Global Change Biology (2010), doi: 10.1111/j.1365-2486.2010.02309.x

Diagnosing and assessing uncertainties of terrestrial ecosystem models in a multimodel ensemble experiment: 1. Primary production

WEILE WANG*†, JENNIFER DUNGAN†, HIROFUMI HASHIMOTO*†, ANDREW R. MICHAELIS*†, CRISTINA MILESI*†, KAZUHITO ICHII‡ and RAMAKRISHNA R. NEMANI†

*California State University, Monterey Bay, Seaside, CA, USA, †NASA Ames Research Center, Moffett Field, CA 94035, USA, ‡Faculty of Symbiotic Systems Science, Fukushima University, Japan

Abstract

We conducted an ensemble modeling exercise using the Terrestrial Observation and Prediction System (TOPS) to evaluate sources of uncertainty in carbon flux estimates resulting from structural differences among ecosystem models. The experiment ran public-domain versions of BIOME-BGC, LPJ, CASA, and TOPS-BGC over North America at 8 km resolution and for the period of 1982–2006. We developed the Hierarchical Framework for Diagnosing Ecosystem Models (HFDEM) to separate the simulated biogeochemistry into a cascade of three functional tiers and sequentially examine their characteristics in climate (temperature–precipitation) and other spaces. Analysis of the simulated annual gross primary production (GPP) in the climate domain indicates a general agreement among the models, all showing optimal GPP in regions where the relationship between annual average temperature (T, °C) and annual total precipitation (P, mm) is defined by $P = 50T + 500$. However, differences in simulated GPP are identified in magnitudes and distribution patterns. For forests, the GPP gradient along $P = 50T + 500$ ranges from $\sim 50\,\mathrm{g\,C\,yr^{-1}\,m^{-2}\,°C^{-1}}$ (CASA) to $\sim 125\,\mathrm{g\,C\,yr^{-1}\,m^{-2}\,°C^{-1}}$ (BIOME-BGC) in cold/temperate regions; for nonforests, the diversity among GPP distributions is even larger. Positive linear relationships are found between annual GPP and annual mean leaf area index (LAI) in all models. For BIOME-BGC and LPJ, such relationships lead to a positive feedback from LAI growth to GPP enhancement. Different approaches to constrain this feedback lead to different sensitivity of the models to disturbances such as fire, which contribute significantly to the diversity in GPP stated above. The ratios between independently simulated NPP and GPP are close to 50% on average; however, their distribution patterns vary significantly between models, reflecting the difficulties in estimating autotrophic respiration across various climate regimes. Although these results are drawn from our experiments with the tested model versions, the developed methodology has potential for other model exercises.

Keywords: biogeochemistry, carbon cycle, model intercomparison, model uncertainty, multimodel ensemble, primary production, terrestrial ecosystem model

Received 7 May 2010 and accepted 6 July 2010

UNCORRECTED PROOF

Introduction

The rising concentration of CO_2 (and other greenhouse gases) in the atmosphere, induced mainly by anthropogenic carbon emissions, represents a major threat to the Earth's climate system. Yet the experienced growth of atmospheric CO_2 does not reflect the full impacts of human's disturbance on the global carbon cycle, as half of the emitted CO_2 is sequestered by natural reservoirs in the ocean and on the land (IPCC, 2007). Between 2000 and 2006, for instance, global anthropogenic carbon emissions was about $9.1\,\mathrm{Pg\,C\,yr^{-1}}$ (petagrams of carbon per year), while carbon uptake by the ocean and the land was 2.2 and $2.8\,\mathrm{Pg\,C\,yr^{-1}}$, respectively (Canadell *et al.*, 2007). In North America, terrestrial ecosystems alone sequestered $0.65\,\mathrm{Pg\,C\,yr^{-1}}$ during the same period, offsetting one-third of carbon emissions from fossil fuel burning and cement manufacturing (Peters *et al.*, 2007).

Compared with their oceanic counterparts, terrestrial carbon sinks have more uncertainties associated with them (Houghton, 2007). Estimates from atmospheric inversion studies (i.e., the 'Top-Down' approach) generally suggest that the major part of the carbon sink (about $2.0\,\mathrm{Pg\,C\,yr^{-1}}$) is located in the northern hemisphere (Tans *et al.*, 1990; Rayner *et al.*, 1999; Gurney *et al.*, 2002; Jacobson *et al.*, 2007). However, estimates from inventory data and satellite observations (i.e., the 'Bottom-Up' approach) indicate the uptake by northern

Correspondence: W. Wang, (c/o Ramakrishna R. Nemani), tel. +1 650 604 6444, fax +1 650 604 6569, e-mail: weile.wang@gmail.com

© 2010 Blackwell Publishing Ltd

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

GCB	2309	B	Dispatch: 23.8.10 Journal: GCB	CE: Guna/Mahendra
Journal Name	Manuscript No.		Author Received: No. of pages: 17	PE: Dheiran/Mini

GCB 2309

forests may only account for 0.7 Pg C yr^{-1} (Myneni *et al.*, 2001; Goodale *et al.*, 2002; Potter *et al.*, 2003). One explanation for this discrepancy is that much of the sink may be contributed by ecosystems other than forests (Pacala *et al.*, 2001), while other studies suggest that carbon uptake by tropical forests may be underestimated (Nemani *et al.*, 2003; Stephens *et al.*, 2007). In addition to the latitudinal discrepancy, uncertainties of carbon budgets at regional scales are even larger (IPCC, 2007).

Critical scientific tools for investigating the terrestrial carbon cycle are ecosystem models. They integrate the understanding of ecological processes obtained from local measurements and apply the knowledge to simulate ecosystem functions over broader regions. Simulation results from ecosystem models also provide important prior information for atmospheric inversion experiments. Therefore, uncertainties in ecosystem models have direct (in the 'Bottom-Up' approach) or indirect (in the 'Top-Down' approach) implications for the estimation of carbon fluxes and carbon stocks.

Endogenous uncertainties of ecosystem models come from initial conditions, model parameters, and model structures. The first two sources of uncertainty are well recognized and routinely addressed in model experiments (e.g., through model spin-up and model calibration). The last source of uncertainty, structural uncertainty, arises from different representations of ecological processes in different models. Because the components of terrestrial ecosystems and the interactions among them are complicated or not well understood, simplifying assumptions must be made to describe them in numerical models. Different modeling strategies may adopt different simplifying assumptions, leading to different model complexity and behavior.

In some simple cases (e.g., a particular aspect of the ecosystem), structural differences of ecosystem models may be examined by directly comparing their mathematical formulation. For instance, Adams *et al.* (2004) analyzed algorithms of net primary production (NPP) in 10 different models and characterized their functional responses to major climate driver and environmental variables (e.g., temperature, soil moisture, and etc.). However, characteristics of isolated model components do not fully reflect their functional behavior within the coupled system, where feedbacks and interactions between the subsystems play a critical role. Mathematical analysis of complicated systems is usually difficult, so numerical experiments with multimodel ensemble (MMEs) have become the main means to tackle the problem. In general, an MME experiment runs a group models with the same input data and under the same initial/boundary conditions; the multimodel means are often treated as the 'best' simulation results, and the intermodel differences are used as a measure of the structural uncertainty (Tebaldi & Knutti, 2007).

Following the successful example of the World Climate Research Programme's Coupled Model Intercomparison Project for the International Panel on Climate Change (IPCC), MME experiments are now broadly adopted in carbon-cycle studies, such as the Atmospheric Tracer Transport Model Intercomparison Project (TransCom; Rayner & Law, 1995; Denning *et al.*, 1999; Gurney *et al.*, 2004). For ecosystem models, in particular, Q1 previous major MME experiments include the Vegetation/Ecosystem Modeling and Analysis Project (VEMAP; VEMAP Members, 1995), the Potsdam NPP Model Intercomparison Project (Cramer *et al.*, 1999), and the Potsdam DGVM Intercomparison Project (Cramer *et al.*, 2001). These experiments utilized ensembles of 6–17 ecosystem models to investigate contemporary patterns of vegetation production and other important carbon fluxes (e.g., NEE) in the United States and at global scales, as well as their responses to future climate change scenarios. Despite the general agreement in the simulation results and the implied scientific significance (e.g., see Schimel, 2007), large differences among the Q3 models were also revealed, for instance, in estimates of contemporary global annual NPP (39.9–80.5 Pg C yr^{-1}; Cramer *et al.*, 1999), or in the sensitivity of carbon storage to future climate change in the United States (−39 to +40%; VEMAP Members, 1995).

The large uncertainties found among ecosystem models highlight the necessity of systematically diagnosing ensemble members in an MME experiment. On one hand, models rarely perform equally well and it is legitimate to weigh more heavily models that produce more realistic estimates. On the other hand, it is also important to gain understanding of the causes of the intermodel discrepancies so that the member models may be improved and the overall uncertainties may be reduced in future model exercises.

Conventionally, ecosystem models are diagnosed by comparisons between simulation outputs and certain reference datasets. For instance, Randerson *et al.* (2009) represented a framework to evaluate simulated terrestrial biogeochemistry in the Carbon-Land Model Intercomparison Project (C-LAMP) using satellite datasets [leaf area index (LAI) and NPP], flux tower measurements (carbon/energy fluxes), and some other data sources. However, this approach is not problem-free: there are measurement uncertainties and scaling issues with the flux tower data (Baldocchi, 2003), while satellite products of LAI and NPP are strictly not observations but model results. Also, since numerical parameter optimization is now a common practice in ecosystem model applications (Raupach *et al.*, 2005), Q4 there are potential risks that the models may be 'tuned'

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

with some of the reference datasets and thus the assumption about the independence between the reference data and the testing samples is violated. Furthermore, simply comparing model outputs with reference data does not necessarily indicate the reasons why they agree (or do not agree) with each other.

In an effort to support the North America Carbon Program (NACP) and by using the Terrestrial Observation and Prediction System (TOPS; Nemani *et al*., 2009), we conducted an MME experiment that includes public versions of the ecosystem models BIOME-BGC, LPJ, CASA, and TOPS-BGC. We developed an alternative diagnostic approach to systematically detect and evaluate structural uncertainties among the tested models. This approach, called as the Hierarchical Framework for Diagnosing Ecosystem Models (HFDEM), tries to separate the simulated biogeochemistry into a cascade of three functional tiers and examine them sequentially. The first tier analyzes the primary production (GPP/NPP) by checking its relationship with major climate drivers as well as vegetation leaf area. The second tier analyzes how carbon is allocated into different vegetation components and how biomass is accumulated through the balance between primary production and biomass turnover. The third tier analyzes the soil carbon balance between the influx from litter fall and the effluxes generated by heterotrophic respirations (HR). Because the framework simplifies the internal logic of models, not only does it help evaluate structural differences between ecosystem models, but also how these differences originate.

We applied the diagnostic framework to analyze structural uncertainties among the ecosystem models tested in our MME experiment and report the results here in two parts: the first part (i.e., this paper) introduces the setup of the model experiment, describes the development of methodology, and discusses the analysis results centered on primary production. The second part (i.e., the companion paper; Wang *et al*., 2010) will focus on detecting and evaluating uncertainties in the simulated carbon budgets.

