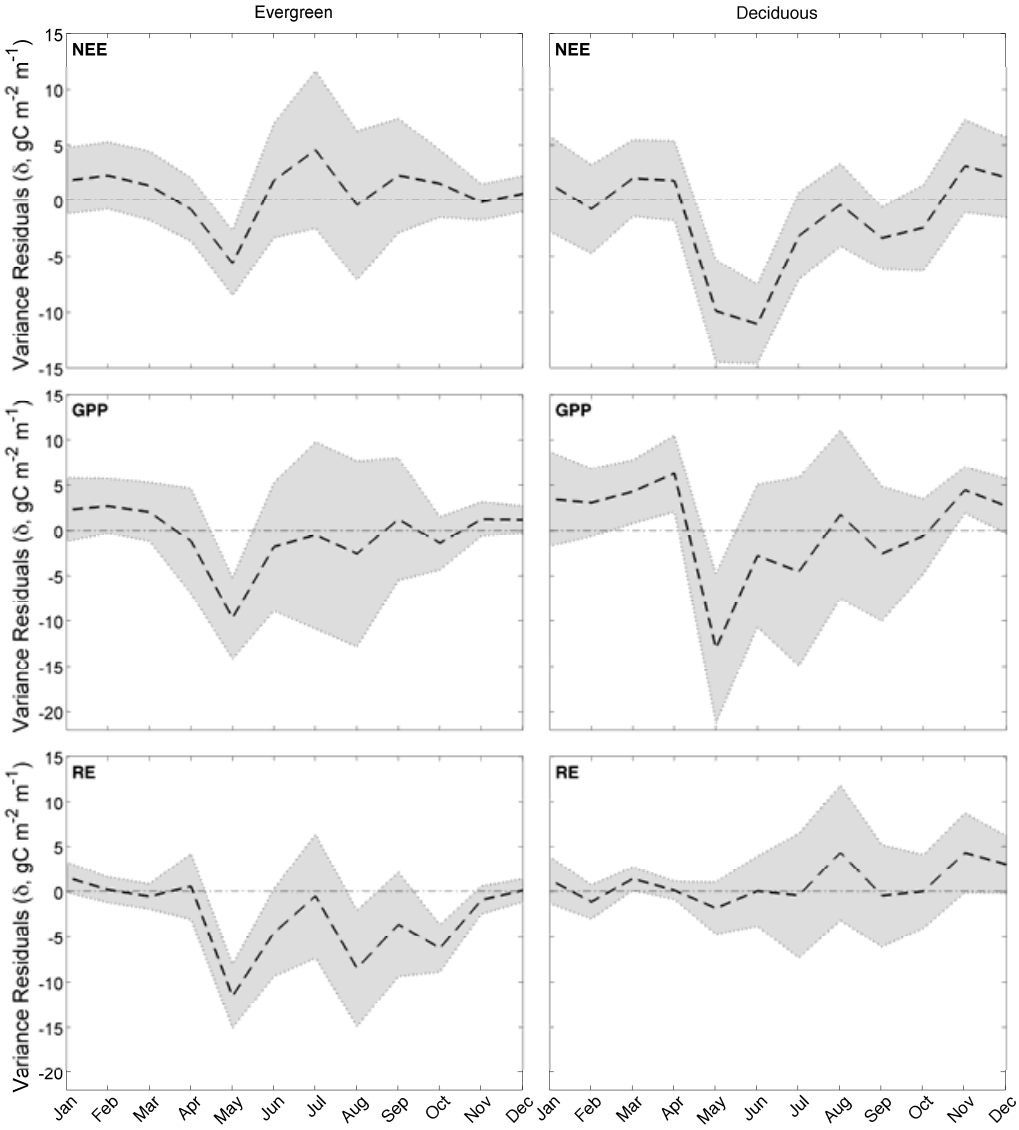
Table 3. Mean standard deviation of observed and modeled interannual variability for net ecosystem exchange (NEE), gross primary productivity (GPP) and ecosystem respiration (RE). Sites are grouped as deciduous broadleaved forest or evergreen needle leaved forest. Mean error of observed annual sums is also given for reference (in brackets). IAV: Interannual variability. Model codes, given in brackets, are those used in Fig. 3.

