JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D17305, doi:10.1029/2007JD008642, 2007

Propagating uncertainty through prognostic carbon cycle data assimilation system simulations

M. Scholze,[1] T. Kaminski,[2] P. Rayner,[3] W. Knorr,[1] and R. Giering[2]

Received 9 March 2007; revised 23 May 2007; accepted 20 June 2007; published 14 September 2007.

[1] One of the major advantages of carbon cycle data assimilation is the possibility to estimate carbon fluxes with uncertainties in a prognostic mode, that is beyond the time period of carbon dioxide (CO_2) observations. The carbon cycle data assimilation system is built around the Biosphere Energy Transfer Hydrology Scheme (BETHY) model, coupled to the atmospheric transport model TM2. It uses about 2 decades of observations of the atmospheric carbon dioxide concentration from a global network to constrain 57 process parameters via an adjoint approach. The model's Hessian matrix of second derivatives provides uncertainty estimates for the optimized process parameters that are consistent with the assumed uncertainties in the observations and the model. With those estimated parameter values, the model can predict the response of the terrestrial biosphere to prescribed climate forcing beyond the assimilation period. We develop a methodological framework that is able to propagate parameter uncertainties through such a prognostic simulation and provide uncertainty estimates for the simulation results. We demonstrate the concept for a 4-year hindcast simulation from 2000 to 2003 following a 21-year assimilation period from 1979 to 1999. We discuss prognostic uncertainties for surface fluxes and atmospheric carbon dioxide.

Citation: 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.

1. Introduction

[2] Anthropogenic emissions (combustion of fossil fuel, biomass burning, and cement manufacture) of CO_2 are mainly driving the observed increase in atmospheric CO_2 concentrations [see *Keeling et al.*, 1995]. However, only about 40% of these emissions, on average, stay in the atmosphere [*Jones and Cox*, 2005], the remainder is taken up by the oceans and terrestrial biosphere. Understanding the mechanisms and the temporal and spatial patterns of the CO_2 uptake and predicting the evolution of these fluxes into the future are therefore of major importance [*Prentice et al.*, 2001].

[3] Currently, predictions about the evolution of future carbon sinks are highly uncertain and efforts to reduce this uncertainty are of substantial scientific and policy interest. The recent developments in coupling atmosphere and ocean general circulation models with prognostic models of the terrestrial carbon cycle allow us to address this challenge in a consistent and comprehensive approach. The pioneering study by *Cox et al.* [2000] predicts that the current terrestrial carbon sink becomes a source at 2050 because of a strong positive feedback between climate and the carbon cycle. *Dufresne et al.* [2002], however, showed in a similar study using another coupled climate carbon cycle model a weakened but continuing terrestrial carbon sink until the end of the twenty-first century. The final CO_2 concentrations in 2100 as simulated by the two models differ by nearly 200 parts per million by volume (ppmv). A comparison between the two model results revealed that the discrepancies are largely caused by differences in three key processes [*Friedlingstein et al.*, 2003]. Two of these three processes are associated with the terrestrial carbon cycle, namely the contrasting responses of vegetation cover, with the model of *Cox et al.* [2000] producing the Amazon die-back under climate change [*Cox et al.*, 2004], and the response of the heterotrophic soil carbon respiration to climate change. The latter was responsible for most of the difference in terrestrial carbon storage. The response depends on the temperature sensitivity of the soil carbon decomposition but also on the allocation of biomass to the various carbon pools in the model. The third process was differences in ocean uptake of CO_2.

[4] The Coupled Carbon Cycle Climate Model Intercomparison Project (C^4MIP [*Fung et al.*, 2000]) has found large differences in the climate-carbon feedback strength as simulated by eleven different coupled models. By the end of the twenty-first century, additional CO_2 due to this positive feedback varied between 20 and 200 ppmv. The higher CO_2 levels led to an additional climate warming ranging between 0.1° and 1.5°C [*Friedlingstein et al.*, 2006]. The strength of this feedback is a major "new"

[1] Quantifying and Understanding the Earth System, Department of Earth Sciences, University of Bristol, Bristol, UK.
[2] FastOpt, Hamburg, Germany.
[3] Laboratoire des Sciences du Climat et l'Environnement, Gif-sur-Yvette, France.