Model experiment setup

We set up our model experiment with the TOPS. TOPS is a data and modeling system developed for ecological monitoring, forecasting, and related ecosystem analyses (Nemani *et al*., 2009). It integrates ecosystem models with meteorological records, climate model outputs, satellite products, and various ancillary datasets from different sources. TOPS also incorporates software tools to process and manage these spatial–temporal datasets. One of the key components of TOPS is the Surface Observation and Gridding System (SOGS), which ingests daily observations of temperature, precipitation and other fields from meteorological stations, and interpolates them to spatially continuous grids following the algorithms developed in Thornton *et al*. (1997) and Jolly *et al*. (2005). SOGS datasets are used to drive ecosystem models integrated within TOPS. What follows is a description of the datasets and models we used with TOPS to construct our MME.

Datasets

The spatial domain of this model experiment covers the entire North American continent at 8 km resolution and the time period simulated is from 1982 to 2006. We used SOGS to generate daily meteorological fields including minimum/maximum temperatures, precipitation, vapor pressure deficit, and incident solar radiation. The input meteorological data were obtained from the Global Summary of the Day (GSOD) and the Cooperative Summary of the Day (TD3200), both archived at the National Climatic Data Center (NCDC). GSOD is a global set based on data exchanged under the World Meteorological Organization (WMO) World Weather Watch Program. It has about 2000 reporting stations over North America, which is relatively sparse for interpolating climate variables such as daily precipitation over the whole continent. For this reason, we combined GSOD with the TD3200 dataset. TD3200 consists of about 8500 reporting stations in the United States, primarily from the National Weather Service (NWS) cooperative station network and principal climatological stations. Based on some informal comparisons of interpolation procedures, we chose the natural-neighbor procedure (Sibson, 1981) to interpolate the station data to the defined model domain because it had the fewest obvious artifacts. As an example, Fig. 1 shows the climatology of the generated annual mean temperature and total precipitation.

For models requiring satellite measured vegetation data, we used the LAI dataset developed by Ganguly *et al*. (2008), which is generated from the multidecade records of the Advanced Very High Resolution Radiometer (AVHRR) sensors but with comparable quality to the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI Collection 5 products. The MODIS land cover product (Friedl *et al*., 2002; Fig. 1c) is used to describe the distribution of plant functional types (PFTs). A global dataset of land surface parameters, ECOCLIMAP (Masson *et al*., 2003), is used to specify soil properties (e.g., texture and depth) and other parameters (e.g., albedo).

Ecosystem models

The architecture of TOPS provides a flexible interface for ecosystem models to be integrated. Here, we present

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

Colour

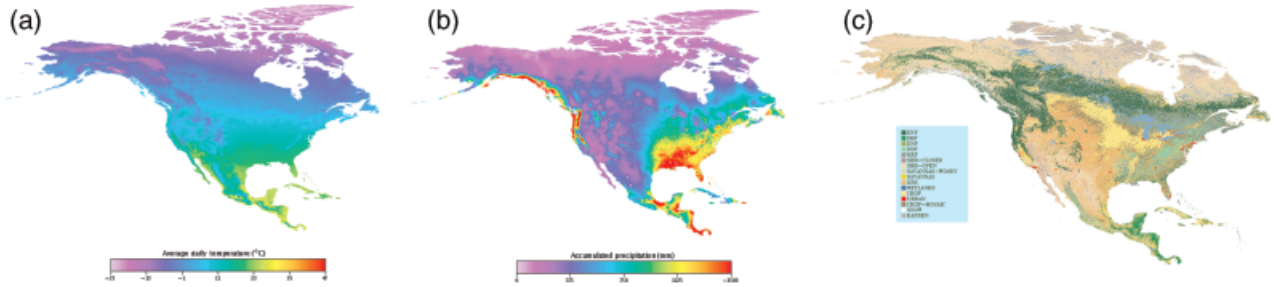

Fig. 1. The domain of the model experiment: (a) annual mean surface temperature; (b) annual total precipitation; and (c) land cover classification (MODIS). Forests on (c) are shown by green colors, whereas nonforests are shown in yellow/brown colors.

Table 1. Selected characteristics and attributes of models used in the MME

	Biome-BGC	LPJ	TOPS-BGC	CASA
Land cover	Prescribed	Simulated	Prescribed	Prescribed
LAI	Simulated	Simulated	Prescribed	Prescribed
Carbon pools	Vegetation and soil	Vegetation and soil	None	Vegetation and soil
GPP algorithm	Farquhar Model	Farquhar Model	LUE	2 × NPP
NPP algorithm	GPP–AR	GPP–AR	0.5 × GPP	LUE
NEE/HR algorithm	NPP–HR; HR estimated from soil pools	NPP–HR; HR estimated from soil pools	NPP–HR; HR estimated from base respiration rates	NPP–HR; HR estimated from soil pools
Dynamic nitrogen cycle	Yes	No	No	No
References	Thornton *et al.* (2002)	Sitch *et al.* (2003)	Nemani *et al.* (2009)	Potter *et al.* (1993)

GPP, gross primary production; LUE, light use efficiency; NPP, net primary production; AR, autotrophic respiration; HR, heterotrophic respiration.

UNCORRECTED PROOF

results from public versions of four process-based ecosystem models, including BIOME-BGC (Thornton *et al.*, 2002), LPJ (Sitch *et al.*, 2003), CASA (Potter *et al.*, 1993), and TOPS-BGC (White & Nemani, 2004). These models cover a range of different approaches to simulate ecological processes and have different complexity. A brief introduction to these models is given below, focusing on carbon cycling aspects. Some of their characteristics, and the associated major publications, are also summarized in Table 1. These models were integrated in TOPS for various other applications; as such, they are not intentionally selected but happen to be the available candidate models for this experiment.

BIOME-BGC (version 4.1.2) is an ecosystem model designed to estimate fluxes of carbon, water, and nitrogen in response to changes in climate and other environmental forcing (Thornton *et al.*, 2002). It incorporates all the major components of the carbon cycle. The radiation regime of canopy is computed by a two-big-leaf scheme (which treats sunlit and shaded leaf area separately), which is followed by the simulation of photosynthesis or gross primary production (GPP) based on the Farquhar model (Farquhar *et al.*, 1980). Maintenance respiration is calculated based on tissue mass, nitrogen content, and temperature, while growth respiration is assumed as a constant proportion of carbon allocated to growth. Allocation of carbon is made by prescribed allometric relationships, and litter is produced as tissues turn over normally or when plants die. Mortality, due to age or induced by fire, occurs at prescribed rates and is uniformly applied to the simulated plant (i.e., no age structure of the plant is recorded). Dead tissues degrade and move through a series of cascading litter/soil pools, during which HR is estimated (Thornton *et al.*, 2002).

One important feature of BIOME-BGC is that the simulated carbon cycle is tightly coupled with, or controlled by, the nitrogen cycle of the ecosystem. Carbon can only be allocated to growth when it is allowed by available nitrogen, and otherwise the surplus carbon is simply subtracted from the GPP. Therefore, GPP in BIOME-BGC is retrospectively adjusted according to nitrogen availability (Thornton *et al.*, 2002; Wang *et al.*, 2009). Finally, there are six main PFTs defined in BIOME-BGC, including various types of forests (e.g., needleleaf or broadleaf, evergreen or deciduous), shrubs, and grassland/cropland. A land cover map is required to assign one of those PFTs to each vegetated grid cell in the domain.

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

LPJ (version 3.1.1) is a dynamic global vegetation model (Sitch *et al*., 2003). The logic of LPJ is similar to that of BIOME-BGC in many aspects of the modeled carbon cycle, although the adopted technique details may be different. For instance, the radiation regime in LPJ is simulated in a one-big-leaf approach and photosynthesis is simulated based on a simplified, semiempirical version of the Farquhar model (Collatz *et al*., 1991; Haxeltine & Prentice, 1996). However, there are also substantial differences in some basic strategies between the two models. Unlike BIOME-BGC, LPJ does not emphasize the regulation of nitrogen availability on carbon allocation, but allows all available carbon to be used for plant growth. Yet it constrains the accumulation of biomass by implementing several scaling rules to regulate the size of individual trees and the density of woody species within a grid cell. When the tree size or tree density exceeds a certain threshold, a self-thinning process is triggered that forces the tree to turn over tissues or to die. Another key feature of LPJ is that it intrinsically supports subgrid cell vegetation structures. For this study, each grid cell in LPJ is treated as a stand that can have as many as 10 PFTs. These PFTs compete for the same natural resources (light, water, etc.) to obtain a population advantage by gaining more NPP. Therefore, LPJ does not need externally specified PFT distributions in the domain.

CASA [version 4.2.2; CASA has several different versions since its original development in Potter *et al*. (1993). The version used in this study is obtained from the Carnegie Institution of Washington, which has a structure close to the original model] is a classic ecosystem model that utilizes satellite measurements of vegetation (Potter *et al*., 1993). Unlike BIOME-BGC or LPJ, CASA does not simulate structural variables of the vegetation canopy (e.g., LAI or FPAR), but requires external inputs (satellite-measured vegetation indices) to estimate them and then to estimate the absorbed photosynthetic active radiation (APAR). It directly translates APAR into biomass (NPP) based on the notion of light use efficiency (LUE), which is a product of an optimal efficiency and regulation functions of environmental factors (e.g., temperature and water stress). The rest of the carbon cycle in CASA is similar to BIOME-BGC or LPJ, as it also simulates the carbon flow of NPP into various biomass compartments, and then into litter/soil pools.

A few revisions were made to the CASA model in this study. The original model requires NDVI as inputs to estimate FPAR and LAI internally. However, when we tested the model with AVHHR, GIMMS, NDVI (version 'G'; Tucker *et al*., 2005) datasets, we found that the internally estimated LAI was often saturated with the maximum values allowed by the model. Accordingly, we changed the model to directly take LAI as input, and estimated FPAR using Beer's law with an average extinction coefficient of 0.5 (Woodward, 1987), which is essentially the same scheme used in LPJ. These revisions allow us to examine the model's behavior in a broader dynamical range.

TOPS-BGC [in previous literature, TOPS-BGC is sometimes called as 'TOPS' for simplicity (e.g., White & Nemani, 2004). Here, 'TOPS' refers to the whole modeling system instead of a particular model running under it] (version 1.0.2) is derived from BIOME-BGC by removing model components dealing with dynamic carbon allocation and the nitrogen cycle (White & Nemani, 2004). It requires satellite measurement of LAI (and FPAR, if available) as inputs, and estimates photosynthesis in a LUE approach similar to the MODIS GPP algorithm (Zhao *et al*., 2005; Heinsch *et al*., 2006). Efforts have been made to improve the hydrological modules (Ichii *et al*., 2008) and to refine the optimal LUE for different land cover types (Yang *et al*., 2007). TOPS-BGC currently does not simulate biomass or other carbon pools. Therefore, it does not estimate autotrophic respiration (AR) independently, but assumes NPP is half of GPP (Waring *et al*., 1998). When a reference rate of HR is provided, TOPS-BGC also has the capacity to estimate the variations of HR under the regulations of soil temperature and soil moisture. This approach of estimating HR is similar to the scheme adopted in the Vegetation Photosynthesis and Respiration Model (VPRM; Mahadevan *et al*., 2008).