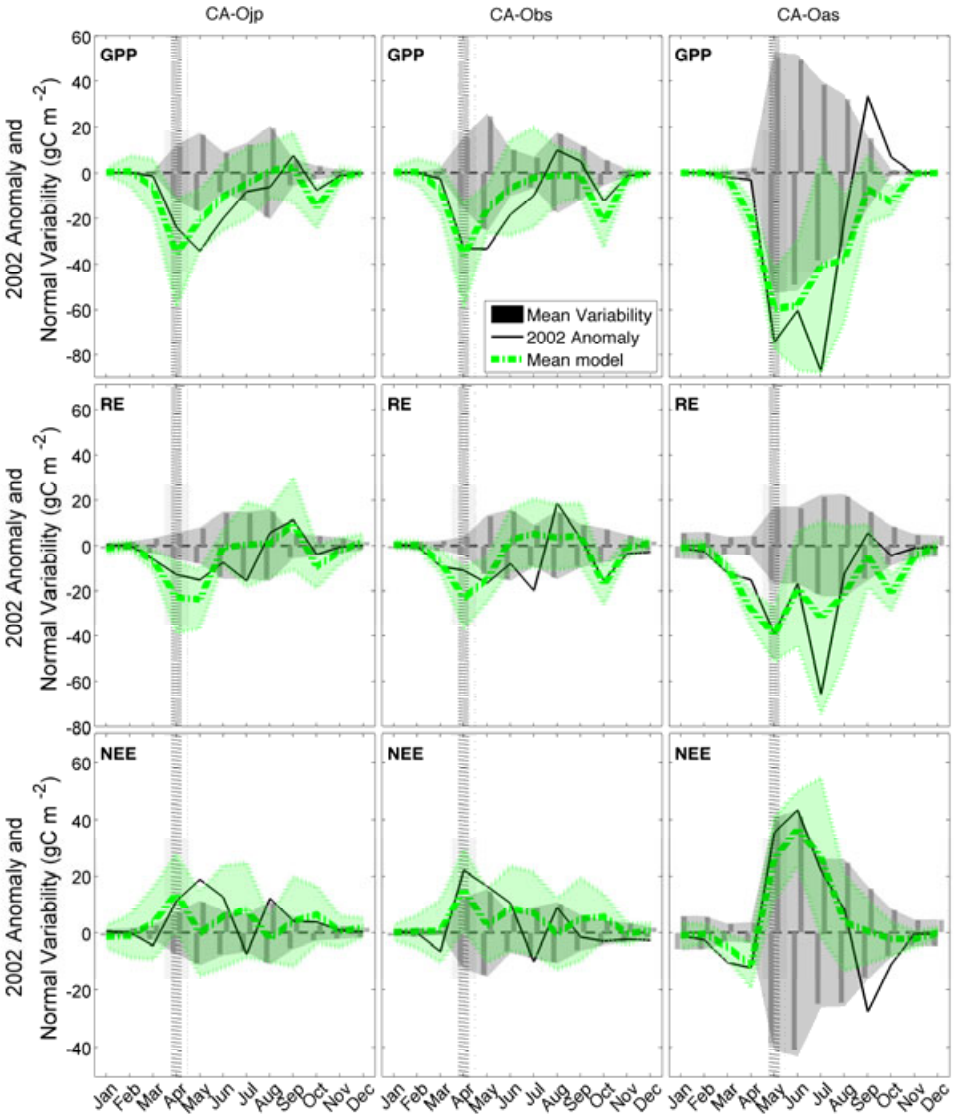
	NEE		GPP		RE	
	DBF	EVG	DBF	EVG	DBF	EVG
Observed IAV	85.53	44.49	115.05	74.30	91.56	74.32
Annual error	(24.73)	(17.12)	(47.73)	(24.54)	(50.89)	(37.50)
Models						
BEPS (A)	213.54	40.36	110.54	69.16	234.33	85.04
Biome-BGC (B)	59.66	77.99	115.54	129.05	103.42	78.51
CanIBIS (C)	57.70	71.36	91.42	78.94	66.11	87.81
CNCLASS (D)	76.98	31.34	55.87	55.84	68.36	39.41
DLEM (E)	81.57	39.28	202.73	81.54	224.81	59.75
EDLUEEDCM (F)	169.20	65.57	240.33	93.72	103.86	43.23
ECOSYS (G)	47.79	59.43	110.59	64.85	86.59	64.58
ED2 (H)	134.66	17.69	156.14	23.85	54.08	14.79
ISAM (I)			178.31	98.10		
LoTEC-DA (J)	88.26	92.03	121.13	56.05	57.77	103.90
LPJml (K)	72.97	104.96	127.76	145.35	86.87	93.58
ORCHIDEE (L)	93.58	26.33	175.23	63.33	113.73	47.96
SiB (M)	0.28	0.54	33.23	32.87	33.14	32.90
SiBCASA (N)	37.93	38.81	54.86	40.53	65.39	56.86
SSIB2 (O)	93.55	53.50	74.59	66.23	124.01	95.89
TECO (P)	64.32	38.87	94.41	57.15	121.13	54.89
BESS (Q)			104.32	178.26		
MODISc5 (R)			82.17	136.14		
MODISc5.1 (S)			48.92	52.06		


NEE
Bias (gC m-2 yr-1)
DBF
EVG
BEPS
Biome-BGC
CanIBIS
CNCLASS
DLEM
EDLUEEDCM
ECOSYS
ED2
ISAM
LoTEC-DA
LPJml
ORCHIDEE
SiB
SiBCASA
SSiB2
TECO
Frequency
Model NEE Bias (gC m-2 y-1)
GPP
Model GPP Bias (%)
RE
Model RE Bias (%)


Model
Measured
Frequency
1
0.8
0.6
0.4
0.2
0
-200
-100
0
100
200
-400
-300
-200
-100
0
100
200
300
400
-400
NEE IAV (gC m-2 yr-1)
GPP IAV (gC m-2 yr-1)


NEE
NRMSE
GPP
DBF
EVG
RE
IAV Chi-Sq


Evergreen
Deciduous
NEE
NEE
GPP
GPP
RE
RE
Variance Residuals (δ, gC m-2 m-1)
Variance Residuals (δ, gC m-2 m-1)
Variance Residuals (δ, gC m-2 m-1)
15
10
5
0
-5
-10
-15
-20
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


CA-Ojp
CA-Obs
CA-Oas
GPP
RE
NEE
2002 Anomaly and Normal Variability (gC m -2)
Mean Variability
2002 Anomaly
Mean model
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec