Atmospheric constraints on gross primary productivity and net ecosystem productivity: Results from a carbon-cycle data assimilation system

E. N. Koffi,[1] P. J. Rayner,[2] M. Scholze,[3] and C. Beer[4]

Received 23 June 2010; revised 8 November 2011; accepted 23 January 2012; published 17 March 2012.

[1] This paper combines an atmospheric transport model and a terrestrial ecosystem model to estimate gross primary productivity (GPP) and net ecosystem productivity (NEP) of the land biosphere. Using atmospheric CO_2 observations in a Carbon Cycle Data Assimilation System (CCDAS) we estimate a terrestrial global GPP of 146 $\pm$ 19 GtC/yr. However, the current observing network cannot distinguish this best estimate from a different assimilation experiment yielding a terrestrial global GPP of 117 GtC/yr. Spatial estimates of GPP agree with data-driven estimates in the extratropics but are overestimated in the poorly observed tropics. The uncertainty analysis of previous studies was extended by using two atmospheric transport models and different CO_2 observing networks. We find that estimates of GPP and NEP are less sensitive to these choices than the form of the prior probability for model parameters. NEP is also found to be significantly sensitive to the transport model and this sensitivity is not greatly reduced compared to direct atmospheric transport inversions, which optimize NEP directly.

Citation: Koffi, E. N., P. J. Rayner, M. Scholze, and C. Beer (2012), Atmospheric constraints on gross primary productivity and net ecosystem productivity: Results from a carbon-cycle data assimilation system, *Global Biogeochem. Cycles*, *26*, GB1024, doi:10.1029/2010GB003900.

1. Introduction

[2] Concern about the current and future behavior of the terrestrial carbon cycle has stimulated the research community to build complex observing systems. These have focused on the net carbon flux since this plays the most direct role in regulating atmospheric CO_2. However, for a range of questions about the carbon cycle, we also need to know about the productivity of the terrestrial biosphere. The gross primary productivity (GPP) is the total amount of CO_2 that is taken up by plants by photosynthesis [e.g., *Farquhar et al.*, 1980]. The net primary productivity (NPP) is the difference between GPP and total plant respiration [e.g., *Lieth*, 1973; *Clark et al.*, 2001a]. NPP is classically regarded as equal to the positive increment in biomass as a result of growth. The net ecosystem production (NEP), which equals the net land-atmosphere CO_2 flux of the undisturbed biosphere, is NPP minus the flux from respiration by heterotrophs and decomposers [e.g., *Reichstein et al.*, 2005]. In contrast to the GPP, the net primary productivity is well known since it is linked to consumption of harvested biomass [e.g., *Malmström et al.*, 1997; *Prince et al.*, 2001; *Imhoff et al.*, 2004; *Imhoff and Bounoua*, 2006]. Moreover, the impact of CO_2 concentration on GPP is subtle [e.g., *Nowak et al.*, 2004; *McMurtrie et al.*, 2008]. It seems worthwhile, therefore, to ask what we can learn from current observations of atmospheric CO_2 about terrestrial productivity.

[3] As soon as we seek to estimate primary productivity we face a different measurement problem than for the net flux. Because of its commercial importance there are relatively direct measures of NPP (e.g., crop and forest inventory data). These, however, are usually measures of above ground biomass increment or harvested biomass, which does not represent all the components of NPP [e.g., *Clark et al.*, 2001b]. Moreover, these measures are not usually simulated by terrestrial ecosystem models. Furthermore, although they are spatially much more extensive than the pointwise net flux derived from eddy covariance measurements [e.g., *Foken and Wichura*, 1996; *Aubinet et al.*, 2000; *Baldocchi*, 2003; *Rebmann et al.*, 2005; *Reichstein et al.*, 2005; *Papale et al.*, 2006; *Lasslop et al.*, 2010], they still under-sample the variability of the unmanaged biosphere. *Clark et al.* [2001b] showed various biases with the inventory method, the biggest one being the problem of measuring fine root turnover.

[4] The large amount of NEP data at flux tower stations measured by the eddy covariance technique has been exploited in concert with satellite and climate data to derive terrestrial GPP [*Beer et al.*, 2010; *Yuan et al.*, 2010]. The NEP data at flux tower sites can be partitioned into GPP and ecosystem respiration by using several methods mainly based on flux partitioning algorithms [e.g., *Reichstein et al.*, 2005; *Lasslop et al.*, 2010]. Empirical relationships between GPP

[1]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France.
[2]School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia.
[3]Department of Earth Sciences, University of Bristol, Bristol, UK.
[4]Biogeochemical Model-Data Integration Group, Max Planck Institute for Biogeochemistry, Jena, Germany.

and explanatory variables such as the fraction of absorbed photosynthetically active radiation, short-wave radiation, precipitation, air temperature, and vapor pressure deficit at the tower sites can then be applied to large-scale grids of these explanatory variables. Since the eddy covariance system also measures latent heat, evapotranspiration can also be an explanatory variable [*Beer et al.*, 2007, 2009] which is at large scale also constrained by runoff data. Such a data-driven approach has the advantage of high spatial (up to 1 km) and temporal resolution (up to 10 days), but this approach has no predictive capability.

[5] In contrast, a process-oriented mechanistic model of photosynthesis at leaf level [*Farquhar et al.*, 1980] in conjunction with assumptions about autotrophic respiration and a model of heterotrophic respiration can be used to estimate large-scale GPP and NEP [*Knorr*, 2000; *Sitch et al.*, 2003; *Krinner et al.*, 2005; *Kattge et al.*, 2009; *Oleson et al.*, 2010]. Such a model requires atmospheric and soil conditions as inputs. The parameters in such models can be constrained by assimilation techniques as employed in the Carbon Cycle Data Assimilation System (CCDAS) [*Rayner et al.*, 2005]. We use this approach in this study.

[6] A data assimilation system has two primary components:

[7] i) a deterministic dynamical model that calculates the evolution of a set of state variables given an initial condition, forcing and a set of process parameters of the model and a series of observation operators that map the state variables of the model onto quantities amenable to observation.

[8] ii) an assimilation system that consists of an algorithm to adjust a subset of the state variables, initial conditions and/or process parameters to reduce the mismatch with observations. Usually any prior information on the variables which are adjusted are also taken into account (see *Kaminski et al.* [2002, 2003] and *Rayner et al.* [2005, and references therein] for the underlying methodology).

[9] Following *Kaminski et al.* [2003] and *Rayner et al.* [2005] our main observable in this work is atmospheric CO_2 concentration. This data set has also been used for estimates of surface flux without the use of dynamical constraints in an approach usually called direct inversion [e.g., *Gurney et al.*, 2002, 2004; *Rayner et al.*, 2008]. In both the direct inversion and the CCDAS approach, net fluxes are coupled to atmospheric CO_2 concentrations by atmospheric transport. Hence their estimates are sensitive to uncertainties in transport models and the choice of network. These uncertainties have been obtained for direct inversions [*Gurney et al.*, 2002; *Law et al.*, 2003; *Gurney et al.*, 2004; *Baker et al.*, 2006]. Additionally, almost all aspects of the net flux pattern, its mean spatial structure, seasonal cycle and interannual variability were found to be sensitive to transport models and networks used in the TransCom exercises [e.g., *Gurney et al.*, 2003, 2004; *Baker et al.*, 2006]. Given the direct links between various model parameters and these large-scale characteristics it seems likely that inferred process parameters of underlying models of a CCDAS will be similarly sensitive, but this has never been tested. This presupposes that the atmospheric CO_2 concentrations contain information on the process parameters of the model. Thus the aim of the paper is twofold: 1) to extend the CCDAS to estimate GPP and 2) to carry out a more detailed analysis of the sensitivity of both NEP and GPP to the configuration of the CCDAS (i.e., transport model, CO_2 observing network, the form of the probability densities on the prior parameters of the underlying biosphere model).

[10] The outline of the paper is as follows:

[11] First, we briefly describe the main principles of the CCDAS (section 2). In section 3, the set-up of CCDAS as used here is described. We discuss the sensitivity of the process parameters of CCDAS to i) two representations of the tracer transport model TM (TM2 [*Heimann*, 1995] and TM3 [*Heimann and Körner*, 2003]) and ii) two observing networks (38 and 68 sites). In addition, we discuss results from two probability distributions for the prior parameters in section 4. The inferred GPPs are presented in section 5 together with the comparison of the regional integrals of our best GPP estimates to the few existing estimates at global and ecosystem scales. Then, the sensitivities of both the inferred NEP and GPP to the configuration of the CCDAS (i.e., transport model, CO_2 observing network, the form of the probability densities on the prior parameters of the underlying biosphere model) are given in section 6. In section 7, results are discussed. Finally, conclusions and remarks for a possible improvement of the system are given in section 8.

2. Methodology

[12] The CCDAS in this study has two primary components: a biosphere model and a transport model. The process parameters of the biosphere model are first optimized by reducing the mismatch between observed CO_2 concentrations and simulations by the transport model. The modeled concentrations are obtained by using the modeled NEP fluxes together with the other components of the CO_2 exchange fluxes with the atmosphere (i.e., ocean flux, land use, and fossil fuel emissions). In this section and in the next, we describe the CCDAS together with all the pieces that compose the system and how these elements are combined to fulfill the objectives of the paper.

2.1. CCDAS System

[13] The methodological detail of the CCDAS is described by *Rayner et al.* [2005] and *Scholze et al.* [2007]. Here we reprise only the generalities of the system and the points where we differ.

[14] CCDAS consists of a biosphere model BETHY (Biosphere Energy Transfer Hydrology) [*Knorr*, 2000] and a versatile atmospheric transport model, together with CO_2 fluxes representing ocean flux, land use change, and fossil fuel emission [*Rayner et al.*, 2005; *Scholze et al.*, 2007]. The process parameters of the biosphere model are optimized by using a Bayesian inference scheme [*Enting*, 2002; *Tarantola*, 2005]. This inference scheme requires the minimization of a cost function representing the negative log likelihood. The cost function includes contributions from the model-observation mismatch and the departure of parameter values from their prior estimates.

[15] The cost function hence includes the uncertainties in both the transport model and observations and the prior parameters in the form of covariance matrices. Parameters are uncorrelated in our formulation. The cost function is minimized using the limited memory Quasi-Newton algorithm M1QN3 V3.2 [*Gilbert and Lemaréchal*, 1989]. The required derivatives are generated automatically from the model code

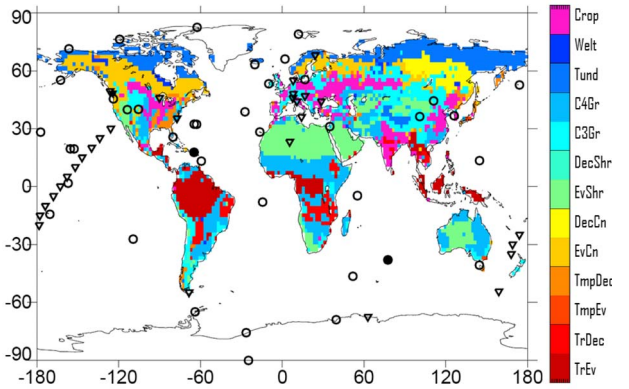

Figure 1. Distribution of the dominant PFT per grid. The labels of the PFTs are: Crop: Crop plant, Welt: Swamp vegetation, Tund: Tundra, C4Gr: C4 grass, C3Gr: C3 grass, DecShr: Deciduous shrub, EvShr: Evergreen shrub, DecCn: Deciduous coniferous tree, EvCn: Evergreen coniferous tree, TmpDec: Temperate broadleaved deciduous tree, TmpEv: Temperate broadleaved evergreen tree, TrDec: Tropical broadleaved deciduous tree, TrEv: Tropical broadleaved evergreen tree. Locations of the 70 CO_2 observational sites used in this study are shown: TM3/38 configuration uses 38 stations (open circles). TM3/68 uses both the 38 stations of TM3/38 plus the 30 stations marked in triangles (both open and filled). TM2/41 uses the 38 stations of TM3/68 and three additional stations (2 filled circles and 1 filled triangle).

using the Transformation of Algorithms in Fortran (TAF) tool [*Giering and Kaminski*, 1998; *Kaminski et al.*, 2003].

2.2. Models and Data

2.2.1. Terrestrial Carbon Cycle Model BETHY

[16] BETHY, the core model of CCDAS, is a process-based model of the terrestrial biosphere which simulates carbon assimilation and plant and soil respiration, embedded within a full energy and water balance [*Knorr*, 2000]. BETHY is structured in four components: (1) energy and water balance, (2) photosynthesis, (3) phenology, and (4) carbon balance. It uses 13 plant functional types (PFT) based on the work of *Wilson and Henderson-Sellers* [1985] (Figure 1) and up to three PFTs can be present within a grid cell with their amount specified by their fractional coverage. The version of BETHY used here is implemented on a $2^\circ \times 2^\circ$ grid box, with 3462 land grid points. BETHY is driven by observed climate and radiation data over the period 1979–1999 [*Nijssen et al.*, 2001].

[17] BETHY is used here in two forms: a full form and a simplified one. In its full form, the model is used to optimize the parameters related to water balance, temperature-limited phenology, and overall fractional vegetation cover against remotely sensed vegetation greenness [*Knorr and Schulz*, 2001]. The optimized parameters from the full version of BETHY are used in the simplified form of BETHY where they are not optimized anymore (see section 2.2.4 for details). A complete description of the simplified version of BETHY used in this study is given by *Rayner et al.* [2005]. Here we review the parts of the formulation on GPP, NPP, and NEP which we need later.

[18] For GPP, we follow the parameterizations of *Farquhar et al.* [1980] and *Collatz et al.* [1992] for C3 and C4 plants, respectively. GPP is calculated as the minimum of an electron transport limited rate, J_E, and a rate, J_C, limited by the carboxylation enzyme Rubisco. The symbols in this section related to the BETHY model are also defined in the notation section.

$$GPP = \begin{cases} \min[J_C, J_E] & \text{for C3 plant} \\ \min[J_e, J_c, J_i] & \text{for C4 plant} \end{cases} \quad (1)$$

where

$$J_C = V_{\max}\left[\frac{C_i - \Gamma^*}{C_i + K_C\left(1 + {}^{O_x}/_{Ko}\right)}\right] \quad (2)$$

$$J_E = \left[\frac{\alpha_q I J_{\max}}{\sqrt{J^2{}_{\max} + \alpha^2{}_q I^2}}\right]\left[\frac{C_i - \Gamma^*}{4(C_i + 2\Gamma^*)}\right] \quad (3)$$

$$\begin{cases} J_e = V_{\max} \\ J_c = kC_i \\ J_i = \alpha_i I \end{cases} \quad (4)$$

With $V_{\max}$ the carboxylation capacity of the leaf, C_i the leaf CO_2 concentration (note that the concentration of the atmospheric CO_2 concentration is kept constant at 355 ppm) minus the net CO_2 assimilated by the leaf, Γ^* the CO_2 compensation point, K_C and Ko the Michaelis-Menten constants for CO_2 and O_2, respectively, O_x the O_2 partial pressure, α_q the quantum efficiency, $J_{\max}$ the maximum electron transport, I the photosynthetically active radiation (PAR) absorption rate, α_i the integrated C4 quantum efficiency, k the PEP case (the initial CO_2 fixating enzyme in C4) CO_2 specificity.

[19] Following *Farquhar* [1988], Γ^* is linearly dependent on vegetation temperature with $a_{\Gamma,T}$ being the dependence parameter. k is linearly dependent on $V_{\max}$ at 25°C with $a_{J,V}$ the linear dependence parameter.

[20] The temperature dependence of $V_{\max}$, K_C and Ko, k are computed based on their respective quantity at 25°C (i.e., $V^{25}_{\max}$, K^{25}_C, K^{25}_o, k^{25}) and their corresponding activation energy E [*Rayner et al.*, 2005].

[21] Two kinds of parameters involved in the computation of GPP are optimized (Table 1): 1) vegetation type dependent parameters ($V^{25}_{\max}$ and $a_{J,V}$) and 2) global parameters (α_q, α_i, $a_{\Gamma,T}$, K^{25}_C, K^{25}_o, $E_{V\max}$, E_{KC}, E_{Ko}, E_k). The exponent 25 stands for the value of the parameter at 25°C. Altogether, 35 parameters relevant to GPP are optimized (Table 1).

[22] The net primary productivity NPP is computed as a gross uptake of CO_2 by the leaves (GPP) minus total autotrophic respiration which includes plant maintenance respiration, R_M, and growth respiration R_G. R_G is parameterized as a function of NPP, while R_M is a function of the leaf dark respiration R_d [*Knorr*, 2000]:

$$R_M = R_d / f_{R,leaf} \quad (5)$$

$$R_G = \left(f_{R,growth} - 1\right)NPP = \left(f_{R,growth} - 1\right)\left(GPP - R_M - R_G\right) \quad (6)$$

Table 1. Controlling Parameters and Their Initial and Optimized Values Obtained From 6 Configurations[a]

Parameter	Prior Value	Prior Uncertainty	TM2/41	TM3/38	TM3/68	TM2/41b	TM3/38b	TM3/68b
V_{max}(TrEv)	60	12	54.1	58.7	58.3	60.6	61.1	63.8
V_{max}(TrDec)	90	18	53.3	48.2	59.5	81.1	78.4	73.5
V_{max}(TmpEv)	41	8.2	45	43.4	44.5	45.4	38.8	39.7
V_{max}(TmpDec)	35	7	95.1	83.3	102.4	124.2	128.4	149.2
V_{max}(EvCn)	29	5.8	30.8	21.2	10.7	36.1	29.1	21.9
V_{max}(DecCn)	53	10.6	122.6	100.7	98.8	184.7	148.1	136.4
V_{max}(EvShr)	52	10.4	101.8	106.7	84.4	162.5	166.3	168.9
V_{max}(DecShr)	160	32	84	75.5	70.9	93.3	107.5	96.1
V_{max}(C3Gr)	42	8.4	25.4	20.1	22.2	21.8	19.2	18.9
V_{max}(C4Gr)	8	1.6	−0.1	−0	−0.1	0.7	0.7	0.7
V_{max}(Tund)	20	4	22.2	5.3	9	24.8	8.2	8.5
V_{max}(Wetl)	20	4	−8.2	−11.7	−7.7	9.6	9.6	9.3
V_{max}(Crop)	117	23.4	24	25.7	19.2	46.8	47.4	47.9
$a_{J,V}$(TrEv)	1.96	0.098	1.98	1.86	1.8	2.02	1.97	1.97
$a_{J,V}$(TrDec)	1.99	0.0995	2.03	1.87	1.96	2.06	2.01	2.02
$a_{J,V}$(TempEv)	2	0.1	2	2	2.01	2.01	2	2
$a_{J,V}$(TmpDec)	2	0.1	1.98	2.17	2.16	1.84	2.05	2.03
$a_{J,V}$(EvCn)	1.79	0.0895	1.87	1.8	1.8	1.92	1.84	1.83
$a_{J,V}$(DecCn)	1.79	0.0895	1.86	1.88	1.88	1.92	1.84	1.83
$a_{J,V}$(EvShr)	1.96	0.098	1.96	1.98	1.95	1.96	2	1.99
$a_{J,V}$(DecShr)	1.66	0.083	1.69	1.67	1.76	1.67	1.67	1.66
$a_{J,V}$(C3Gr)	1.9	0.095	1.9	1.94	1.92	1.89	1.9	1.9
$a_{J,V}$(C4Gr)	140	28	127	135	138	133	137	130
$a_{J,V}$(Tund)	1.85	0.0925	1.86	1.85	1.86	1.87	1.84	1.84
$a_{J,V}$(Wetl)	1.85	0.0925	1.84	1.85	1.85	1.83	1.82	1.83
$a_{J,V}$(Crop)	1.88	0.094	1.85	1.88	1.87	1.87	1.88	1.88
$f_{R.leaf}$	0.4	0.1	0.3	0.2	0.2	0.2	0.3	0.3
$f_{R.growth}$	1.25	0.0625	0.56	0.84	0.71	1.11	1.12	1.11
Q_{10f}	1.5	1.5	1.38	1.44	1.84	1.37	1.3	1.32
Q_{10s}	1.5	1.5	3.61	2.24	0.72	3.09	3.11	3
τ_f	1.5	3	8.72	9.01	20.88	8.52	7.6	7.57
κ	1	10	0.83	0.49	0.83	0.78	0.5	0.54
f_s	0.2	2	0.44	0.44	0.29	0.53	0.57	0.56
E_{Rd}	45000	2250	33648	42984	34430	32518	39272	39829
E_{Vmax}	58520	2926	82137	83197	82299	62363	70479	67212
E_{KO}	35948	1797.4	35843	35655	35403	37383	39183	39433
E_{KC}	59356	2967.8	77266	66258	70869	54559	53222	53901
E_k	50967	2548.35	50833	50890	50943	50804	50803	50760
α_q	0.28	0.014	0.44	0.38	0.43	0.56	0.48	0.51
α_{i25}	0.04	0.002	0.04	0.04	0.04	0.04	0.04	0.04
K_{C25}	460	23	413	373	369	460	474	478
K_O	330	16.5	342.4	353	352.9	326	319.8	320.6
$a_{\Gamma,T}$	1.7	0.085	0.81	1.22	1.08	0.73	1.01	0.94
β(TrEv)	1	0.25	0.88	0.82	0.79	0.93	0.86	0.88
β(TrDec)	1	0.25	0.7	1.06	1.01	1.01	1.2	1.15
β(TmpEv)	1	0.25	0.02	0.68	0.13	0.57	1.04	1.02
β(TmpDec)	1	0.25	0.64	0.67	0.83	0.83	0.75	0.79
β(EvCn)	1	0.25	1.34	1.14	0.91	1.29	1.17	1.12
β(DecCn)	1	0.25	0.15	0.87	1.82	0.35	0.75	0.84
β (EvShr)	1	0.25	0.77	0.43	0.02	0.56	0.65	0.58
β(DecShr)	1	0.25	2.4	1.94	2.84	6.47	3.23	2.66
β (C3Gr)	1	0.25	1	0.88	0.7	0.82	0.73	0.82
β (C4Gr)	1	0.25	0.98	0.97	0.72	0.52	0.98	1.01
β (Tund)	1	0.25	0.92	0.63	0.66	0.9	0.79	1.03
β (Wetl)	1	0.25	0.39	0.64	0.48	4.05	1.31	1.42
β (Crop)	1	0.25	0.59	1.33	1.46	0.57	1.16	1.14
Offset	338	1	336	337	338	336	337	337

[a]TM2/41 denotes optimization by using TM2 data and 41 stations. TM3/38 (TM3/68) denotes optimization by using TM3 model and 38 (68) sites. Units are: V_{max}, μmol$(CO_2)m^{-2}$ s^{-1}; $a_{J,T}$ activation parameter; $a_{\Gamma,T}$ μmol(CO_2)mol(air)$^{-1}(C)^{-1}$; activation energies E, J/mol; τ_f, years; offset, ppm; all other parameters are unitless and correspond to values at 25°C. K_C is multiplied by 10^6. TM2/41, TM3/38, and TM3/68 stand for optimizations by using Gaussian PDF on prior parameters, while TM2/41b, TM3/38b, and TM3/68b stand for lognormal PDF. When considering the lognormal PDF, we use a Gaussian PDF for 15 parameters: $a_{\Gamma,T}$, $a_{J,V}$ (PFT dependent thus 13 parameters), and the offset (i.e., the atmospheric CO_2 at the first year of the study). Note that for $a_{\Gamma,T}$, $a_{J,V}$, and the offset, there is not any reason to use lognormal PDF since these parameters can have negative values. The acronyms relevant for PFT are described in Figure 1. For other details on the definition of the parameters, see the notation section.

where $f_{R,leaf}$ is the leaf fraction of the maintenance respiration and $f_{R,growth}$ stands for the amount of carbon respired for a unit gain in vegetation biomass. They are both treated as parameters. Relationships between R_d and V_{max} at 25°C and their dependence on both temperature and its activation energy E_{Rd} are detailed by *Rayner et al.* [2005]. The parameters $f_{R,leaf}$, $f_{R,growth}$ and E_{Rd} are optimized (Table 1).

[23] The net ecosystem productivity *NEP* which quantifies the net CO_2 flux between the atmosphere and the biosphere is defined as the net primary productivity *NPP* minus

heterotrophic respiration (R_S), which is composed of the respiration from short-lived (fast) litter pool $R_{S,f}$ and a long-lived (slow) soil carbon pool $R_{S,s}$

$$NEP = NPP - R_S = NPP - R_{S,f} - R_{S,s} \tag{7}$$

where

$$\begin{cases} R_{S,f} = C_f(1-f_s)\left(\omega^{\kappa} Q_{10f}{}^{Ta/10}\right)/\tau_f \\ R_{S,s} = C_s\left(\omega^{\kappa} Q_{10s}{}^{Ta/10}\right)/\tau_s \end{cases} \tag{8}$$

with f_s, the fraction of decomposition from the fast pool that goes to the long-lived soil carbon pool. C_f and C_s stand for the sizes of the short-lived and the slow litter pools, respectively. Other variables are: ω the soil moisture, T_a the air temperature, κ the soil moisture dependence parameter, Q_{10f} and Q_{10s} temperature dependence parameters relative for fast and slow pools, respectively. τ_f and τ_s the pool turnover times at 25°C relative to fast and slow pools, respectively. The parameters f_s, Q_{10f}, Q_{10s}, κ, and τ_f are optimized (Table 1).

[24] Following *Rayner et al.* [2005] we formulate NEP as a PFT-dependent storage efficiency multiplied by NPP (which varies by grid point). By further assuming that the size of the slow pool does not change over the 21-year study period we can diagnose the mean respiration from the slow pool:

$$\overline{NEP} = \overline{NPP}\left(1 - \frac{1}{\beta}\right) \tag{9}$$

The overlying bar denotes the time averages of the quantity in question over the simulation period at each grid cell. The scaling factor β (taken as a parameter) is vegetation type dependent with $\beta = 1$ (first guess) represents a carbon neutral terrestrial biosphere, $\beta > 1$ characterizes a net uptake of CO_2 by vegetation and $0 < \beta < 1$, a net source of CO_2 from the vegetation into the atmosphere. The parameter β which is PFT dependent is optimized. Altogether, 18 parameters related to NEP are optimized (Table 1).

[25] In total we optimize 56 biospheric parameters relevant for GPP, NPP, and NEP plus one initial condition representing the atmospheric CO_2 at the first year of the study (Table 1).

2.2.2. Calculation of Uncertainty in BETHY Fluxes

[26] *Rayner et al.* [2005] demonstrated how to project the parameters' uncertainties forward through the computed carbon fluxes from BETHY. The uncertainty $C(y)$ in the simulated quantity of interest $y(x)$ (i.e., *GPP, NPP, NEP*) is approximated to a first order by:

$$C(y) = \frac{dy(x_{opt})}{dx}\,\frac{1}{H}\left(\frac{dy(x_{opt})}{dx}\right)^{T} \tag{10}$$

H (the Hessian matrix) is the second derivative of the cost function with respect to the parameters x at value x_{opt}. The symbol ^{T} denotes the transposed matrix. The inverse of H gives the uncertainty in x. We consider both the observations and the prior of the parameters x to have a Gaussian Probability Density Function (PDF), which then also results in a Gaussian PDF for the optimized values of x [*Tarantola*, 1987]. The posterior uncertainty is expressed by:

$$H = \frac{d^2 J(x_{opt})}{dx^2} \tag{11}$$

With $J(x_{opt})$ the cost function computed for the optimized values of the parameter x_{opt}.

2.2.3. Transport Models

[27] For the tracer transport we use pre-computed transport Jacobians, i. e. sensitivities of concentrations to fluxes at every pixel. We use the Jacobians of *Kaminski et al.* [1999] for TM2 and *Rödenbeck et al.* [2003] for TM3. The key difference between the models is the spatial resolution. TM2 has a resolution of 8 degrees latitude by 10 degrees longitude with 9 levels while TM3 uses 4 degrees of latitude by 5 degrees of longitude with 19 levels. TM2 and TM3 use ECMWF (European Centre for Medium-Range Weather Forecasts) and NCEP (National Centers for Environmental Prediction) meteorological fields as input, respectively.

[28] The Jacobians were computed over the 1979–1999 period by using only one year of winds for TM2, while TM3 used meteorological forcing that varied each year. The base models are described by *Heimann* [1995] for TM2 and *Heimann and Körner* [2003] for TM3. Both versions TM2 and TM3 have participated in TransCom, a model intercomparison experiment based on fossil fuel CO_2 emissions and annually balanced CO_2 fluxes [*Law et al.*, 1996; *Gurney et al.*, 2003].

[29] Since TM2 or TM3 allow only Jacobians at particular stations, we have used the global climate model LMD_Z-INCA [e.g., *Hauglustaine et al.*, 2004] to compute CO_2 concentrations at any other stations. The model has 19 levels and a horizontal resolution of 2.5° in latitude and 3.75° in longitude. LMD_Z is an online model i.e., it generates its dynamics internally along with tracer transport. To ensure realistic simulation of actual meteorological conditions the model is nudged toward ECMWF reanalyses.

2.2.4. Background Fluxes

[30] One problem in drawing any inferences about the terrestrial biosphere from atmospheric observations is the contribution of other CO_2 fluxes. These sources are fossil fuel emissions, ocean fluxes, and fluxes related to land use change. They are not perfectly known, and they impact the concentration of CO_2. Since we use the same type of data as in the work by *Scholze et al.* [2007], only the main patterns of these background fluxes are described here. For the fossil fuel fluxes, we use the flux magnitudes from *Marland et al.* [2001] for the years 1979 to 1995 and a constant magnitude of 6.5 GtC/yr for the years 1996–1999. We use two different patterns distributing the fossil emissions corresponding to the years 1990 and 1995. Fields are taken from *Andres et al.* [1996] for 1990 and *Brenkert* [1998] for 1995. For the years prior to 1991, we use the 1990 pattern solely and for the years following 1994 we use the pattern for 1995. For the years 1991 to 1994, we linearly interpolate between the two spatial patterns. The shortcoming of these data, as we use them, may be the lack of seasonality in the flux. The ocean fluxes are taken from two sources: i) The flux pattern and magnitude from *Takahashi et al.* [1999] are used to describe the flux climatology (both annual mean and seasonal cycle)

and ii) an estimate of inter-annual variability in ocean flux taken from the study of *Le Quéré et al.* [2003] for years 1980 to 1998 is then added. The flux caused by land-use change is taken from the study of *Houghton et al.* [1987] that estimates the annual mean flux. Following *Scholze et al.* [2007], we divide the total annual values in 12 to have monthly values as multipliers of the annual spatial pattern [*Houghton*, 2008].

2.2.5. Other Data

[31] As mentioned earlier, BETHY is optimized in two stages with two different data sets. First, to optimize the parameters related to water balance, temperature-limited phenology, and overall fractional vegetation, remotely sensed fAPAR (fraction of Absorbed Photosynthetically Active Radiation) data are used [*Knorr and Schulz*, 2001]. The authors estimated the error in fAPAR to lie between 0.05 and 0.1. This step provides leaf area index (LAI) and plant available soil moisture ω (as a fraction of maximum soil water capacity), which are prescribed in the simplified form of the model to assimilate atmospheric CO_2 concentration observations. Thus, LAI and ω parameters are not optimized with respect to atmospheric CO_2 concentrations.

[32] Second, to optimize the parameters related to BETHY carbon fluxes, the monthly mean atmospheric CO_2 concentration data from the GLOBAVIEW database [*GLOBALVIEW-CO2*, 2004, hereinafter GV] and some additional CO_2 measurement sites are considered (Figure 1). The processing of the GLOBALVIEW data for 41 sites used for the TM2 model is described by *Rayner et al.* [2005], while the treatment of these data for the 68 sites for TM3 is given by *Rödenbeck et al.* [2003]. For the 27 additional sites allowed for the TM3 model, we do not have any precomputed Jacobians for TM2. Consequently, for the study of the sensitivity of our results to transport model, only a common subset of the sites allowed by both TM2 and TM3 models are considered, as described in section 3. The treatment of uncertainties for CO_2 concentrations follows that of *Rayner et al.* [2005]. The GLOBALVIEW-CO_2 uncertainties in the monthly means are derived from the residual standard deviations [*Masarie and Tans*, 1995]. Following *Rayner et al.* [2005], we add a minimum uncertainty of 0.5 ppm to account for additional transport model error. Our data uncertainties range from 0.51 ppm to 4.9 ppm.

2.2.6. Prior Values and Uncertainties

[33] According to the general Bayesian methodology we incorporate prior information on model parameters (Table 1). For physical parameters the values are taken from literature summarized by *Knorr* [2000]. For other values such as the beta storage efficiency (β), we commence with reasonable values, which do not take account of previous atmospheric studies (since it is important to avoid double-counting information). Thus we start with the assumption of a balanced biosphere ($\beta = 1$) with uncertainties in β generally assumed to be large since there are little data on which to base these (Table 1). Finally prior information not only includes results of previous studies but also knowledge of the physical limits of parameters. For example many parameters are physically limited to positive values. Such a bounding is not implemented in the optimization algorithm M1QN3 we use. For these we implement an exponential transform so that, while the parameter follows a Gaussian probability distribution with mean μ and standard deviation σ the model transforms this as $\exp(\mu \pm \sigma)$, that is the physical parameter follows a lognormal distribution. We implement this for 42 of the 56 biosphere parameters (Table 1).

3. Experimental Setup

[34] Three configurations are built to optimize the parameters of BETHY: TM2/41: TM2 Jacobians with observations at 41 sites from GV network. TM3/38: TM3 Jacobians with 38 sites used in both TM2 and TM3. The CO_2 concentration data for TM2 and TM3 are computed slightly differently. Thus, we repeated TM3/38 by using an exact subset of TM2/41 observations (hereafter TM3/38*) and produced very similar results to TM3/38. When quantifying the root mean square deviation both in space and time (rsmd) between the optimized NEPs from these three above mentioned configurations, results show smallest rmsd between TM3/38 and TM3/38*. Therefore, the optimized parameters of TM3/38* will not be considered in the following sections. TM3/68: TM3 Jacobians with observations at 68 sites, which is considered as the preferred formulation for the optimization.

[35] The locations of the sites under study are shown in Figure 1. We append "b" to the experiment name when we consider the lognormal transform discussed previously.

4. Optimized Parameters

[36] The overall quality of the fit to data is embodied in the value of the cost function at the minimum. The final values of the cost function obtained after about 2200 iterations (i.e., number of iterations from which the value of the cost function does not change much) of each of the 6 configurations of model/data under study are considered: These configurations are TM2/41, TM3/38, and TM3/68 relevant for optimizations by using a Gaussian PDF on prior parameters and TM2/41b, TM3/38b, and TM3/68b when using the lognormal PDF (Table 1). The reduced chi square (χ^2) which describes the average misfit of optimized simulations and observations plus the departures is computed for each configuration of the model/data. It should be around 1 for an optimal system while our optimizations yield values around 2. Stations more subject to background fluxes (e.g., oceanic stations) contribute disproportionately to this value suggesting some of the problem may lie with the background fluxes rather than the biosphere model. We do not adjust observational uncertainties to account for this mismatch but the result should be borne in mind when considering posterior uncertainties.

[37] Table 1 shows the optimized parameters of BETHY for the six configurations of model/data. All optimizations yield large shifts from the prior parameters for i) the photosynthesis parameter V_{max} of some PFTs (DecCn, EvShr, C4Gr, Wetl, Crop; see Figure 1 for location and definition of the acronyms), ii) some global parameters related to photosynthesis (i.e., E_{Vmax}, α_q, $a_{\Gamma,T}$; referred to section 2.2.1. for the definition of these parameters). When using the Gaussian distribution for the prior parameters, V_{max} is found to be more sensitive to both transport model and observing network for the PFTs C4Gr, Crop, Wetl, DecCn, TmpDec. In detail, V_{max} for TmpDec, EvShr, and EvCn PFTs are more sensitive to observing network, while V_{max} for Tund and DecCn PFTs are more sensitive to transport model (Table 1). Overall, the use of a lognormal PDF on the prior parameters tends to increase the shifts from the prior parameters,

Table 2. Mean Biosphere Fluxes (GtC/yr) Inferred From the BETHY Model Using Prior and Optimized Parameters Over the Time Period 1980–1999[a]

Region	Prior	TM2/41	TM3/38	TM3/68	TM2/41b	TM3/38b	TM3/68b	K09	B10
				NEP					
Global	0.5	2.08	2.14	2.19	2.07	2.13	2.13		
NH	−0.01	0.75	0.85	0.73	0.83	1.01	1.04		
Tropics	0.5	1.48	1.27	1.45	2.46	1.54	1.37		
				GPP					
Global	145	127	109	117	164	144	146	163	118
NH	38	36	27	27	45	36	36	46	34
Tropics	101	88	79	86	114	104	106	112	81

[a]GPP denotes Gross Primary Productivity and NEP, Net Ecosystem Productivity. Results from six optimized data configurations are shown: TM2/41 (TM2 model with observations from 41 Global View sites), TM3/38 (TM3 model with 38 sites used in TM2/41), and TM3/68 (TM3 model with 68 sites). TM2/41, TM3/38, and TM3/68 stand for optimizations by using Gaussian PDF on prior parameters, while TM2/41b, TM3/38b, and TM3/68b stand for lognormal PDF. Estimates of GPP from *Kattge et al.* [2009] (K09) and *Beer et al.* [2010] (B10) are also given. Global, northern extra-tropical [NH: 30°N–90°N], and tropical [30°S–30°N] fluxes are calculated for all these configurations.

especially for the global parameters α_q, fs, Q_{10s}, τ_f, $a_{\Gamma,T}$, and V_{max} for most of the PFTs (Table 1). In addition, the use of lognormal PDFs on the prior parameters reduces the sensitivity of the parameters to the observing network. Among the optimized parameters, the choice of prior PDF has more influence on the final parameter values than either network or transport model (Table 1): the differences between optimized parameters for the same PDF are smaller than the differences obtained between the same model/data set up with different PDFs. As an example, the difference in optimized V_{max} for the TmpDec PFT between TM2/41 and TM3/38 is about four times the one between TM3/38 and TM3/38b (Table 1). Thus, optimized parameters are more sensitive to the choice of prior PDF for parameters than either network or transport model.

[38] Some optimized parameters show values which are not physically reasonable when optimizations are performed with Gaussian PDFs for prior parameters. The striking examples are the photosynthesis parameters V_{max} for C4 grass and wetland vegetation, respectively, showing negative optimized values regardless of the configuration (Table 1). Such a V_{max} gives a negative GPP for this PFT (see equations (1) and (2)). For these V_{max}, the use of lognormal PDFs gives also a large reduction (Table 1). To a lesser extent, there is the relatively small V_{max} for crops in all the optimizations. Indeed, such a small V_{max} suggests low

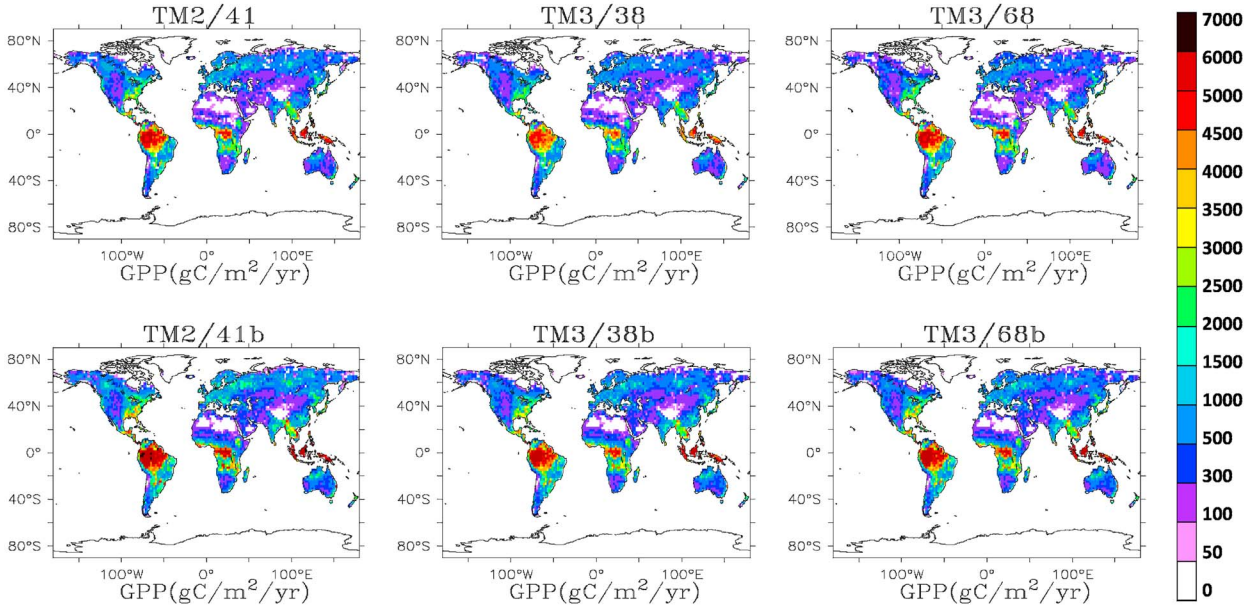

Figure 2. Spatial distribution of mean gross primary productivity (GPP) inferred from 6 configurations of data and for the period 1980–1999. (left) TM2/41 shows inferred fluxes obtained from optimization but using TM2 data and 41 stations. (right) TM3/68 shows results from optimization when using TM3 model and 68 sites. (middle) TM3/38 gives results when using TM3 model and 38 sites. TM2/41, TM3/38, and TM3/68 are obtained from a Gaussian distribution on prior parameters, while TM2/41b, TM3/38b, and TM3/68b stand for the lognormal distribution.

Table 3. Gross Primary Productivity (GPP in GtC/yr) Obtained From Each of the 13 Plant Functional Types (PFT) Used in BETHY Model Reported When Using the 6 Configurations of Optimized Parameters, as Described in Table 1[a]

	GPP (GtC/yr)							Parametric Uncertainty GPP (GtC/yr)	
PFT	Prior	TM2/41	TM3/38	TM3/68	TM2/41b	TM3/38b	TM3/68b	Prior	TM3/68b
TrEv	36	53	49.2	53	59.2	53.6	56.1	3	2
TrDec	18.6	20.9	18	21.8	27.3	24.2	23.8	2.7	2
TmpEv	0.5	0.6	0,5	0,6	0.6	0.5	0.5	2.6	2.1
TmpDec	4	9.9	8.3	10.1	12.4	11	12.4	2.7	2.2
EvCn	8.3	9.7	6.5	3.7	11.5	8.1	6.5	3.2	1.8
DecCn	1.1	2.3	1.8	2	3.1	2.3	2.3	2.6	2.1
EvShr	4.7	8.6	8.2	7.6	11.6	10.6	11	2.6	2
DecShr	2.5	2.4	2	2.1	2.6	2.5	2.4	2.6	2.1
C3Gr	20.1	18	14.3	16.3	16.1	13	13.1	3.3	1.9
C4Gr	34.5	−2.3	−0.5	−2.3	6.5	7.2	6.5	3.1	2.3
Tund	2	2.3	0.6	1	2.7	0.8	0.9	2.7	2
Wetl	0.6	−4.5	−6.5	−4.3	0.5	0.4	0.4	2.6	2.1
Crop	12	6.1	6.3	5.2	10.1	9.3	9.5	2.7	2
all PFTs	145	127	109	117	164	144	146	4.7	2

[a]The parametric uncertainties on GPP inferred from prior and optimized parameters by using TM3/68b data set are shown. TM2/41, TM3/38, and TM3/68 stand for optimizations by using Gaussian PDF on prior parameters, while TM2/41b, TM3/38b, and TM3/68b stand for lognormal PDF. See Figure 1 for a description of the PFTs.

productivity of crops contradicting in situ observations [e.g., *Beerling and Quick*, 1995].

[39] All the three optimizations when using Gaussian PDFs for the prior parameters give values of $f_{R,growth}$ less than 1 (Table 1). We see from equation (6) that a value of $f_{R,growth} < 1$ yields unrealistic negative growth respiration. We note that even in this unrealistic case the sum of maintenance and growth respiration (the difference between NPP and GPP) can still be reasonable even if the individual components are not. When we impose a lognormal distribution on the parameter to bound its value above 1, results show also a large reduction (close to 1; see Table 1).

[40] Finally, it is worth noting the large shifts of the quantum efficiency α_q (increase from its prior) and the slope $a_{\Gamma,T}$ (reduced from its prior) (Table 1). Such a reduction of $a_{\Gamma,T}$ decreases the value of the CO_2 compensation point Γ^* (see section 2.2.1), which in turn can increase GPP for C3 plants (equations (1)–(3)). The impact of these changes in GPP is discussed in more details in section 5.1.1.

[41] We found that the unrealistic negative values of V_{max} for C4-grass and swamp vegetation (Wetl) are mainly due to the contribution of Antarctic and Southern Ocean stations, far distant from the subtropical savannas described by the parameter. This immediately casts doubt on the result. It is almost certain, for example, that the large-scale transport between the savannas and these stations contains correlated errors so that the inclusion of many such stations with errors assumed to be independent overestimates their contribution. Both high frequency continuous observations and stations placed nearby C4 grass regions would reduce this unreasonable dependence.

[42] The uncertainties in the optimized parameters have been computed by using the inverse of the Hessian (i.e., the second derivative; see equation (11)) of the cost function with respect to the parameter x, as performed by *Rayner et al.* [2005] and *Scholze et al.* [2007]. The results are not shown for brevity sake. Overall, for the majority of the parameters, there is only a slight reduction of the prior uncertainty (up to 5%). The largest reductions are obtained for the scaling parameter β (equation (9)). As an example, the uncertainty in β for the tropical evergreen forest (TrE*v*) which was initially 25% of its prior value, dropped to 12% for TM3/68.

5. Gross Primary Productivity GPP

5.1. Optimized GPP From CCDAS

5.1.1. GPP Patterns

[43] Table 2 shows a range of estimates of GPP for different regions and configurations. Most configurations show a substantial increase in global GPP from the prior value, dominated by increases in tropical forests (Figure 2). Overall, about 92% of the pixels show GPP less than 3500 gC/m^2/yr. About 5% of the total pixels show GPP between 3500 gC/m^2/yr and 5000 gC/m^2/yr and they are in the tropics. Very few pixels (e.g., 3 for TM3/68b) show very large GPP (>6000 gC/m^2/yr) in the tropics. Such extreme GPP values are also reported by *Kattge et al.* [2009] when using V_{max} for tropical forests similar to our optimizations. In this study, the large GPP found in tropics are due to both optimized V_{max} and the global parameters relative to the energy activation α_q (i.e., quantum efficiency) and the slope $a_{\Gamma,T}$ between the CO_2 compensation point Γ^* and the vegetation temperature (see equations (1)–(3)). Thus, as obtained from the optimizations, the combination of both the increase of α_q and the decrease of $a_{\Gamma,T}$ (see Table 1) leads to more primary productivity. Indeed, the increase of α_q and decrease of $a_{\Gamma,T}$ (hence the decrease of the CO_2 compensation point Γ^*) increase both the electron transport, J_E, and the carboxylation enzyme Rubisco, J_C, limiting rates (equations (1)–(3)). Given this, only a strong reduction of V_{max} can significantly decrease the GPP. As an example, in TM2/41, the strong reduction of prior V_{max} for crop (about 80%) decreases the prior GPP for this PFT only to about half of its value (Tables 1 and 3).

[44] Our preferred configuration (TM3/68b) yields an annual global GPP of 146 GtC/yr (Tables 2 and 3). The configurations do show considerable variability even though each of them is reasonably able to fit the station observations, as shown by the similar values of the cost function obtained

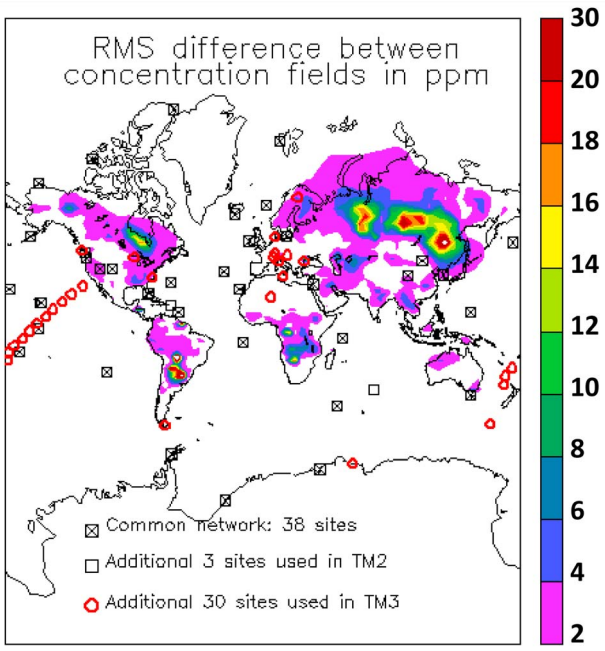

Figure 3. Root mean square difference deviation RMSD (ppm) between CO_2 concentration fields using NEP of TM3 with 68 sites and with Gaussian PDF on prior parameters (i.e., TM3/68) in contrast to TM3 with 68 sites but with lognormal PDF (i.e., TM3/68b). Simulations are performed through the global transport model LMDz-INCA.

from the 6 configurations (not shown). This suggests that the current network is not well placed to observe GPP, in particular in the tropics. Some of the range of estimates can be discounted since the unbounded parameter cases can produce negative GPP for some PFTs (Table 3).

5.1.2. Uncertainty in GPP

[45] For optimizations obtained by using the same PDF on the priors (i.e., Gaussian or lognormal), the differences among the optimized GPPs are mainly due to the transport model (Table 3 and Figure 2). As an example, the largest difference of about 18 GtC/yr is between TM2/41 and TM3/38, while similar GPP is derived from the TM3 configurations (Table 3 and Figure 2). The largest differences occur for C4 grass and swamp vegetation PFTs where relatively large negative GPP are computed when using Gaussian PDFs on prior parameters for optimizations (Table 3). Added to this uncertainty stemming from the configuration, there is also the so-called internal uncertainty due to the poor constraint of some parameters. We calculate this parameter uncertainty following *Rayner et al.* [2005] and propagate it to uncertainties in global and regional GPP following *Scholze et al.* [2007] (see equation (10)). Since the uncertainties on posterior GPPs obtained from the 6 configurations are similar, only results from the TM3/68b set up along with the prior uncertainties are reported in Table 3. The largest uncertainties are found for tropical evergreen forest, evergreen coniferous forest, C3 grass, and crop vegetation types. These uncertainties are significantly reduced after the optimization: the prior uncertainties are reduced up to 57%. Quadratically summing the uncertainties from different configurations (i.e., the standard deviation from the optimized GPPs reported in Tables 2 and 3) and the parametric uncertainty (Table 3) yields a total uncertainty in global GPP of about 19 GtC/yr.

5.1.3. Testing the Network With LMDz

[46] The sensitivity of the estimated GPP to the configuration is striking given that all configurations fit the observations equally well. Indeed, as earlier mentioned we find similar values of the cost functions from all the configurations. This prompts the question whether the poor constraint on GPP is a fundamental problem of atmospheric networks or a property of the particular networks we have used. Recall that atmospheric CO_2 concentrations are directly sensitive to NEP rather than component fluxes such as GPP. One can assess the constraint available from the atmosphere by asking whether atmospheric concentrations anywhere are sensitive to the range of GPP estimates we calculate. We cannot test this with TM2 or TM3 since we only have access to the Jacobians at particular stations. We simulated surface CO_2 concentrations using the global climate model LMDz-INCA

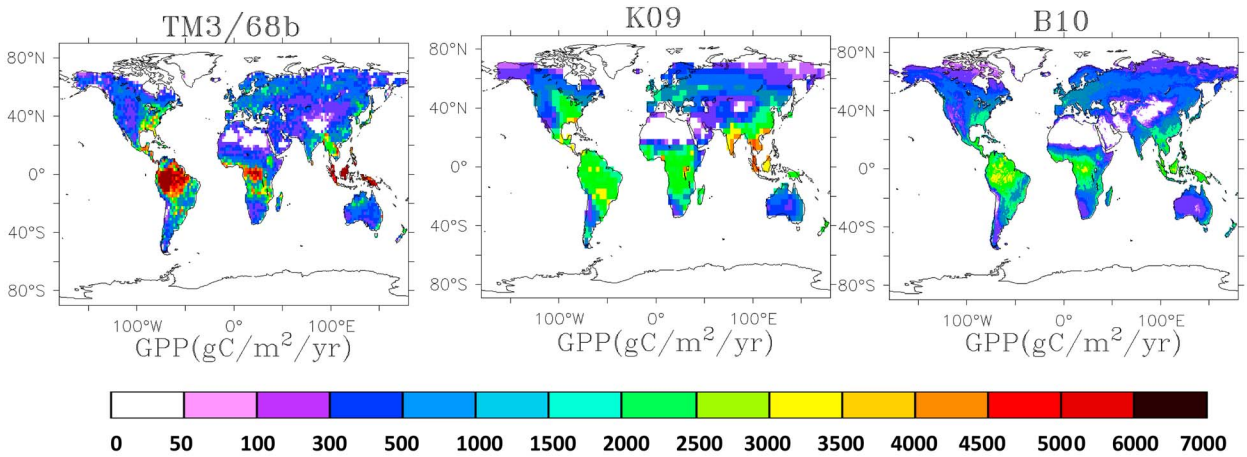

Figure 4. Comparison of mean annual GPP from the TM3/68b configuration (2° × 2° spatial resolution) and the JSBACH-based estimate by *Kattge et al.* [2009] (K09; 3.75° × 3.75° spatial resolution) and to the data-driven estimate by *Beer et al.* [2010] (B10; 0.5° × 0.5° spatial resolution).

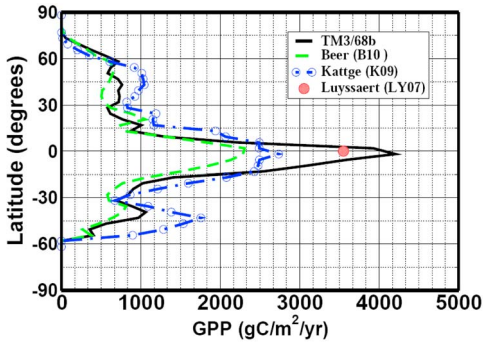

Figure 5. Latitudinal distribution of the mean gross primary productivity (GPP) within a 5° latitudinal band. TM3/68b stands for GPP obtained from optimized BETHY by using TM3 model and 68 CO_2 concentration observational sites by using a lognormal PDF on prior parameters. B10 corresponds to GPP obtained from *Beer et al.* [2010] and K09, those estimated from *Kattge et al.* [2009]. LY07 stands for GPP for tropical forests reported by *Luyssaert et al.* [2007].

(see section 2.2.2 for further details on the model) using NEP obtained from the cases TM3/68 and TM3/68b. We created a monthly mean climatology of the NEP fields for the two configurations. We ran LMD_Z for 6 years using these two cases as inputs then discarded the first three years as spin-up. The simulated mixing ratios at the surface were analyzed. The differences between the two simulations were quantified by the root mean square difference computed both in space and time. Results first confirm the CCDAS simulation in the sense that only small differences are found at stations of the studied networks (Figure 3). There are, however, large differences elsewhere such as Central Asia, South America, Hudson's Bay in Canada, and part of Africa. This suggests that the observation of productivity is possible using an atmospheric network but requires quite different placement of stations.

[47] While it would be possible to design an observational network specifically for GPP this is likely to be dependent on details of modeled transport. The results do reinforce the utility of the spatially dense measurements of CO_2 provided by satellites.

5.2. Comparison of Our Inferred GPP to Previous Findings

[48] *Kattge et al.* [2009], using leaf level nitrogen content and photosynthesis measurements, optimized the parameters of the photosynthesis model of *Farquhar et al.* [1980] (see equations (1)–(6) in this study for details). In a Bayesian inversion approach using a Markov Chain Monte Carlo sampling technique, they optimized the parameters relevant for the GPP computation. Their optimization substantially reduced the productivity of tropical trees and moderately increased the productivities for temperate trees compared to their prior estimate. GPP estimates of their model (JSBACH) were validated against a data sets of forest biome ecosystem measurement sites [*Luyssaert et al.*, 2007]. They produced a global estimate for the period 1980–1999, which was 163 GtC/yr. The study of *Kattge et al.* [2009] is called hereafter K09.

[49] *Beer et al.* [2010] estimated mean terrestrial GPP for the period 1998–2005 on the basis of NEP measurements by the eddy covariance technique at flux towers and large-scale explanatory variables from remote sensing and climate re-analyses. Flux tower NEP was partitioned into GPP and ecosystem respiration based on both nighttime and day-time NEP data. Such derived GPP was further scaled from flux tower sites to the globe by using several schemes including machine learning algorithms or regressions to the intercepted light or transpired water. For doing so, explanatory variables at large scales such as land cover, fraction of absorbed photosynthetically active radiation, leaf area index, short-wave radiation, vapor pressure deficit, air temperature, precipitation, and runoff were used. The median map used here for comparison was calculated from the full GPP distribution

Table 4. Comparison of GPP Obtained From the Optimization From TM3/68b to Results of *Beer et al.* [2010] (B10)[a]

BETHY Map: TM3/68b			B10 Biome Map				
				TM3/68b		B10	
PFTs	Area (10^{12} m^2)	Mean GPP per Area ($gC/m^2/yr$)	Biome Types	Area (10^{12} m^2)	Mean GPP per Area ($gC/m^2/yr$)	Area (10^{12} m^2)	Median GPP per Area ($gC/m^2/yr$)
TrEv+TrDec	19.4	4127.2	tropical forests	18.1	3447.9	20.3	2332.9
TmpEv+TmpDec	4.0	3288.4	temperate forests	12.3	1094.6	13.6	954.4
EvCn+DecCn	11.0	800.3	boreal forests	14.3	605.6	14.6	604.6
C4 grass	25.5	254.3	tropical savannahs and grasslands	18.8	1249.6	18.8	1134.9
C3 grass+DecShr	24.6	632.1	temperate grasslands and shrublands	10.3	496.4	11.2	478.2
EvShr	19.3	569.8	deserts	22.7	379.6	17.4	231.5
Tundra	7.6	118.4	tundra	5.0	227.8	7.6	292.9
Crop	9.7	977.3	croplands	14.8	1057.7	15.8	1094.6
Welt	1.2	346.7	wetlands	0.9	1056.7	0.9	1051.2
-	-	-	not classified	5.1	1191.9	-	-
Total/mean	122.3	1190.0	total/mean	122.3	1190.0	120.4	999.6

[a]TM3/68b stands for GPP obtained from optimized BETHY by using TM3 model and 68 CO_2 concentration observational sites and by using a lognormal PDF on prior parameters. Two methods are used for the comparison: i) BETHY map: correspondence of BETHY PFTs to B10 biome types and ii) B10 biome map: GPP derived from TM3/68b mapped onto B10 biome types. The mean GPP at global scale is given by [mean GPP per area] * [Area of the biome] / [Total area]. When using the method ii), some pixels of BETHY were not associated to any of the biome types. The corresponding GPP from TM3/68b are reported as "Not classified." The definition of the BETHY PFTs acronym can be found in Figure 1.

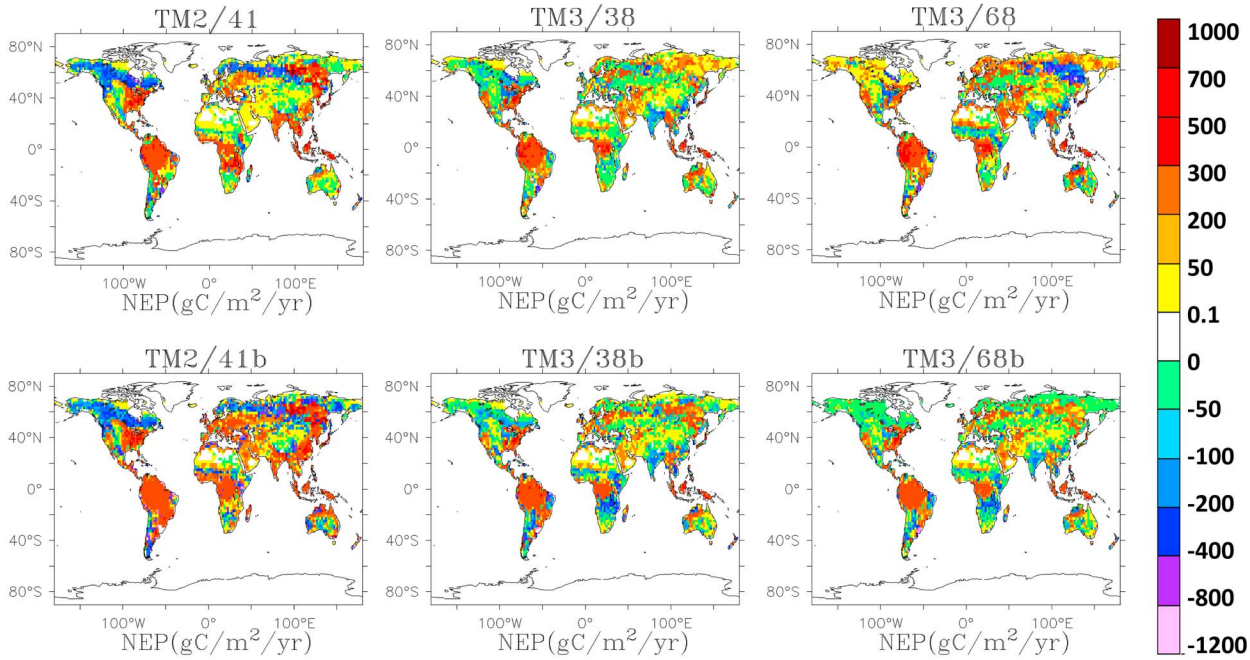

Figure 6. As in Figure 2, but for the net ecosystem productivity NEP. Positive values stand for uptake of CO_2 by the terrestrial biosphere.

representing uncertainties due to flux partitioning, model parameters, explanatory variables, model structure and scaling approach. The overall terrestrial GPP was estimated at 123 GtC/yr with a median absolute deviation of 8 GtC/yr and a 95% confidence interval between 102 and 135 GtC/yr [*Beer et al.*, 2010, hereinafter B10].

[50] We compare global and regionally aggregated estimates from the case TM3/68b to K09 and B10 estimates. Since the approach in this study shares a common photosynthesis model with K09, we focus on the more independent estimates from B10.

[51] Figure 4 presents the maps of the mean GPP for TM3/68b, K09, and B10. Regional aggregates are displayed in Table 2 together with their latitudinal distributions in Figure 5. TM3/68b shows similar spatial patterns with B10 in the extra-tropics, while it exhibits large differences with both B10 and K09 in the Tropics (Figures 4 and 5). Our GPP estimates in the tropics are greater than all the previous findings. Our results are similar to the GPP reported by *Luyssaert et al.* [2007] (Figure 5). The study of *Luyssaert et al.* [2007] probably overestimates regional GPP since it used a few productive sites. The differences between TM3/68b and K09 arise from the low V_{max} for C3 grass, crop, and C4 grass (C3 herbaceous in K09) PFTs and the high V_{max} of both tropical evergreen forest and temperate deciduous forest PFTs obtained from the TM3/68b optimization (Table 1). In detail, V_{max} from TM3/68b for both the tropical forests (63.8 μmol m^{-2} s^{-1}) and crop (47.9 μmol m^{-2} s^{-1}) are about twice their corresponding values reported in K09. For temperate deciduous forests, V_{max} of 149.2 μmol m^{-2} s^{-1} is obtained from TM3/68b, while it is only 29 μmol m^{-2} s^{-1} in K09. As already discussed in section 4, TM3/68b gives a very low V_{max} for C4 grass (0.7 μmol m^{-2} s^{-1}) compared to 78.2 μmol m^{-2} s^{-1} used in K09. Moreover, the global parameters controlling energy activation such as α_q and $a_{\Gamma,T}$ contribute to the differences in GPPs between K09 and TM3/68b (see section 5.1.1 for details).

[52] Surprisingly, all four analyzed cases show good agreement on the partitioning of GPP between the Northern Hemisphere and Tropics with a ratio around 0.40 (Table 2).

[53] We have computed mean GPP per unit area for 9 biome types as in B10. Thus, for a given B10 biome type, all the pixels of BETHY that fall in the pixels of this biome are considered. We obtain differences in the two maps for which we lose 4.2% of the total GPP from BETHY. In fact, the two vegetation maps can be different at some grid points. As an example, TM3/68b can describe grassland for a given grid point in its vegetation map, while B10 considers it as forest in its map. Thus, we have also made a correspondence (i.e., grouping all the PFTs that better represent a given biome type) between BETHY PFTs and B10 biome types and compute means and totals for GPP derived from BETHY. The main driver of GPP at a point is the vegetation type rather than climate, as performed when mapping BETHY PFTs on B10 biome map. These two GPP characterizations give both the lower and upper limits of GPP derived from BETHY for a given biome type.

[54] For a given biome type, the mean GPP per unit area is then computed as the total GPP for this biome divided by the area covered by the biome (PFT or group of PFTs) (Table 4). Again we use TM3/68b. Overall, a good agreement is found between TM3/68b and B10 estimates. Except for tropical and temperate forests, and savannah and grasslands, the differences between TM3/68b and B10 total GPP can be explained by the differences in their respective biome areas (Table 4). For the tropical forest biome, our mean GPP per unit area is at most 1.8 times higher than B10 estimates. As already discussed in section 5.1, this result is mainly explained by

the high productivity of the tropical forest PFTs in BETHY. The same reason can be evoked for temperate forests for which the dominant PFT in BETHY (temperate deciduous) also shows high productivity (Tables 1 and 4).

6. Sensitivity to Configuration

[55] Previous studies have found considerable sensitivity of inverse flux estimates to the set-up [*Gurney et al.*, 2002; *Law et al.*, 2003; *Gurney et al.*, 2003]. Important details include the transport model, the observing network, the prior and the discretization of the source regions. A CCDAS needs to make many of the same choices, to which we can add the choice of biosphere model. Here we explore some of these sensitivities, both for the net flux and the GPP estimates made earlier.

[56] Although we do not have two biosphere models, the choice of the prior probability distribution is, in one sense, a change to the model. As noted earlier (see section 2.2.6), the exponential transform carried out on most parameters in the "b" cases is effectively a change in the function mapping parameters to observations, i.e., the model.

[57] Figure 6 (summarized in Table 2) shows NEP derived by using the six configurations of optimized parameters. Similar NEP distributions are obtained for TM3/38 and TM3/68, which are quite different from TM2/41 results. For a given configuration, large differences also occur between optimizations using Gaussian PDFs and lognormal PDFs (Figure 6 and Table 2). We quantify the differences in NEP between two configurations by calculating the root mean square difference (rmsd) over all grid cells of BETHY and all months in the study period 1980–1999. The results show that NEP is more sensitive to the choice of PDF on the form of the prior parameters, but it also significantly sensitive to the choice of transport model. The rmsd between TM3/68 and TM3/68b (sensitivity to PDF) is 143 gC/m^2/yr, while 111 gC/m^2/yr is obtained between TM2/41 and TM3/38 (sensitivity to transport model). The rmsd between TM3/38 and TM3/68 (sensitivity to observing network) is 42 gC/m^2/yr. The use of a lognormal PDF on prior parameters reduced the sensitivity of NEP to both transport model and network. The rmsd between TM2/41b and TM3/38b is 80 gC/m^2/yr, while a small rmsd (16 gC/m^2/yr) is obtained between TM3/38b and TM3/68b. All the six optimizations give the largest uptake over the Tropics (Table 2). In general, the large-scale patterns of NEP produced with CCDAS do not mirror those of direct inversions using the same models. *Gurney et al.* [2002] estimated a northern extra-tropical terrestrial flux to the atmosphere of −0.74 GtC/yr for TM2 and −3.15 GtC/yr for TM3. For the tropics these fluxes were −0.18 GtC/yr for TM2 and +1.40 GtC/yr for TM3. These figures cannot be compared directly with CCDAS because of different region boundaries and treatments of land-use change, but we can note the differences in the partitioning of NEP between the Northern Hemisphere and Tropics.

[58] In summary we see that the patterns of NEP within CCDAS are equally sensitive to the choice of PDF on prior parameters and the choice of transport model. This transport model sensitivity is not greatly reduced compared to direct inversions.

[59] The case for GPP is rather different. As one might expect, the terrestrial model BETHY intervenes more strongly in the transfer of information from atmospheric measurements to indirectly observed quantities like GPP than it does for the NEP. Thus there are large differences in all cases between the Gaussian and lognormal prior PDFs (Tables 2 and 3). The differences between the optimized GPPs were quantified by the root mean square difference (rmsd) computed both in space and time. Results show that the largest differences are driven by the PDF on the prior parameters. A rmsd of 242 gC/m^2/yr is found between TM3/68 and TM3/68b. Differences between TM2/41 and TM3/38 (sensitivity to transport model) give a rmsd of 130 gC/m^2/yr which is similar to 126 gC/m^2/yr obtained between TM3/38 and TM68 (sensitivity to observing network). Again, the use of a lognormal PDF on prior parameters reduced the sensitivity of GPP to both transport model and observing network, but with a larger reduction to the network: a rmsd of 104 gC/m^2/yr is now found between TM2/41b and TM3/38b, while a small rmsd (33 gC/m^2/yr) is found between TM3/38b and TM3/68b.