Model experiment protocol

Before normal model simulations, spinup runs are required to bring the states of the above models into equilibrium with the provided climate and ancillary datasets. This is generally accomplished by running the models for many simulation years (in which the climate data are recycled), yet different models may have different spinup algorithms (e.g., Thornton & Rosenbloom, 2005). BIOME-BGC implements a specific procedure to check whether the simulated ecosystem reaches its steady state at the end of the spinup run, which takes about 2000 simulation years on average. By default, LPJ runs for 1000 simulation years for the spinup process; it also has a procedure to analytically calculate the balance of soil carbon pools during the process. However, we noticed that 1000 simulation years were not long enough to bring the model into steady state; therefore, the model was run for an additional 1000 simulation years. CASA does not have a particular spinup procedure, but the model becomes stable after 100 climate cycles (i.e., 2500 simulation years). Finally, because TOPS-BGC has no carbon pools simulated, it only needs to bring soil moisture to contemporary status, which usually takes one climate cycle

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

(i.e., 25 simulation years). This stage of the model experiment does not consider climate change scenarios; a constant level of atmospheric CO_2 at 360 ppm is assumed in all the spinup runs. No disturbances are externally specified in the simulations. Fire is internally simulated by BIOME-BGC and LPJ.

With the endpoints of the spinup runs as initial conditions, we ran the models for the climate years from 1982 to 2006. Standard model outputs include monthly carbon fluxes (e.g., GPP, NPP, NEP, and respiration), monthly LAI (in the case of BIOME-BGC and LPJ), and annual averaged carbon stocks (biomass and soil carbon pools). For LPJ, results of a grid cell represent the proportionally weighted sum of all PFTs within it. We calculate the mean seasonal cycles of monthly variables and annual averages of annual variables over the simulation period.

Method of diagnosis

The basic idea of the HFDEM was developed from our analysis of the BIOME-BGC model (Wang *et al*., 2009), and was motivated by the fact that today's ecosystem models are so sophisticated that their calibration and analysis have become a very difficult task. To address this problem, a solution is to construct a model hierarchy by sequentially removing functional components from the original model (Held, 2005). Following this approach, Wang *et al*. (2009) identified three key functional tiers in BIOME-BGC, which include: (1) photosynthesis and evapotranspiration at the leaf level; (2) carbon (and nitrogen) allocation and respiration at the whole-plant level; and (3) carbon/nitrogen cycles in litter/soil pools. A model hierarchy was then developed and applied to successfully calibrate BIOME-BGC at multiple Fluxnet sites (Wang *et al*., 2009).

HFDEM applies the idea of Wang *et al*. (2009) to examine biogeochemistry simulated by ecosystem models in a more general fashion. With a focus on the carbon cycling, HFDEM first examines how carbon is assimilated into the system (GPP) and how GPP is partitioned into AR and net carbon gain (NPP). Next, it evaluates how NPP is allocated into different vegetation tissues and how biomass is accumulated. Finally, HFDEM checks how carbon is stored in soil pools and the carbon balance of the whole system.

For the purpose of diagnosis, it is not necessary to derive analytical expressions for the developed model hierarchies [as did in Wang *et al*. (2009)]. Instead, HFDEM utilizes diagrams that relate model outputs with their major drivers to help simplify the model logic. Because most ecological processes are driven by climate, climate-domain diagrams of carbon fluxes can be used to good effect in this context. Climate-domain diagrams organize model results by their climate context instead of their geographic locations, which represent intrinsic characteristics of the models in a more informative way than conventional geographic maps. Owing to their dominant importance to terrestrial ecosystems, temperature and precipitation are usually chosen as the two dimensions to define the climate domain (e.g., Churkina & Running, 1998; Schloss *et al*., 1999).

The climate domain adopted in this study spans across annual mean temperature between −10 and 30 °C (as the x-axis) and across annual total precipitation from 0 to 2000 mm yr^{-1} (as the y-axis; referenced as the T–P domain hereafter). This domain covers climate conditions of most vegetated areas over North America. To generate 2D diagrams, we divide the T–P domain into 100 × 100 bins, each having a 'climate resolution' of 0.4 °C and 20 mm yr^{-1}. Geographic grid cells are projected into these climate bins according to their annual temperature and precipitation (as shown in Fig. 1). Because the geographically defined (latitude–longitude) grid is not an equal-area projection, results from grid cells are weighted by the cosine of their latitude before they are summed and averaged in the T–P domain. It should be noted that the distributions of the grid cells in the T–P domain are generally not homogeneous, and thus results representing the climate bins may have different sample sizes. However, because the total sample base is large (~480 000 vegetated grid cells), major patterns on the T–P diagrams are generally supported by sufficient samples and thus statistically meaningful. Estimates of total GPP and NPP over the NACP domain are summarized in Table 2.

As an example, Fig. 2 shows the density distributions of vegetated grid cells for forests and nonforests, as classified by the MODIS land cover map (Fig. 1c), respectively, in the T–P domain. The distributions are rather heterogeneous and the density of the grid cells ranges from zero to over 200 in different climate regions, rendered by different colors (from blue to red). To facilitate the discussion, a boundary [The 'dry–moist' boundary adopted here is based on the Köppen climate classification system for the case where precipitation is evenly distributed throughout the year (Strahler & Strahler, 1978). For cases of precipitation concentrated in summer or winter, the offset of the boundary should change from 140 to 280 mm or 0 mm, respectively] between dry and moist climate conditions is shown in the diagrams, which is defined by the line of $P = 20T + 140$ (where T is in °C and P is in millimeters). In comparison, also shown is the upward ('lower/left-to-upper/right') diagonal of the T–P diagrams, defined by $P = 50T + 500$ (Fig. 2). In the discussions, we roughly divide the climate domain along the temperature dimension in three main regions, referred to as 'cold'

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

Table 2. Mean annual GPP* and NPP* estimated for North America

	GPP			NPP		
	Forest regions	Nonforest regions	Total	Forest regions	Nonforest regions	Total
Biome-BGC	6.79	3.88	10.7	2.36	1.64	4.00
LPJ	4.95	7.30	12.3	2.73	4.07	6.80
TOPS	4.99	5.28	10.3	2.49	2.64	5.13
CASA	3.79	5.04	8.83	1.90	2.52	4.42

*Units: $10^{15}\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}$.
†Based on total vegetated area around $1.77\times10^{13}\,\mathrm{m^2}$, including $0.55\times10^{13}\,\mathrm{m^2}$ for forest regions and $1.22\times10^{13}\,\mathrm{m^2}$ for nonforest regions.
GPP, gross primary production; NPP, net primary production.

($T<5\,^\circ$C), 'warm' ($5\,^\circ\mathrm{C}<T<20\,^\circ\mathrm{C}$), and 'hot' ($T>20\,^\circ$C), respectively.

As shown in Fig. 2a, forests are mostly distributed in moist climate regions, with the distribution centers largely aligned with $P = 50T + 500$. The highest density of forests is found in the cold climate zone (around 0 °C and $400\,\mathrm{mm\,yr^{-1}}$), contributed mainly by boreal forests. The centers of distribution spread rather evenly over regions where annual precipitation is between 1000 and $1500\,\mathrm{mm\,yr^{-1}}$ in the warm climate zone, corresponding to eastern temperate forests. The hot climate zone is mainly occupied by the tropical forests of Central America. In addition to these major groups, forests located in the cool ($\sim5\,^\circ$C) and very moist ($>1500\,\mathrm{mm\,yr^{-1}}$) climate regions are dominated by the marine west coast forests.

In contrast to forests, the distribution of nonforests is shifted towards arid climate regimes, and is mostly below the upward diagonal in warm and hot climate zones (Fig. 2b). Three major centers of distribution can be identified on the diagram: the first one is located in the very lower-left corner ($<-5\,^\circ$C), which is contributed by subarctic shrubs; secondly, a broad band is distributed close to the moist/arid boundary in the warm climate zone, which reflects the grasslands on the Great Plains and open shrubs over the North American deserts; the third center, less intensive but still recognizable, is located in warm and moist climate zones (centered at 10 °C and $1000\,\mathrm{mm\,yr^{-1}}$) and is mainly contributed by croplands (Fig. 2b).

Because forests and nonforests (e.g., shrubs and grasses) have different characteristics, we may discuss the two sets of results separately when it is necessary. In these cases, we used the MODIS land cover map (Fig. 1c) to determine the vegetation type of grid cells (as in Fig. 2). However, because vegetation distributions are internally simulated and solely regulated by climate in LPJ (see section on 'Model experiment protocol'), they essentially represent a scenario of 'potential vegetation', which may be different from the classifications of the MODIS land cover map. Such differences are generally small for forests but more evident for nonforests. For instance, because LPJ does not simulate human-induced land use and land cover changes, a large portion of the agricultural region in the mid-west of the continent (Fig. 1) is occupied by forests (Sitch *et al.*, 2003). Therefore, the above classification of nonforests essentially does not apply to the results of LPJ. When necessary, we use the term of 'vegetation simulated in nonforest regions' in the discussions to clarify the difference for LPJ.

In addition to the climate-domain diagrams, another tool adopted in this study is density-scatter plots, which are scatter plots with a color scheme indicating the density of the points. It is used to illustrate the relationship between two variables, for instance, GPP and LAI (see section on 'GPP vs. LAI'). To create the density-scatter plots, we divide the ranges of the two variables of interests (e.g., GPP and LAI) to 100 intervals along the x- and the y-axis, which thus divides the plotting area into 100×100 bins. The number of points that fall into a bin defines its density and is shown by a color. Density-scatter plots are more suitable than traditional scatter plots to illustrate the distribution of a large number of points ($\sim10^5$ in our case) on the x–y plane.

Results and discussion

Annual GPP

Annual GPP of forests and nonforests in the T–P domain are shown in Figs 3 and 4, respectively. For forests, GPP ranges from 100 to above $2000\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}$, except that GPP of CASA (GPP of CASA here is estimated as twice NPP – see Table 1) is mostly below $1500\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}$. These values are generally consistent with those of Beer *et al.* (2010) over North America. In general, lower GPP is distributed in colder/drier climate zones while higher GPP is found over warmer/moister regimes, indicating a positive relationship between GPP and both climate variables.