0148-0227/07/2007JD008642$09.00

uncertainty in the relationship between CO_2 emissions and climate. Using an integrated assessment model *Smith and Edwards* [2006] demonstrated that uncertainty in the response of the carbon cycle to climate has significant impacts on the costs of a stabilization policy: A carbon cycle feedback of, for example, 100 ppmv is estimated to increase this cost by a factor of five, underpinning the need for a soundly based uncertainty specification and reduction.

[5] Various attempts have been undertaken to quantify the uncertainty in model predictions. For instance, the above mentioned model intercomparison project addresses structural uncertainty due to differences in the process description. Structural uncertainty can also be quantified by testing different process formulations in one modeling framework [e.g., *Knorr and Heimann*, 2001]. Uncertainty in model simulations arises also from uncertainty in the parameter set in the mathematical descriptions of individual processes. Estimates of this parameter-based uncertainty are usually limited to a few parameters because of computational constraints. *Murphy et al.* [2004] report on a systematic attempt to determine the range of climate sensitivity of an atmosphere-mixed layer ocean GCM on the basis of a 53-member ensemble of model versions constructed by varying model parameters. The ''climate*prediction*.net'' experiment analyzes uncertainty in climate sensitivity by a perturbed physics multi-thousand-member grand ensemble of simulations using a general circulation model [*Stainforth et al.*, 2005]. For both experiments, model parameters are set to alternative values considered plausible by experts in the relevant parameterization schemes.

[6] So far, only a few studies have investigated the effects of propagating parameter uncertainty in global terrestrial ecosystem models. *Knorr* [2000] and *Knorr and Heimann* [2001] have quantified the uncertainties in terrestrial CO_2 fluxes simulated by the terrestrial ecosystem model BETHY by varying one model parameter at a time around a standard value. In a more systematic approach, *Zaehle et al.* [2005] used a Monte Carlo type sampling strategy to identify functionally important parameters and simultaneously estimate the uncertainty range of modeled results from the Lund-Potsdam-Jena dynamic global vegetation model. While these studies only infer uncertainties in model results around prior parameter values, other studies have used inverse methods to optimize parameters in terrestrial ecosystem models from observations [*Wang et al.*, 2001; *Vukicevic et al.*, 2001; *Randerson et al.*, 2002; *Barrett*, 2002; *Knorr and Kattge*, 2005].

[7] The study of *Kaminski et al.* [2002] combines both: They optimized the controlling process parameters of the Simple Diagnostic Biosphere Model (SDBM) with respect to the seasonal cycle of atmospheric CO_2 concentrations using a variational data assimilation approach. The model was then run with these optimized parameters to predict some diagnostic quantities of interest such as net fluxes and net primary productivity. They also estimated an uncertainty range for the optimized parameters from the uncertainty in the observed concentration and, finally, uncertainty of the monthly net flux.

[8] The carbon cycle data assimilation system (CCDAS [*Scholze*, 2003; *Rayner et al.*, 2005b]) builds upon the study by *Kaminski et al.* [2002] with SDBM replaced by the Biosphere Energy Transfer Hydrology Scheme (BETHY [*Knorr and Heimann*, 2001]). BETHY is fully prognostic and can, if driving fields are available, be run for future scenarios. Therefore, knowledge about the current terrestrial carbon cycle can be applied to predict its evolution into the future as has been demonstrated by *Rayner et al.* [2005a]. The model can also be coupled to atmospheric general circulation models for climate change studies, and it constitutes the core of the terrestrial carbon cycle component of the coupled carbon cycle-climate ECHAM5-JSBACH model version [*Raddatz et al.*, 2007].