7. Discussions

[60] One of the two goals of this study was to understand the interplay between the dynamical constraint offered by BETHY and the observational constraint from the atmosphere. We see that as we change observational network, transport model or the form of the prior probability distribution it is this last which has the biggest impact. We should emphasize that this is not a question of the choice of prior values for parameters since these are the same in all cases. Rather the prior probability distribution affects the mapping between parameters within the optimization and the observables. In one important sense this is a vindication for the use of methods like data assimilation for carbon-cycle studies. Had the observational constraint produced results that did not depend on the underlying model we would avoid the extra effort and assumptions of data assimilation and choose a direct inversion. The relative dominance of the dynamical constraint is more a measure of the weakness of the observational constraint than of the inherent power of the dynamics. This is shown by the similarity of the fit to the observations generated by quite different simulations (e.g., different GPP). It is worth noting that NEP is also significantly sensitive to transport model and this is not greatly reduced compared to direct inversion.

[61] The results provide an important context for previous CCDAS studies like *Rayner et al.* [2005] and *Scholze et al.* [2007]. In particular these previous studies produce impressively small uncertainties for calculated quantities such as net fluxes or concentrations. These studies acknowledge that this is a result of the small number of unknown parameters (57 for this study) and the consequent over-determination of the problem. The sensitivity of our results to the prior probability distributions suggests that these parametric uncertainties do not capture the overall uncertainty of the system. While the projection of parametric uncertainty onto calculated quantities remains useful it is now clear that other, less precise but more complete methods will also be necessary to characterize the total uncertainty.

[62] It is worth mentioning that uncertainties in our inferred GPPs may depend on uncertainties on the

background fluxes used to compute CO_2 concentrations as well as uncertainties on the parameters related to e.g., fAPAR.

[63] Part of the motivation for this study was to extend the diagnostic application of CCDAS beyond NEP. The dominance of the dynamical constraint also holds for GPP. Quite large changes in GPP do not project onto significant changes in concentration. Our mapping of the concentrations arising from the two different prior PDFs suggests the problem is not inherent in the use of concentration data but a result of a particular network. This is good news but poses its own problem since a network optimal for constraining one quantity is little use for another. Indeed, NEP is clearly related to the atmospheric CO_2 concentration, while GPP depend on its interaction with the plant during the photosynthesis process and hence to the characteristics of this plant. We suggest that a station relevant for GPP may capture much of the singularities of each type of vegetation, while for NEP this can be less important. Consequently, for the GPP, one needs to place stations close or within each of the regions that characterize the main vegetation type. This supports the use of measurement approaches with good spatial coverage such as the current and future measurements of CO_2 from space [*Crisp et al.*, 2004; *Hamazaki et al.*, 2004]. Of course this does not reduce the value of measurements elsewhere in the chain of processes linking photosynthesis to NEP.

[64] This study has only considered one of the potential observational constraints on GPP. The methodology we have used is capable of treating the same data as used by *Kattge et al.* [2009] and *Beer et al.* [2010]. It remains an important question whether the joint observational constraint of these data sets used together will further refine our estimates, i.e., are they mutually consistent with the dynamics imposed by BETHY.

[65] Our results have clearly shown that parameters relevant for GPP are poorly constrained in some sensitive areas due to the coarse observing network of CO_2 concentration measurements we are using. Therefore, the immediate effort to be invested to improve the current CCDAS should be on the characterization of the observations relevant to assimilate the parameters of BETHY. We recommend the use of i) the new prior parameters relevant to GPP [*Kattge et al.*, 2009], ii) the fluorescence data from the GOSAT satellite [e.g., *Frankenberg et al.*, 2011], iii) the high frequency and continuous CO_2 concentrations from the current network, and iv) the leaf level observations [*Ziehn et al.*, 2011] to constrain these process parameters.

8. Conclusions

[66] We have studied the sensitivity of estimates using a carbon-cycle data assimilation system to choices of atmospheric concentration network, transport model and the choice of prior probability distributions of parameters. We have extended this sensitivity analysis to GPP as well as NEP or net flux. Our conclusions can be summarized as follows:

[67] 1) Estimates of NEP and GPP are more sensitive to the choice of prior PDF for the parameters than the choices of transport model and observing network, however, we must note that our alternative choice of prior PDF was motivated by restricting some of the control parameters to their physically meaningful domains. It is worth noting that NEP is also significantly sensitive to transport model and this is not greatly reduced compared to direct inversion.

[68] 2) Our best-case estimate of GPP is 146 $\pm$ 19 GtC/yr globally. This cannot be distinguished by the current atmospheric network from the lower estimate of 117 $\pm$ 19 GtC/yr from the alternative choice of PDF. Our GPP estimates span a wide range, especially in the Tropics where our approach gives higher values for all the configurations than most other approaches.

[69] 3) The atmospheric constraint of GPP could be improved by additional CO_2 observations in Asia, parts of America, and African rain forest as well as in C4 grass regions.

Notation

[70] The following notation is relevant to the BETHY model description (section 2.2.1).

GPP gross primary productivity (μmol m^{-2} s^{-1})
J_C minimum rate of electron transport limited by the carboxylation enzyme Rubisco (μmol m^{-2} s^{-1}).
J_E minimum rate of electron transport (μmol m^{-2} s^{-1}).
J_c gross photosynthetic product for CO_2 limited capacity (μmol m^{-2} s^{-1}).
J_e gross photosynthetic rate limited by the photosynthetic enzyme Rubisco (μmol m^{-2} s^{-1}).
J_i gross photosynthetic rate limited by the amount of available light (μmol m^{-2} s^{-1}).
$V_{\max}$ maximum rate of Rubisco carboxylation (μmol m^{-2} s^{-1}).
C_i intercellular CO_2 concentration (μmol mol^{-1}).
Γ^* CO_2 compensation point in the absence of dark respiration (μmol mol^{-1}).
Kc Michaelis-Menten constant for CO_2 (μmol mol^{-1}).
K_O Michaelis-Menten constant for O_2 (μmol mol^{-1}).
O_x partial pressure of oxygen (hPa).
α_q the quantum efficiency (electrons s^{-1} / photons s^{-1}).
α_{i} the integrated C4 quantum efficiency (electrons s^{-1} / photons s^{-1}).
I photosynthetically active radiation (PAR) absorption rate (μmol m^{-2} s^{-1}).
$J_{\max}$ maximum rate of electron transport (μmol m^{-2} s^{-1}).
k PEP case (the initial CO_2 fixating enzyme in C4) CO_2 specificity (mol m^{-2} s^{-1}).
$a_{J,V}$ the slope between k and $V_{\max}$ (dimensionless).
$a_{\Gamma,T}$ the slope between Γ^* and the vegetation temperature (μmol mol^{-1} °C^{-1}).
E activation energy (J mol^{-1}).
NPP net primary productivity (μmol m^{-2} s^{-1}).
R_M maintenance respiration of the plant (μmol m^{-2} s^{-1}).
R_G growth respiration of the plant (μmol m^{-2} s^{-1}).
R_d dark (mitochondrial) respiration (μmol m^{-2} s^{-1}).
$f_{R,leaf}$ the fraction of the maintenance respiration due to the leaf (dimensionless).
$f_{R,growth}$ the amount of carbon respired for a unit gain in vegetation biomass (dimensionless).
NEP net ecosystem productivity (μmol m^{-2} s^{-1}).
R_S heterotrophic respiration (μmol m^{-2} s^{-1}).
$R_{S,f}$ heterotrophic respiration due to litter pool respiration (μmol m^{-2} s^{-1}).

$R_{S,s}$ heterotrophic respiration due to slow carbon respiration carbon pool (μmol m^{-2} s^{-1}).
f_s fraction of decomposition from the litter pool that goes to slow pool (dimensionless).
C_f size of the litter pool (μmol m^{-2} °C^{-1}).
C_s size of the slow litter pool (μmol m^{-2} °C^{-1}).
ω soil moisture (dimensionless).
T_a air temperature (°C).
κ soil moisture dependence parameter (dimensionless).
Q_{10f} temperature dependence parameter relative for litter pool (dimensionless).
Q_{10s} temperature dependence parameter relative for slow pool (dimensionless).
τ_f the litter pool turnover time at 25°C (s).
τ_s the slow pool turnover time at 25°C (s).
β scaling factor (dimensionless).

[71] **Acknowledgments.** We acknowledge the financial support of the European Commission through the IMECC Integrated Infrastructure Initiative (I3) project under the 6th Framework Program (contract 026188). PR is in receipt of an ARC Professorial Fellowship (DP1096309). We would like to sincerely thank both reviewers for thoughtful, thorough and constructive comments on the manuscript.

References

Andres, R. J., G. Marland, I. Fung, and E. Matthews (1996), A 1° × 1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cycles*, *10*, 419–429, doi:10.1029/96GB01523.

Aubinet, M., et al. (2000), Estimates of the annual net carbon and water exchange of forest: The EUROFLUX methodology, *Adv. Ecol. Res.*, *30*, 114–173.

Baker, D. F., et al. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003, *Global Biogeochem. Cycles*, *20*, GB1002, doi:10.1029/2004GB002439.

Baldocchi, D. D. (2003), Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future, *Global Change Biol.*, *9*, 479–492, doi:10.1046/j.1365-2486.2003.00629.x.

Beer, C., M. Reichstein, P. Ciais, G. D. Farquhar, and D. Papale (2007), Mean annual GPP of Europe derived from its water balance, *Geophys. Res. Lett.*, *34*, L05401, doi:10.1029/2006GL029006.

Beer, C., et al. (2009), Temporal and among-site variability of inherent water use efficiency at the ecosystem level, *Global Biogeochem. Cycles*, *23*, GB2018, doi:10.1029/2008GB003233.

Beer, C., M. Reichstein, E. Tomelleri, P. Ciais, and M. Jung (2010), Terrestrial gross carbon dioxide uptake: Global distribution and co-variation with climate, *Science*, *329*, 834–838, doi:10.1126/science.1184984.

Beerling, D. J., and W. P. Quick (1995), A new technique for estimating rates of carboxylation and electron transport in leaves of C3 plants for use in dynamic global vegetation models, *Global Change Biol.*, *1*, 289–294, doi:10.1111/j.1365-2486.1995.tb00027.x.

Brenkert, A. L. (1998), Carbon dioxide emission estimates from fossil fuel burning, hydraulic cement production, and gas flaring for 1995 on a one degree grid cell bases, http://cdiac.esd.ornl.gov/ftp/ndp058a/, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Clark, D. A., S. Brown, D. W. Kicklighter, J. Q. Chambers, J. R. Thomlinson, and J. Ni (2001a), Measuring net primary production in forests: Concepts and field methods, *Ecol. Appl.*, *11*, 356–370, doi:10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2.

Clark, D. A., S. Brown, D. W. Kicklighter, J. Q. Chambers, J. R. Thomlinson, J. Ni, and E. A. Holland (2001b), Net primary production in tropical forests: An evaluation and synthesis of existing data, *Ecol. Appl.*, *11*, 371–384, doi:10.1890/1051-0761(2001)011[0371:NPPITF]2.0.CO;2.

Collatz, G. J., M. Ribas-Carbo, and J. A. Berry (1992), Coupled photosynthesis-stomatal conductance model of leaves for C4 plants, *J. Aust. Plant Physiol.*, *19*, 519–538, doi:10.1071/PP9920519.