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

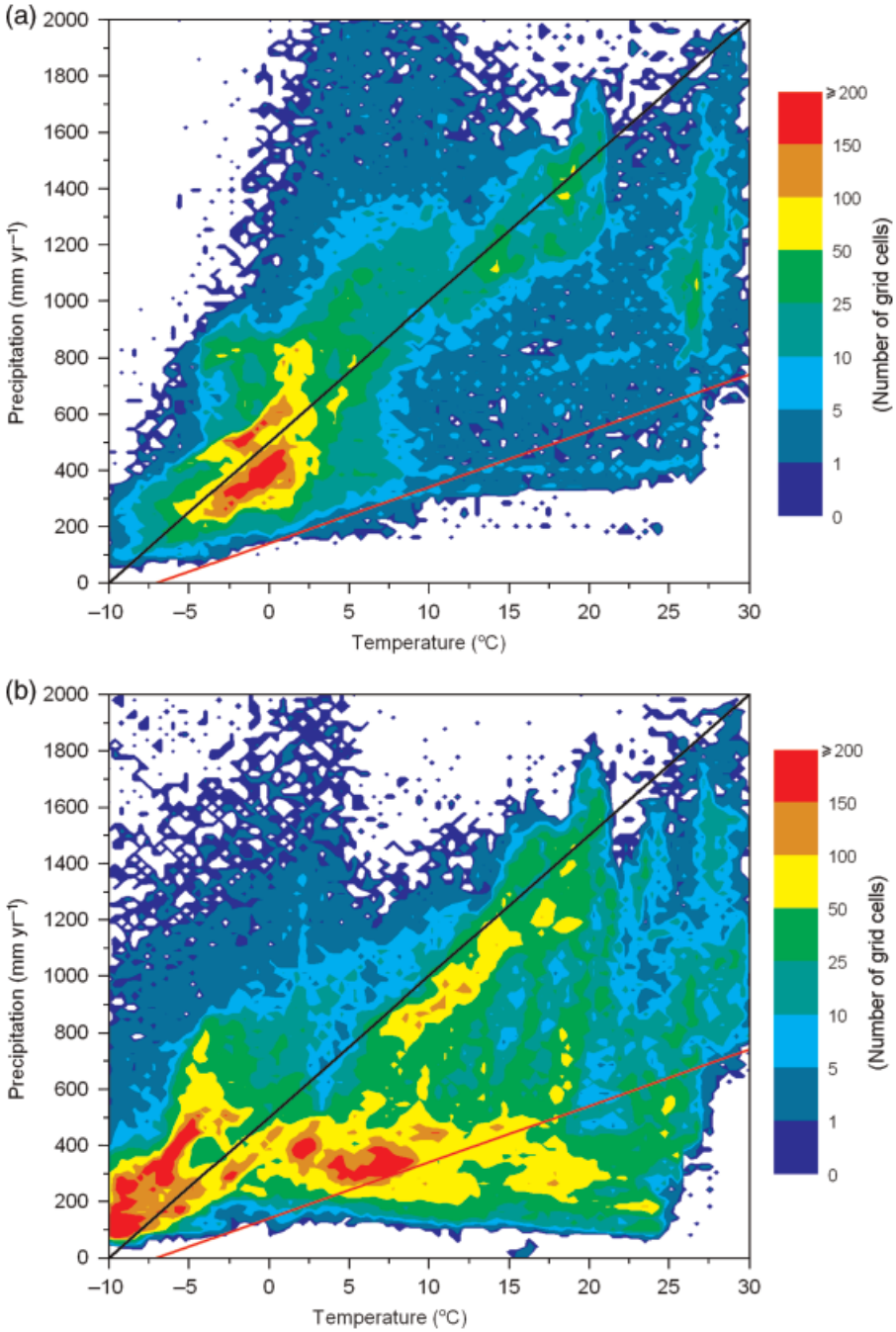

Fig. 2 Density distributions of geographic grid cells in the climate (T–P) domain: (a) forests; (b) nonforests. The color bar indicates the numbers of grid cells per T–P bin. The red line represents the dry/moist climate boundary (defined by $P = 20T + 140$) and the black line is the diagonal of the diagram (defined by $P = 50T + 500$).

The center of the contour levels coincides with the upward diagonal of $P = 50T + 500$, along which the changes of GPP, or the gradient, are rather uniform (though they vary among different models). The GPP contour lines do not generally follow lines of constant temperature or precipitation, but tilt towards the upward diagonal: below the diagonal (i.e., the 'warm–dry' corner) the horizontal contour lines are inclined towards higher precipitation as temperature increases, indicating that for equal GPP more water is required in warmer environments; above the diagonal (the 'cold–wet' corner) the vertical contour lines are inclined

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

Colour

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

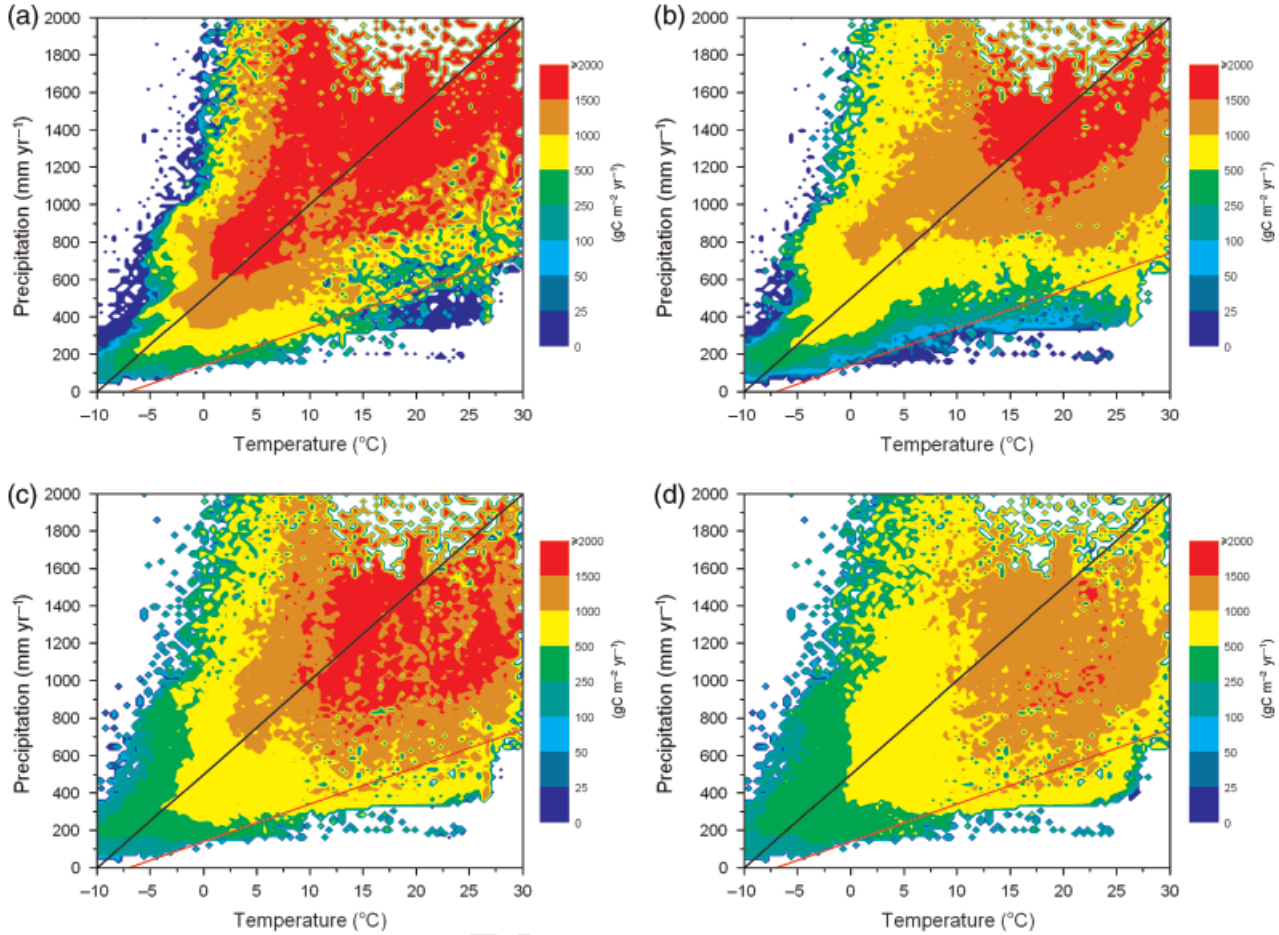

Fig. 3 Annual gross primary production (GPP, g C m^{-2} yr^{-1}) of forests in the *T–P* domain: (a) BIOME-BGC; (b) LPJ; (c) TOPS-BGC; and (d) CASA.

towards higher temperature as precipitation increases, likely induced by lower solar radiation levels associated with extra rainfall.

In Fig. 3, regions with the highest GPP values reflect the most favorable climate conditions for forest productivity. Because these regions are located close to $P = 50T + 500$, their characteristics may be quantified by examining GPP variations along this line. Starting from the origin (−10 °C and 0 mm), GPP increases as temperature and precipitation increase: the gradient of GPP is about 125, 70, and 50 g C m^{-2} yr^{-1} °C^{-1} in BIOME-BGC, LPJ and TOPS-BGC, and CASA, respectively (the corresponding gradients with regard to precipitation are 2.5, 1.4, and 1.0 g C m^{-2} yr^{-1} mm^{-1}, respectively). As such, the most favorable climate region (where GPP reaches above 1500 g C m^{-2} yr^{-1}) starts shortly above 0 °C in BIOME-BGC, but it starts around 10 °C in LPJ and TOPS-BGC. In CASA, because the highest GPP level is actually 1000–1500 g C m^{-2} yr^{-1}, its favorable climate regions also start around 10 °C. When in the favorable climate regions, GPP becomes steady within certain temperature ranges. As temperature increases further, it may eventually become too hot and unfavorable for vegetation productivity. However, the effects of high temperature on GPP are not very evident in the chosen *T–P* domain, as the upper temperature limit of the favorable climate regions is about 27 °C for TOPS-BGC and CASA and up to 30 °C for BIOME-BGC and LPJ.

Compared with forests, the GPP patterns of nonforests show more diversity among the models (Fig. 4). LPJ stands out from the other models in that the distribution patterns of its GPP over these nonforest regions are almost the same as those of forests. As mentioned earlier, this is because vegetation simulated by LPJ is solely regulated by climate, and such 'potential vegetation' would have similar behavior under similar climate conditions. In the other models, GPP of nonforests is clearly different from that of forests. For TOPS-BGC and CASA, the main differences are in magnitude, which ranges between 500 and 1000 g C m^{-2} yr^{-1} in warm/

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

Colour

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

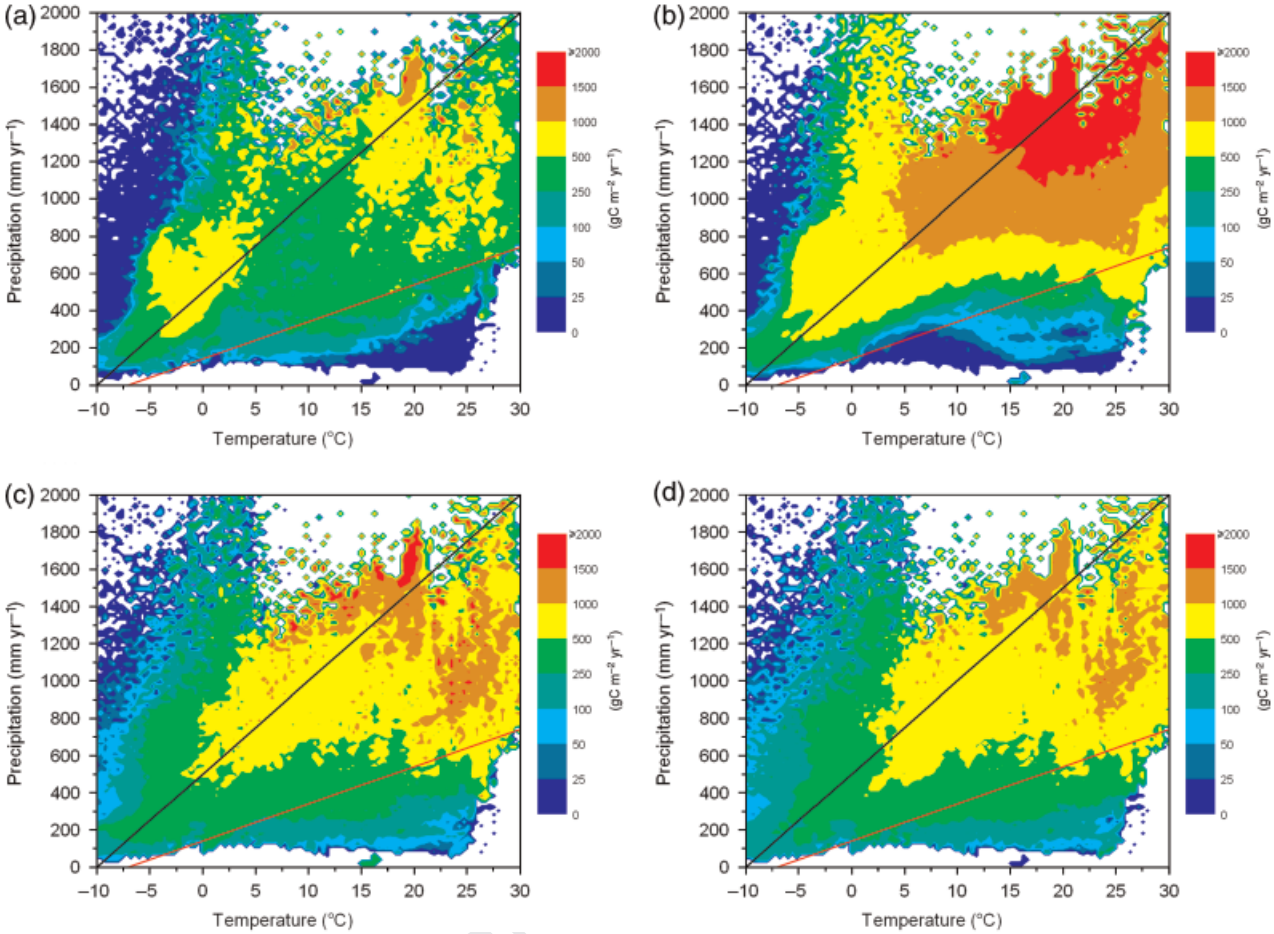

Fig. 4 Same as Fig. 3 but for nonforests or, in the case of LPJ, vegetation simulated in nonforest regions: (a) BIOME-BGC; (b) LPJ; (c) TOPS-BGC; and (d) CASA.

moist regions and rarely reaches above 1500 g C m^{-2} yr^{-1}; low GPP values (under 100 g C m^{-2} yr^{-1}) also become evident in unfavorable (cold/wet or hot/dry) climate regions. On the other hand, the distribution patterns of nonforest GPP in TOPS-BGC and CASA share a characteristic with those of forests: the centers of contours levels are largely aligned with $P = 50T + 500$, on which the GPP gradient is about 30 g C m^{-2} yr^{-1} 1 °C^{-1}. Indeed, the distribution patterns are very similar between TOPS-BGC and CASA (Fig. 4).

The strongest contrast in GPP between nonforests and forests is found in BIOME-BGC. Although BIOME-BGC has the highest forest GPP among the models, its nonforest GPP is the lowest (Table 2), which is under 1000 g C m^{-2} yr^{-1} even in some favorable climate conditions (Fig. 4). The distribution patterns of nonforest GPP also differ from those of forests. Along $P = 50T + 500$, for instance, there is a notable region (5–15 °C) where GPP is under 500 g C m^{-2} yr^{-1}, while it is above 500 g C m^{-2} yr^{-1} in adjacent colder (−5–5 °C) or warmer (15–25 °C) climate regions. This feature gives the distribution of nonforest GPP in BIOME-BGC a bimodal appearance.

The major ecosystems in the low productive regions discussed above are grasslands/croplands. Therefore, the results of Fig. 4a suggest some issues of BIOME-BGC in parameterizing these PFTs. Additional analysis reveals that the low GPP of grasslands and croplands are indeed induced by their relatively high fire mortality, which is assumed 10% per year (in comparison, the fire mortality of forests is about 0.2–0.5% per year). Reducing fire rates can significantly increase the productivity of these PFTs. (Reasons why vegetation in BIOME-BGC is so sensitive to fire in the first place will be discussed in NPP and CUE).

NPP and CUE

Because NPP in TOPS-BGC and CASA is estimated as a constant proportion (i.e., 50%) of GPP, their distribution patterns are identical to those of GPP and thus need not

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

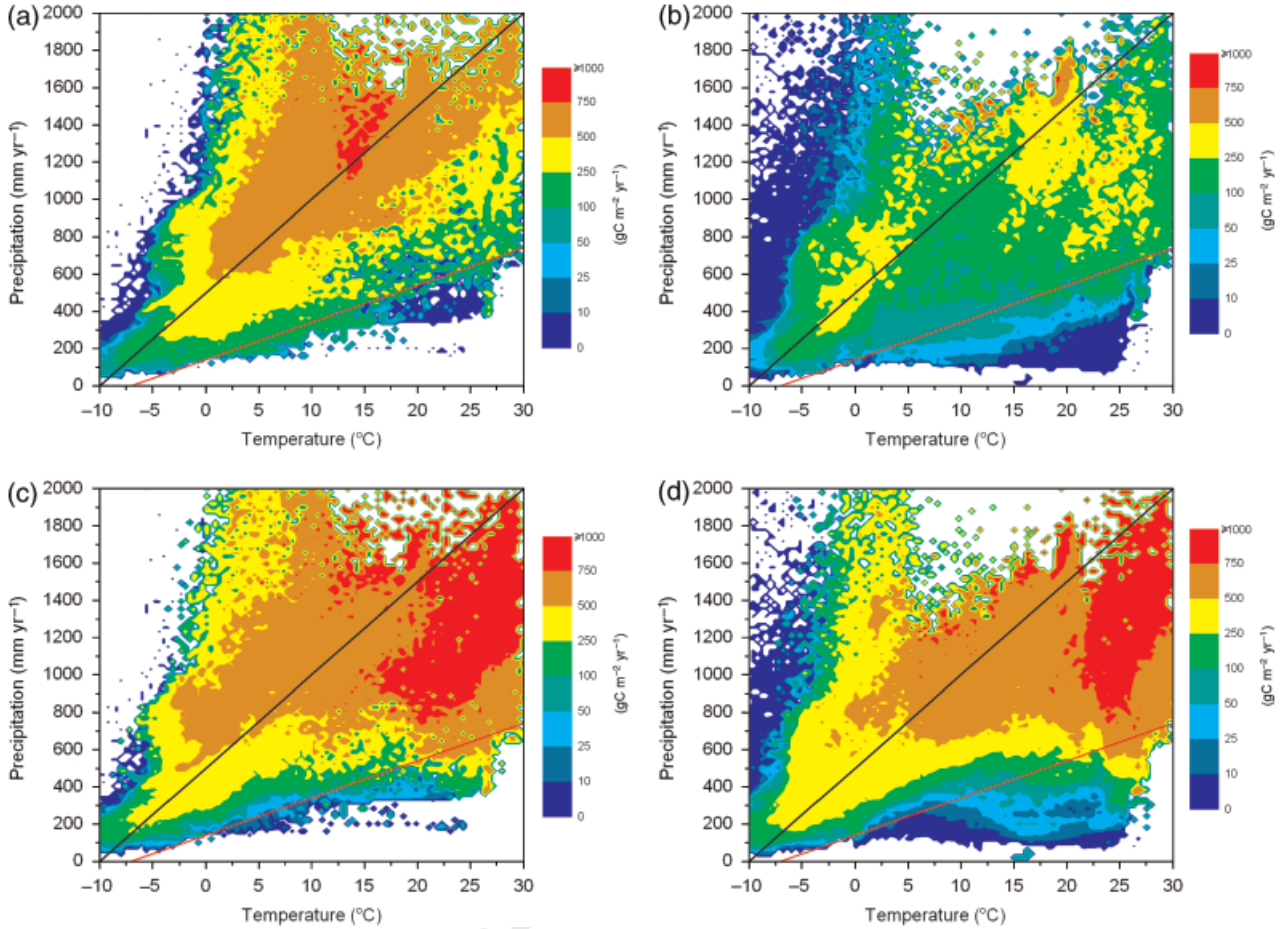

Fig. 5 Annual net primary production (NPP, g C m^{-2} yr^{-1}) in the *T*–*P* domain: (a) forests of BIOME-BGC; (b) non-forests of BIOME-BGC; (c) forests of LPJ; (d) nonforests of LPJ (vegetation simulated in nonforest regions).

to be further discussed. Below, we focus only on NPP of BIOME-BGC and LPJ (Fig. 5).

For both BIOME-BGC and LPJ, NPP ranges mostly between 0 and 1000 g C m^{-2} yr^{-1}, about half of the magnitude of GPP. The distribution patterns of NPP generally follow those of GPP, with the centers of contour levels largely aligned with $P = 50T + 500$. However, there are also noticeable differences. In LPJ, for example, the highest NPP (>750 g C m^{-2} yr^{-1}) is mostly distributed in climate regions with temperature above 20 °C and precipitation above 700 mm yr^{-1}. Compared with the corresponding patterns of GPP, high NPP is apparently shifted towards hot climate zones, and implies greater tolerance of drought conditions. For BIOME-BGC, the most remarkable difference is the reduction in the highest NPP: even for forests, NPP reaches above 750 g C m^{-2} yr^{-1} only in limited climate conditions (*T* around 15 °C and *P* around 1000–1400 mm yr^{-1}). Therefore, BIOME-BGC has lower overall NPP than LPJ, although its overall GPP is higher (Table 2).

The relative changes between NPP and GPP can be evaluated with the carbon use efficiency (CUE; DeLucia *et al.*, 2007), that is, the NPP-to-GPP ratio (Fig. 6). The CUE of BIOME-BGC ranges from 10% to 60%, but is mostly below 50% (Fig. 6a and b). In general, CUE is higher (~50%) in regions close to $P = 50T + 500$ and is lower (<20%) in unfavorable climate conditions. For forests; however, CUE also decreases to <30% when temperature is above 20 °C (Fig. 6a). This temperature-dependent feature is not evident for CUE of nonforests, which is rather uniform (40–50%) in most climate conditions. The overall CUE of BIOME-BGC is ~35% and ~42% for forests and nonforests, respectively.

The CUE of LPJ ranges from 40% to 70% in most climate regions (Fig. 6c and d). It is about 56% for both forest and nonforest regions, higher than BIOME-BGC. In addition, LPJs distribution of CUE shows apparent dependence on climate variables (temperature, in particular). CUE is mostly above 60% when temperature is below 0 °C; it drops to <60% for 0–12 °C, and further