Crisp, D., et al. (2004), The Orbiting Carbon Observatory (OCO) mission, *Adv. Space Res.*, *34*, 700–709, doi:10.1016/j.asr.2003.08.062.

Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, 392 pp., Cambridge Univ. Press, Cambridge, U. K., doi:10.1017/CBO9780511535741.

Farquhar, G. D. (1988), Models relating subcellular effects of temperature to whole plant response, *Symp. Soc. Exp. Biol.*, *42*, 395–409.

Farquhar, G. D., S. V. Cammerer, and J. A. Berry (1980), A biochemical model of photosynthesis in leaves of C4 species, *Planta*, *149*, 78–90, doi:10.1007/BF00386231.

Foken, T., and B. Wichura (1996), Tools for quality assessment of surface-based flux measurements, *Agric. For. Meteorol.*, *78*, 83–105, doi:10.1016/0168-1923(95)02248-1.

Frankenberg, C., et al. (2011), New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, *Geophys. Res. Lett.*, *38*, L17706, doi:10.1029/2011GL048738.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *Trans. Math. Software*, *24*, 437–474, doi:10.1145/293686.293695.

Gilbert, J. C., and C. Lemaréchal (1989), Some numerical experiments with variable-storage quasi-Newton algorithms, *Math. Program.*, *45*, 407–435, doi:10.1007/BF01589113.

GLOBALVIEW-CO2 (2004), Cooperative Atmospheric Data Integration Project—Carbon dioxide [CD-ROM], Global Monit. Div., Earth Syst. Res. Lab., NOAA, Boulder, Colo. [Available at ftp://ftp.cmdl.noaa.gov/ccg/co2/GLOBALVIEW]

Gurney, K. R., et al. (2002), Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, *Nature*, *415*, 626–630, doi:10.1038/415626a.

Gurney, K. R., et al. (2003), TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior fluxes, *Tellus, Ser. B*, *55*, 555–579, doi:10.1034/j.1600-0560.2003.00049.x.

Gurney, K. R., et al. (2004), Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, *Global Biogeochem. Cycles*, *18*, GB1010, doi:10.1029/2003GB002111.

Hamazaki, T., Y. Kaneko, and A. Kuze (2004), Carbon dioxide monitoring from the GOSAT satellite, paper presented at XXth ISPRS Conference, Int. Soc. for Photogramm. and Remote Sens., Istanbul. [Available at http://www.cartesia.org/geodoc/isprs2004/comm7/papers/43.pdf]

Hauglustaine, D. A., F. Hourdin, L. Jourdain, M. A. Filiberti, S. Walters, J. F. Lamarque, and E. A. Holland (2004), Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, *J. Geophys. Res.*, *109*, D04314, doi:10.1029/2003JD003957.

Heimann, M. (1995), The global atmospheric tracer model TM2, *Tech. Rep., 10*, Dtsch. Klimarechenzent., Hamburg, Germany.

Heimann, M., and S. Körner (2003), The global atmospheric tracer model TM3, *Tech. Rep., 5*, 131 pp., Max-Planck-Inst. für Biogeochem., Jena, Germany.

Houghton, R. A. (2008), Carbon flux to the atmosphere from land-use changes: 1850–2005, in *Trends: A Compendium of Data on Global Change* [online], Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.

Houghton, R. A., et al. (1987), The flux of carbon from terrestrial ecosystems to the atmosphere in 1980 due to changes in land use: Geographic distribution of the global flux, *Tellus, Ser. B*, *39*, 122–139, doi:10.1111/j.1600-0889.1987.tb00277.x.

Imhoff, M. L., and L. Bounoua (2006), Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data, *J. Geophys. Res.*, *111*, D22S12, doi:10.1029/2006JD007377.

Imhoff, M. L., L. Bounoua, T. Ricketts, C. Loucks, R. Harriss, and W. T. Lawrence (2004), Global patterns in human consumption of net primary production, *Nature*, *429*, 870–873, doi:10.1038/nature02619.

Kaminski, T., M. Heimann, and R. Giering (1999), A coarse grid three-dimensional global inverse model of the atmospheric transport: 2. Inversion of the transport of CO_2 in the 1980s, *J. Geophys. Res.*, *104*(D15), 18,555–18,581, doi:10.1029/1999JD900146.

Kaminski, T., W. Knorr, P. Rayner, and M. Heimann (2002), Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochem. Cycles*, *16*(4), 1066, doi:10.1029/2001GB001463.

Kaminski, T., R. Giering, M. Scholze, P. Rayner, and W. Knorr (2003), An example of an automatic differentiation-based modelling system, in *Computational Science and Its Applications—ICCSA 2003*, *Lect. Notes in Comput. Sci.*, vol. 2668, edited by V. Kumar et al., pp. 95–104, Springer, Berlin.

Kattge, J., W. Knorr, T. Raddatz, and C. Wirth (2009), Quantifying photosynthesis capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models, *Global Change Biol.*, *15*(4), 976–991, doi:10.1111/j.1365-2486.2008.01744.x.

Knorr, W. (2000), Annual and interannual CO_2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, *Global Ecol. Biogeogr.*, *9*, 225–252, doi:10.1046/j.1365-2699.2000.00159.x.

Knorr, W., and J. P. Schulz (2001), Using satellite data assimilation to infer global soil moisture and vegetation feedback to climate, in *Remote Sensing and Climate Modeling: Synergies and Limitations*, edited by M. Beniston and M. Verstraete, pp. 273–306, Kluwer Acad., Dordrecht, Netherlands.

Krinner, G., N. Viovy, N. de Noblet-Ducoure, 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 Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Lasslop, G., M. Reichstein, D. Papale, A. Richardson, A. Arneth, A. Barr, P. Stoy, and G. Wohlfahrt (2010), Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation, *Global Change Biol.*, *16*, 187–208, doi:10.1111/j.1365-2486.2009.02041.x.

Law, R. M., et al. (1996), Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, *Global Biogeochem. Cycles*, *10*(4), 783–796, doi:10.1029/96GB01892.

Law, R. M., C. Yu-Chan, K. R. Gurney, P. Rayner, A. S. Denning, and TransCom3 Modelers (2003), TransCom 3 CO_2 inversion intercomparison: 2. Sensitivity of annual mean results to data choices, *Tellus, Ser. B*, *55*(2), 580–595, doi:10.1034/j.1600-0889.2003.00053.x.

Le Quéré, C., et al. (2003), Two decades of ocean CO_2 sink and variability, *Tellus, Ser. B*, *55*(2), 649–656, doi:10.1034/j.1600-0560.2003.00043.x.

Lieth, H. F. H. (1973), Primary production: Terrestrial ecosystems, *Hum. Ecol.*, *1*, 303–332, doi:10.1007/BF01536729.

Luyssaert, S., et al. (2007), CO_2 balance of boreal, temperate, and tropical forest derived from global database, *Global Change Biol.*, *13*, 2509–2537, doi:10.1111/j.1365-2486.2007.01439.x.

Malmström, C. M., M. V. Thompson, G. Juday, S. O. Los, J. T. Randerson, and C. B. Field (1997), Interannual variation in global scale net primary production: Testing model estimates, *Global Biogeochem. Cycles*, *11*, 367–392.

Marland, G., T. A. Boden, and R. J. Andres (2001), Global, regional, and national CO_2 emission, in *Trends: A Compendium of Data on Global Change* [online], Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.

Masarie, K. A., and P. P. Tans (1995), Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record, *J. Geophys. Res.*, *100*(D6), 11,593–11,610, doi:10.1029/95JD00859.

McMurtrie, R. E., R. J. Norby, B. E. Medlyn, R. C. Dewar, D. A. Pepper, P. B. Reich, and C. V. M. Barton (2008), Why is plant-growth response to elevated CO_2 amplified when water is limiting but reduced when nitrogen is limiting? A growth-optimisation hypothesis, *Funct. Plant Biol.*, *35*, 521–534, doi:10.1071/FP08128.

Nijssen, B., R. Schnur, and D. Lettenmaier (2001), Retrospective estimation of soil moisture using the VIC land surface model, 1980–1993, *J. Clim.*, *14*(8), 1790–1808, doi:10.1175/1520-0442(2001)014<1790:GREOSM>2.0.CO;2.

Nowak, R. S., D. S. Ellsworth, and S. D. Smith (2004), Functional responses of plants to elevated atmospheric CO_2—Do photosynthetic and productivity data from FACE experiments support early predictions?, *New Phytol.*, *162*, 253–280, doi:10.1111/j.1469-8137.2004.01033.x.

Oleson, K. W., et al. (2010), Technical description of version 4.0 of the Community L, Model (CLM), *NCAR Tech. Note, NCAR/TN-478+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo. [Available at http://www.cesm.ucar.edu/models/ccsm4.0/clm/CLM4_Tech_Note.pdf]

Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*, 571–583, doi:10.5194/bg-3-571-2006.

Prince, S. D., J. Haskett, M. Steininger, H. Strand, and R. Wright (2001), Net primary production of US Midwest croplands from agricultural harvest yield data, *Ecol. Appl.*, *11*, 1194–1205, doi:10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2.

Rayner, P. J., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochem. Cycles*, *19*, GB2026, doi:10.1029/2004GB002254.

Rayner, P. J., R. M. Law, C. E. Allison, R. J. Francey, C. M. Trudinger, and C. Pickett-Heaps (2008), Interannual variability of the global carbon cycle (1992–2005) inferred by inversion of atmospheric CO_2 and $\delta^{13}CO_2$ measurements, *Global Biogeochem. Cycles*, *22*, GB3008, doi:10.1029/2007GB003068.

Rebmann, C., et al. (2005), Quality analysis applied on eddy covariance measurements at complex forest sites using footprint modelling, *Theor. Appl. Climatol.*, *80*(2–4), 121–141, doi:10.1007/s00704-004-0095-y.

Reichstein, M., et al. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, *Global Change Biol.*, *11*, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x.

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964, doi:10.5194/acp-3-1919-2003.

Scholze, M., T. Kaminski, P. Rayner, W. Knorr, and R. Giering (2007), Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *J. Geophys. Res.*, *112*, D17305, doi:10.1029/2007JD008642.

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

Takahashi, T., R. H. Wanninkhof, R. A. Feely, R. F. Weiss, D. W. Chipman, N. Bates, J. Olafsson, C. Sabine, and S. C. Sutherland (1999), Net sea-air CO_2 flux over the global oceans: An improved estimate based on the sea-air pCO2 difference, in *Proceedings of the Second International Symposium, CO_2 in the Oceans*, edited by Y. Nojiri, pp. 9–14, Cent. for Global Environ. Res., Natl. Inst. for Environ. Stud., Tsukuba, Japan.

Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimation*, Elsevier, Amsterdam.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Wilson, M. F., and A. Henderson-Sellers (1985), A global archive of land cover and soil data for use in general circulation climate models, *J. Climatol.*, *5*, 119–143.

Yuan, W., et al. (2010), Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data, *Remote Sens. Environ.*, *114*, 1416–1431, doi:10.1016/j.rse.2010.01.022.

Ziehn, T., J. Kattge, W. Knorr, and M. Scholze (2011), Improving the predictability of global CO_2 assimilation rates under climate change, *Geophys. Res. Lett.*, *38*, L10404, doi:10.1029/2011GL047182.

C. Beer, Biogeochemical Model-Data Integration Group, Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, D-07745 Jena, Germany.

E. N. Koffi, Laboratoire des Sciences du Climat et de l'Environnement, Ormes des merisiers, F-91191 Gif-sur-Yvette, France. (ernest.koffi@lsce.ipsl.fr)

P. J. Rayner, School of Earth Sciences, University of Melbourne, Melbourne, Vic 3010, Australia.

M. Scholze, Department of Earth Sciences, University of Bristol, Queen's Road, Bristol BS8 1RJ, UK.