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

Colour
(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

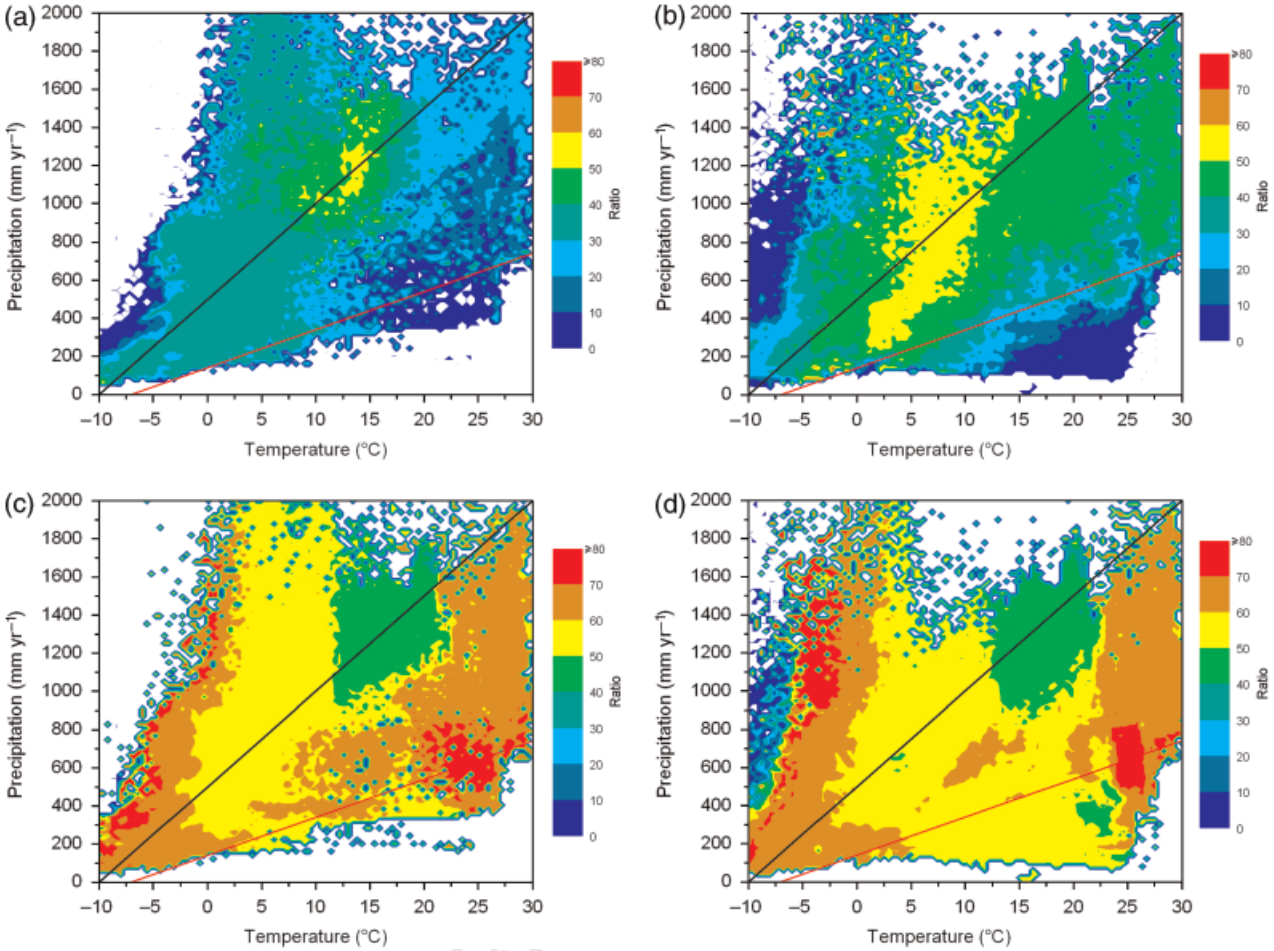

Fig. 6 Carbon use efficiency (CUE, unitless) in the *T*–*P* domain: (a) forests of BIOME-BGC; (b) nonforests of BIOME-BGC; (c) forests of LPJ; (d) nonforests of LPJ (vegetation simulated in nonforest regions).

decreases to 40–50% when temperature increases to 12–22 °C (with precipitation above 800 mm yr^{-1}). Therefore, the CUE of LPJ generally decreases with increases of temperature in cold and temperate climate zones. However, this trend is sharply reversed in hot climate regions, where CUE jumps to 60–70% when temperature is above 22 °C (Fig. 6c and d).

Because NPP in these models is calculated as GPP less AR, we explore the variations of CUE from the aspect of AR. In general, AR increases exponentially with temperature (at a rate specified by the Q_{10} coefficient). This explains the negative gradient of CUE vs. temperature as seen in the results of BIOME-BGC (forests, in particular) over hot climate zones (Fig. 6a), and in the results of LPJ over cold-temperate climate regions (Fig. 6c and d). In contrast, the sharp increase of CUE in LPJ over hot climate zones goes against this reasoning, suggesting parameter changes in the calculation of AR. Indeed, in LPJ the reference respiration rate for tropical forests is only about 20% of that for temperate and boreal forests (Sitch *et al.*, 2003). On the other hand, the rather uniform CUE of BIOME-BGC is induced by the fact that GPP in BIOME-BGC is often retrospectively adjusted based on nitrogen availability (see section 'Model experiment protocol'), which decreases the sensitivity of CUE to AR.

GPP vs. LAI

It has been long recognized that a linear relationship generally exists between annual GPP and leaf area duration, that is, the product of LAI and growing season length (Kira & Shidei, 1967; Waring & Schlesinger, 1985). Therefore, below we examine the GPP–LAI relationship among the tested models. In particular, we compare annual GPP both to annual mean LAI and to annual maximum LAI. Because the results of forests and nonforests either have similar features (TOPS-BGC and CASA) or their features do not overlap with one another (BIOME-BGC and LPJ), they are represented in the

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

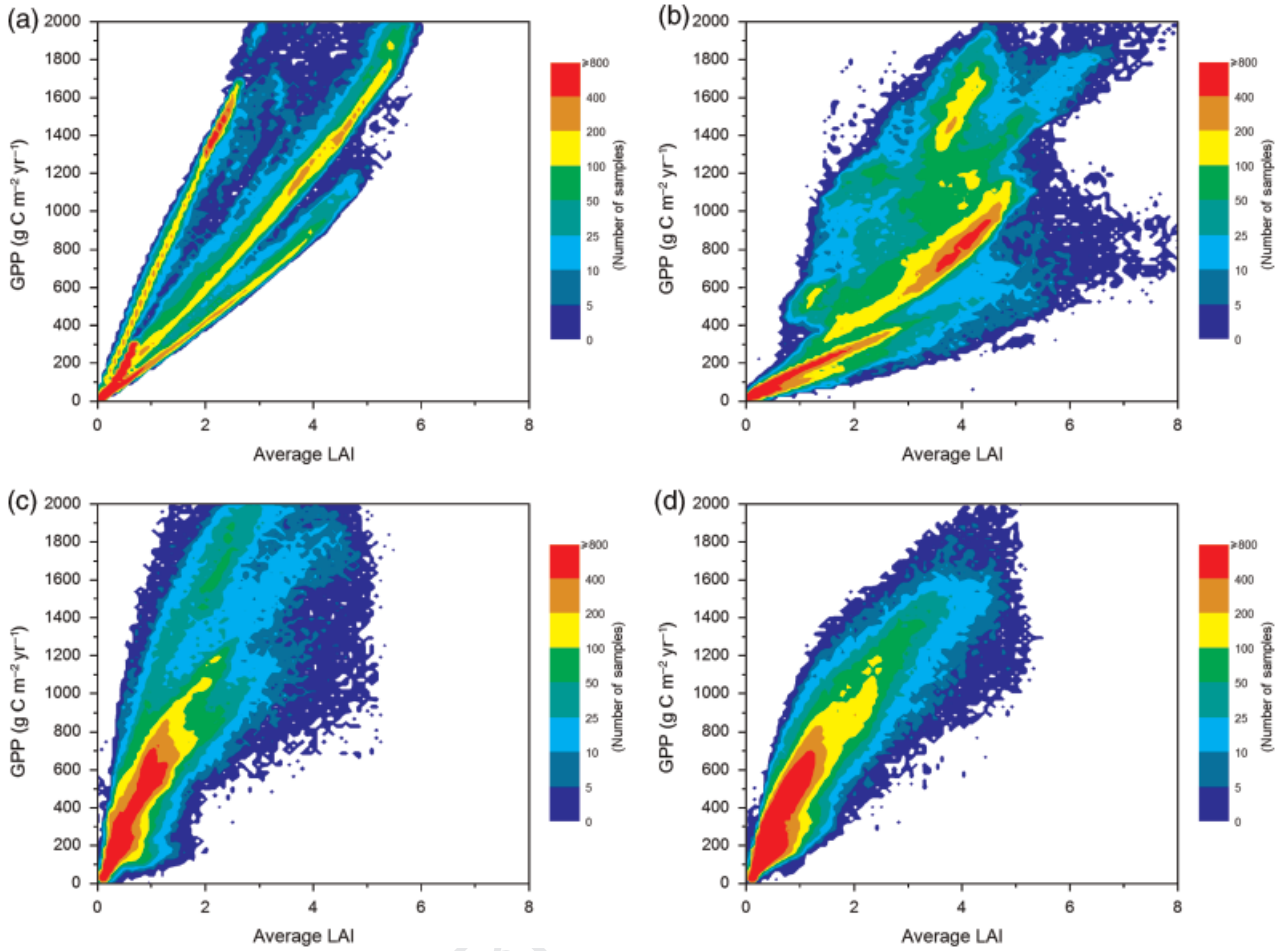

Fig. 7 Density scatter plots between annual gross primary production (GPP, g C m^{-2} yr^{-1}) and annual mean leaf area index (LAI, unitless): (a) BIOME-BGC; (b) LPJ; (c) TOPS-BGC; and (d) CASA. The color scheme shows the number of samples represented by each point of the scatter plot (i.e., the density).

same plots. For reference, diagrams of maximum LAI in the *T*–*P* domain are given in the Supporting Information (Figs S1 and S2).

Overall, the relationship between annual GPP and annual mean LAI (Fig. 7) is largely linear, although it has different characteristics among the models. The GPP–LAI (mean) relationship of BIOME-BGC (Fig. 7a) is composed by multiple strong linear components that represent different PFTs. Components contributed by forests are most distinguishable: the one representing deciduous broadleaf forests has LAI mostly under 3 but its GPP/LAI slope is the highest of all the components, about 600 g C m^{-2} yr^{-1} per LAI; those representing needleleaf/broadleaf evergreen forests have LAI reaching 5–6 and lower slopes about 300 g C m^{-2} yr^{-1} per LAI. Another less distinctive component that represents nonforests (grasses/crops, in particular) has annual mean LAI mostly <1, though its GPP/LAI slope may be higher than that of the evergreen forests.

The multicomponent linear relationship is also evident in the scatter plot of LPJ (Fig. 7b). However, these components are much more dispersed than those of BIOME-BGC and are characterized by centers with high scatter densities. For instance, there is one major component with LAI under 3 and GPP under 400 g C m^{-2} yr^{-1}. Its low productivity suggests that this is mainly comprised of nonforests. In comparison, another major component is distributed with its center at LAI of 4 and GPP of 800 g C m^{-2} yr^{-1}, contributed by samples with more forest dominance. There is also a minor component centered at LAI of 4 and GPP of 1500 g C m^{-2} yr^{-1}, representing cells dominated by forests. The overall GPP/LAI slope of the scatter plot is about 200–300 g C m^{-2} yr^{-1} per LAI. Another feature of LPJs plot is that, although the mean LAI is mostly under 6, it reaches values as high as 8 in some cases.

The GPP/LAI relationships of TOPS-BGC and CASA look similar (Fig. 7c and d). They show few features that can

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

Colour

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

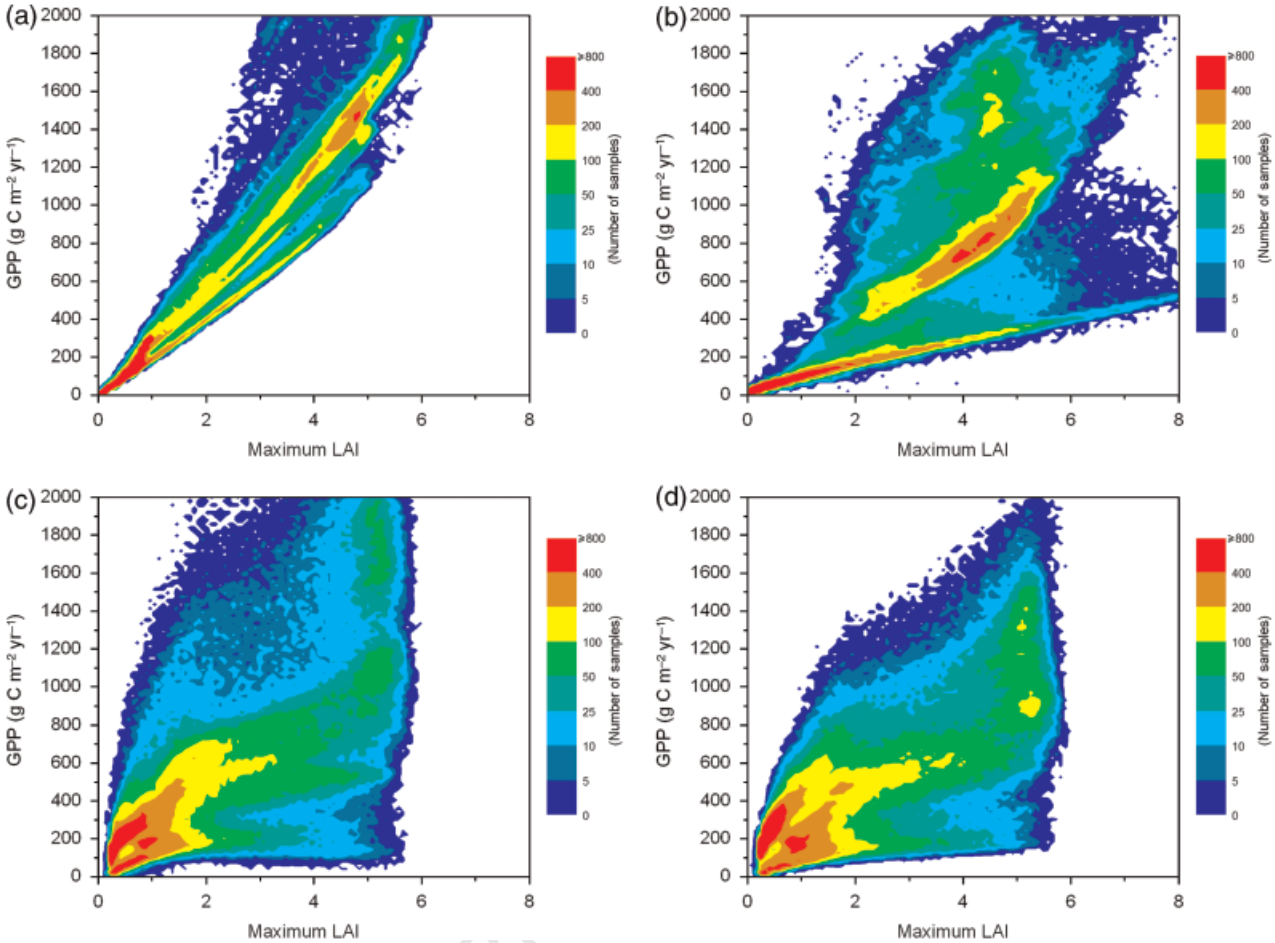

Fig. 8 Same as Fig. 7 but between annual gross primary production (GPP, g C m^{-2} yr^{-1}) and annual maximum leaf area index (LAI, unitless): (a) BIOME-BGC; (b) LPJ; (c) TOPS-BGC; and (d) CASA.

be tied to specific PFTs, but have an overall nearly linear shape. These variables are less tightly coupled than in BIOME-BGC but less dispersed than those in LPJ. The GPP/LAI slope is about 500 g C m^{-2} yr^{-1} per LAI in both models, except that in CASA, the slope is slightly reduced when mean LAI is above 2, reflecting lower productivity of forests in the model.

We further tested the relationship between annual GPP and annual maximum LAI of all the models (Fig. 8). The outstanding feature of the results is that they reveal the difference between models that simulate LAI (BIOME-BGC and LPJ) and models that do not (TOPS-BGC and CASA). The multicomponent linear relationship is preserved between GPP and maximum LAI in BIOME-BGC and LPJ (Fig. 8a and b). Indeed, the two scatter plots are very similar to the corresponding plots between GPP and mean LAI (Fig. 7a and b). The differences are mainly induced by rescaling (or stretching) the components of deciduous forests (BIOME-BGC) or nonforests (LPJ) along the 'x' axis (representing LAI) such that the GPP–LAI slopes of these deciduous components are lowered by a factor of about 2 (Fig. 8a and b). At the same time, the evergreen components largely remain unchanged.

On the other hand, the scatter plots between annual GPP and annual maximum LAI of TOPS-BGC and CASA (Fig. 8c and d) are distinctively different from those between GPP and mean LAI (Fig. 7c and d). Indeed, the scatter plots of Fig. 8c and d are so dispersed that GPP may range from almost 0 to above 800 g C m^{-2} yr^{-1} with LAI around 1 (Fig. 8c and d). Such a wide range of values prevents any meaningful relationship to be inferred between the two variables.

Because the difference between annual mean LAI and annual maximum LAI is that the former integrates information about the seasonal cycle of LAI (i.e., leaf phenology) whereas the later does not, the different behaviors between the two groups of models reflect their differences in modeling leaf phenology. Leaf phenology is explicitly simulated in BIOME-BGC and LPJ but is

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

implicitly inferred from satellite measurements (LAI) in TOPS-BGC and CASA. The results of Figs 7 and 8 indicate that the simulated phenology is temporally more regular and/or spatially more homogeneous, so that mean LAI can be inferred from maximum LAI by a simple relationship. For instance, comparing Figs 7a and b with 8a and b suggests that in BIOME-BGC and LPJ there is little seasonal variation in LAI of evergreen PFTs, and the mean LAI is largely half of the corresponding maximum LAI for deciduous PFTs. In contrast, phenology implied by the satellite data is more irregular and heterogeneous (White *et al.*, 2009), which makes the scatter plots using maximum LAI (Fig. 8c and d) much more dispersed than those using mean LAI (Fig. 7c and d).

There are other processes and interactions that play an important role in shaping the GPP–LAI relationship in BIOME-BGC and LPJ. The simulation of LAI may be explained using the analogy of 'carbon investment'. That is, plants invest carbon assimilated through photosynthesis (GPP) to develop more leaves (and supporting tissues) in the expectation of more carbon (increases in GPP) as the return; the cost associated with the investment is AR. Because the marginal gain of GPP decreases as LAI increases (due to the saturation of FPAR) while AR increases during the same course, a balance point may be reached when the rate of return diminishes and NPP reaches its maximum. However, a balance point of this kind is not checked in BIOME-BGC or LPJ, likely for the reason that the simulated LAI could be unrealistically high even before the balance point is reached. Instead, BIOME-BGC and LPJ implements other strategies to constrain the growth of LAI.

In BIOME-BGC, such a constraint is provided by soil nitrogen availability. In determining how much carbon to allocate, BIOME-BGC needs to evaluate both carbon and nitrogen budgets. If nitrogen is limited (in most cases), only a proportion of available carbon is actually allocated. The surplus of carbon is then subtracted from the original GPP (see section on 'Model experiment protocol'). As such, the GPP of BIOME-BGC is substantially influenced by the actually allocated carbon, and this is the reason for the strong linear relationship between GPP and LAI. On the other hand, the constraint to LAI growth in LPJ is implemented as a self-thinning process: when the density of forests is above a certain limit, or when the size of an individual tree is too big, the extra portion of vegetation tissue is forced to be shed as litter (Sitch *et al.*, 2003). However, nonwoody species (nonforests) in LPJ are not limited by their density and are able to grow without limit. As such, LAI of nonforests in LPJ can be higher than that of forests; also, the GPP–LAI relationship of nonforests is more linear (Fig. 8b).

The discussions of GPP–LAI interactions in BIOME-BGC help explain its different GPP patterns between forests and nonforests. Among the tested models, BIOME-BGC has the highest GPP over forests but the lowest GPP over nonforests (Table 2). A direct reason for the low production of nonforests is that BIOME-BGC can not efficiently produce leaf area for those ecosystems, even when the production is not limited by climate (Figs 4 and S2). Yet the deeper reason that nonforests (grasses and crops, in particular) fail to generate sufficient leaf area is that they have a much higher rate of fire (10% per year) than those of forests (0.2–0.5% per year), through which a lot more nitrogen is lost. Therefore, grasslands and croplands in BIOME-BGC are generally in greater need of exogenous nitrogen than forests, and the tight coupling of carbon and nitrogen cycles in BIOME-BGC makes the model very sensitive to parameters such as annual fire rate and nitrogen deposition rate. This example, along with all the other issues of different models discussed above, emphasizes the impacts of different modeling strategies on model behaviors and highlights the structural uncertainties that can be induced in the simulated results.

Finally, limitations of the current study should be recognized. Although we aim at structural differences among ecosystem models, some of the discussed results are clearly dependent on the parameter values chosen for each ensemble member. For instance, if we increase the LUE coefficients of forests in CASA by 20%, the magnitudes of its simulated GPP will be more comparable to the other three models. Also, the above discussions mainly focus on the mean of ecological variables in the *T*–*P* space, while their variance has barely been analyzed. Solutions to these issues are very important for us to further quantify uncertainties among ecosystem models, and we should address them in future studies.

Conclusions

This study reports results from a MME experiment designed to diagnose and assess sources of uncertainties among terrestrial ecosystem models. The experiment was conducted under the TOPS, and the tested models include public versions of BIOME-BGC, LPJ, TOPS-BGC, and CASA. We developed the HFDEM to systematically analyze outputs of the model experiments. HFDEM separates the simulated biogeochemistry into a cascade of three functional tiers and sequentially examines their characteristics in climate (temperature–precipitation) and other spaces. This paper focuses on how carbon is assimilated through GPP and how it is partitioned between AR and net carbon gain.

Analyzing model results in the climate (*T*–*P*) domain indicates a general agreement on GPP among the models, which is particularly evident in the distribution

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x

patterns. GPP tends to be optimal in climate regions where the relationship between annual temperature and precipitation is defined by $P = 50T + 500$. Along this line, GPP generally increases by $\sim 70\,\mathrm{g\,C\,yr^{-1}\,m^{-2}}$ (forests) or by $\sim 30\,\mathrm{g\,C\,yr^{-1}\,m^{-2}}$ (nonforests) for a 1 °C increase in temperature (or 50 mm increase in precipitation) in cold/temperate climate regimes. The consensus of such broad patterns reflects the dominant control of climate variables on simulated ecosystem production. Also, strong relationships are found between annual GPP and annual mean LAI in all the models, suggesting the latter as a good indicator to estimate vegetation production.

Differences in patterns and magnitudes of GPP are also identified. For instance, BIOME-BGC has the highest GPP for forests but the lowest GPP for nonforests among the tested models, while the corresponding contrast of GPP is much smaller in the other models. A major source of such differences can be traced back to how LAI is simulated in the models. For BIOME-BGC and LPJ, in particular, there is a positive feedback between the growth of LAI and GPP. Different strategies (e.g., nitrogen limitation vs. self-thinning) to constrain such positive feedbacks lead to distinctive characteristics of the two models. The assimilation of satellite measurement of LAI in TOPS-BGC and CASA relieves them from the difficulty of modeling GPP–LAI interactions, yet uncertainties may be introduced by the satellite datasets themselves.

NPP in TOPS-BGC and CASA is assumed as a constant ratio (50%) of GPP, whereas in BIOME-BGC and LPJ it is simulated through independent estimation of AR. The independently simulated NPP largely follows the patterns of GPP, and the corresponding NPP-to-GPP ratios (CUE) are close to 50% on average. However, this coincidence could be a result of model calibration. In spite of the above agreement, separating respiration from GPP remains a challenge for both BIOME-BGC and LPJ. In BIOME-BGC, this process is complicated by the regulation of nitrogen availability; while in LPJ, different parameterizations of AR in temperate and tropical forests result in discrepancies of CUE between the two PFTs.

The goal of this study is mainly to develop a general diagnostic framework (i.e., HFDEM) to systematically analyze and compare the characteristics of different ecosystem models. The developed methodology can also be applied to analyze observational datasets and facilitate model-data comparisons. For instance, the same *T–P* diagrams can be used to describe distributions of observed ecological variables in climate space. By neglecting geographic coordinates, the *T–P* diagrams allow model results to be compared with reference data from different regions, which helps increase the sample size of observations in the comparison. Indeed, this feature also allows model results from different experiments to be compared. Therefore, by increasing the sample size of ecosystem models and reference data in ensemble experiments, the potential exists to gain a more comprehensive understanding of the causes of structural uncertainty and therefore constrain models more intelligently in the future.

Acknowledgements

The dataset GSOD is from NOAA NCDC (http://www.ncdc.noaa.gov). The dataset TD3200 is from the Research Data Archive (RDA), which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation (NSF). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds510.0. This research was funded by grants from the NASA Earth Science program. The authors thank two anonymous reviewers for their insightful comments and constructive suggestions.

References

Adams B, White A, Lenton TM (2004) An analysis of some diverse approaches to modelling terrestrial net primary productivity. *Ecological Modelling*, **177**, 353–391.

Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology*, **9**, 479–492.

Beer C, Reichstein M, Tomelleri E *et al*. (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* (in press). Q7

Canadell JG, Gifford R, Houghton R *et al*. (2007) Saturation of the terrestrial carbon sink. In: *Terrestrial Ecosystems in a Changing World* (eds Canadell JG, Pataki D, Pitelka L), pp. 59–78. Springer-Verlag, Berlin.

Churkina G, Running SW (1998) Contrasting climatic controls on the estimated productivity of global terrestrial biomes. *Ecosystems*, **1**, 206–215.

Collatz GJ, Ball JT, Grivet C, Berry JA (1991) Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. *Agricultural and Forest meteorology*, **54**, 107–136.

Cramer W, Bondeau A, Woodward FI *et al*. (2001) Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373.

Cramer W, Kicklighter DW, Bondeau A *et al*. (1999) Comparing global models of terrestrial net primary productivity (NPP): overview and key results. *Global Change Biology*, **5** (Suppl. 1), 1–15.

DeLucia EH, Drake JE, Thomas RB, Gonzalez-Meler M (2007) Forest carbon use efficiency: is respiration a constant fraction of grass primary production? *Global Change Biology*, **13**, 1157–1167.

Denning AS, Holzer M, Gurney KR *et al*. (1999) Three-dimensional transport and concentration of SF6: a model intercomparison study (TransCom 2). *Tellus*, **51B**, 266–297.

Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta*, **149**, 78–90.

Friedl MA, McIver DK, Hodges JCF *et al*. (2002) Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, **83** (1–2), 287–302.

Ganguly S, Mitchell AS, Samanta A *et al*. (2008) Generating vegetation leaf area index earth system data records from multiple sensors. Part 1: theory. *Remote Sensing of Environment*, **112**, 4333–4343.

Goodale CL, Apps MJ, Birdsey RA *et al*. (2002) Forest carbon sinks in the northern hemisphere. *Ecological Applications*, **12**, 891–899.

Gurney KR, Law RM, Denning AS *et al*. (2002) Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, **415**, 626–630.

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

Q8 Gurney KR, Law RM, Denning AS *et al*. (2004) Transcom 3 inversion intercomparison: model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochemical Cycles*, **18**, doi: 10.1029/2003GB002111.

Haxeltine A, Prentice IC (1996) BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles*, **10**, 693–709.

Heinsch FA, Zhao M, Running SW *et al*. (2006) Evaluation of remote sensing based terrestrial productivity from MODIS using tower eddy flux network observations. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1908–1925.

Held IM (2005) The gap between simulation and understanding in climate modeling. *Bulletin of American Meteorological Society*, **86**, 1609–1614.

Houghton RA (2007) Balancing the global carbon budget. *Annual Review of Earth and Planetary Sciences*, **35**, 13–47.

Ichii K, White MA, Votava P, Michaelis A, Nemani RR (2008) Evaluation of snow models in terrestrial biosphere models using ground observation and satellite data: impact on terrestrial ecosystem processes. *Hydrological Processes*, **22**, 347–355.

IPCC (2007) *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, Cambridge, UK.

Q9 Jacobson AR, Mikaloff Fletcher SE, Gruber N, Sarmiento JL, Gloor M (2007) A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Global Biogeochemical Cycles*, **21**, doi: 10.1029/2005GB002556.

Jolly WM, Graham JM, Michaelis A, Nemani RR, Running SW (2005) A flexible, integrated system for generating meteorological surfaces derived from point sources across multiple geographic scales. *Environmental Modeling and Software*, **20**, 873–882.

Kira T, Shidei T (1967) Primary production and turnover of organic matter in different forest ecosystems of the western pacific. *Japanese Journal of Ecology*, **17**, 70–87.

Q10 Luyssaert S, Inglima I, Jung M *et al*. (2007) CO_2 balance of boreal, temperate, and tropical forests derived from a global database. *Global Change Biology*, **13**, 2509–2537.

Q11 Mahadevan P, Wofsy SC, Matross DM *et al*. (2008) A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: vegetation photosynthesis and respiration model (VPRM). *Global Biogeochemical Cycles*, **22**, doi: 10.1029/2006GB002735.

Masson V, Champeaux J-L, Chauvin F, Meriguet C, Lacaze R (2003) A global database of land surface parameters at 1-km resolution in meteorological and climate models. *Journal of Climate*, **16**, 1261–1282.

Myneni RB, Dong J, Tucker CJ *et al*. (2001) A large carbon sink in the woody biomass of northern forests. *Proceedings of the National Academy of Sciences USA*, **98**, 14,784–14,789.

Nemani R, Hashimoto H, Votava P *et al*. (2009) Monitoring and forecasting protected area ecosystem dynamics using the Terrestrial Observation and Prediction System (TOPS). *Remote Sensing of Environment*, **113**, 1497–1509.

Nemani RR, Keeling CD, Hashimoto H *et al*. (2003) Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, **300**, 1560–1563.

Pacala SW, Hurtt GC, Baker D *et al*. (2001) Consistent land- and atmosphere-based U.S. carbon sink estimates. *Science*, **292**, 2316–2320.

Peters W, Jacobson AR, Sweeney C *et al*. (2007) An atmospheric perspective on North American carbon dioxide exchange: carbontracker. *Proceedings of the National Academy of Sciences USA*, **104**, 18925–18930.

Potter C, Klooster S, Myneni R, Genovese V, Tan P-N, Kumar V (2003) Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982–1998. *Global and Planetary Change*, **39**, 201–213.

Potter CS, Randerson JT, Field CB, Matson PA, Vitousek PM, Mooney HA, Klooster SA (1993) Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles*, **7**, 811–841.

Randerson JT, Hoffman FM, Thornton PE *et al*. (2009) Systematic assessment of terrestrial biogeochemistry in coupled climate-carbon models. *Global Change Biology*, **15**, 2462–2484.

Rayner PJ, Enting IG, Francey RJ, Langenfelds R (1999) Reconstructing the recent carbon cycles from atmospheric CO_2 $\delta^{13}C$ and O_2/N_2 observations. *Tellus*, **51B**, 213–232.

Rayner PJ, Law RM (1995) *A comparison of modelled responses to prescribed CO_2 sources*. CSIRO Division of Atmospheric Research Technical Paper No. 36.

Schimel D (2007) Carbon cycle conundrums. *Proceedings of the National Academy of Sciences USA*, **104**, 18353–18354.

Schloss AL, Kicklighter DW, Kaduk J, Wittenberg Uand the participants of the Potsdam NPP Model Intercomparison (1999) Comparing global models of terrestrial net primary productivity (NPP): comparison of NPP to climate and the normalized difference vegetation index. *Global Change Biology*, **5** (Suppl. 1), 25–34.

Sibson R (1981) A brief description of natural neighbor interpolation. In: *Interpreting Multivariate Data* (ed. Barnett V), pp. 21–36. John Wiley, Chichester, UK.

Sitch S, Smith B, Prentice IC *et al*. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161–185.

Stephens BB, Gurney KR, Tans PP *et al*. (2007) Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2. *Science*, **316**, 1732–1735.

Strahler AN, Strahler AH (1978) *Modern Physical Geography*. John Wiley & Sons, New York.

Tans PP, Fung IY, Takahashu T (1990) Observational constraints on the global atmospheric CO_2 budget. *Science*, **247**, 1431–1439.

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A*, **365**, 2053–2075.

Thornton PE, Law BE, Gholz HL *et al*. (2002) Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests. *Agricultural and Forest Meteorology*, **113**, 185–222.

Thornton PE, Rosenbloom NA (2005) Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecological Modelling*, **189**, 25–48.

Thornton PE, Running SW, White MA (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214–251.

Tucker CJ, Pinzon JE, Brown ME *et al*. (2005) An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, **26**, 4485–5598.

VEMAP Members (1995) Vegetation/Ecosystem Modeling and Analysis Project (VEMAP): comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. *Global Biogeochemical Cycles*, **9**, 407–437.

Wang W, Ichii K, Hashimoto H, Michaelis AR, Thornton PE, Law BE, Nemani RR (2009) A hierarchical analysis of terrestrial ecosystem model Biome-BGC: equilibrium analysis and model calibration. *Ecological Modelling*, **220**, 2009–2023.

Waring RH, Landsberg JJ, Williams M (1998) Net primary production of forests: a constant fraction gross primary production? *Tree Physiology*, **18**, 129–134.

Waring RH, Schlesinger WH (1985) *Forest Ecosystems: Concepts and Management*. Academic Press, New York.

White MA, De Beurs KM, Didan K (2009) Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology* doi: 10.1111/j.1365-2486.2009.01910.x. Q12

White MA, Nemani RR (2004) Soil water forecasting in the continental United States: relative forcing by meteorology versus leaf area index and the effects of meteorological forecast errors. *Canadian Journal of Remote Sensing*, **30**, 717–730.

Woodward FI (1987) *Climate and Plant Distribution*. Cambridge University Press, Cambridge, UK.

Yang F, Ichii K, White MA *et al*. (2007) Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach. *Remote Sensing of Environment*, **110**, 109–122.

Zhao M, 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.

Zheng D, Prince S, Wright R (2003) Terrestrial net primary production estimates for 0.5° grid cells from field observations – a contribution to global biogeochemical modeling. *Global Change Biology*, **9**, 46–64. Q13

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Annual maximum LAI (unitless) of forests in the T-P domain: (a) Biome-BGC; (b) LPJ; and (c) AVHRR (input for TOPS-BGC and CASA).

Figure S2. Same as Fig. SA1 but for non-forests or vegetation simulated over non-forest regions: (a) Biome-BGC; (b) LPJ; and (c) AVHRR (input for TOPS-BGC and CASA).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

UNCORRECTED PROOF

(BWUK GCB 2309 Webpdf:=08/23/2010 04:05:12 2583606 Bytes 17 PAGES n operator=S.arsathali) 8/23/2010 4:07:08 PM

© 2010 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2010.02309.x