[9] Here, we briefly summarize modifications and updates of CCDAS since *Rayner et al.* [2005b]. The greater part of the paper, however, reports on propagation of the formal uncertainty estimates gained during the parameter optimization step to calculate uncertainties of the main predicted quantities such as the net terrestrial CO_2 flux and atmospheric CO_2 concentrations. We present results of this uncertainty projection for both diagnostic and prognostic calculations. Although there is a whole suite of results available from prognostic CCDAS simulations, we focus the discussion here on the prognostic uncertainties for surface net fluxes and atmospheric carbon dioxide.

2. Model Description

[10] The setup, data and models used in CCDAS have been described by *Scholze* [2003] and *Rayner et al.* [2005b], to which we refer for details. In brief, BETHY, the core CCDAS model, is a process-based model of the terrestrial biosphere [*Knorr*, 2000]. It simulates carbon assimilation and plant and soil respiration embedded within a full energy and water balance and phenology scheme. BETHY is a fully prognostic model, and is thus able to predict the future evolution of the terrestrial carbon cycle under a prescribed climate scenario. In the present study, BETHY is driven by observed climate data for the period 1979 to 2003 [*Nijssen et al.*, 2001] (which have been extended to the year 2003 (R. Schnur, personal communication, 2005)). It is run on a $2^\circ \times 2^\circ$ grid resolution. Global vegetation is mapped onto 13 plant functional types (PFT) based on *Wilson and Henderson-Sellers* [1985]. A grid cell can contain up to three different PFTs, with the amount specified by their fractional coverage.

[11] We use the full BETHY model to assimilate in a first step global monthly fields of the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) derived from satellite data for optimizing parameters controlling soil moisture and phenology. Second, a reduced version of BETHY, Carbon-BETHY, which has no phenology scheme and no water balance, is used to assimilate atmospheric CO_2 concentration observations for the period 1979 to 1999. This simplified form of the model uses the optimized leaf area index (LAI) and plant available soil moisture fields from the first assimilation step as input data for the second step (see Figure 1). The present study focuses on Carbon-BETHY and the second assimilation step. Control parameters affect the photosynthesis scheme, and both the autotrophic and heterotrophic respiration schemes. Carbon-BETHY is coupled to the atmospheric transport model TM2 [*Heimann*, 1995] for calculating atmospheric CO_2 concentrations. We use the Jacobian matrices generated by the adjoint form of the transport model driven with winds

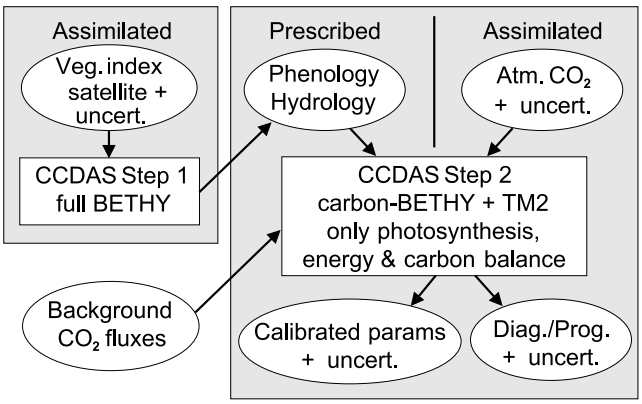

Figure 1. Sketch of the CCDAS setup: ovals represent input and output data, and boxes represent calculation steps. Background fluxes comprise ocean CO_2 fluxes, anthropogenic emissions and land use change.

from 1986 to efficiently represent the transport [*Kaminski et al.*, 1999a]. The parameters are listed with their a priori values and uncertainties in Table 1.

[12] CCDAS has helped to identify weaknesses in the formulation of the biosphere model and prompted a number of modifications. One of the modifications concerns the discretization scheme for the carbon cycling between the pools within BETHY leading to improved numerical behavior. The modified scheme led to slightly different results with a much more stable optimization procedure. We also identified one redundant parameter, namely the vegetation temperature dependence of the maximum electron transport ($a_{J,T}$), which only appeared in one equation as a product together with another parameter, namely the ratio of the maximum carboxylation rate to the maximum electron transport ($a_{J,V}$) [*Rayner et al.*, 2005b, equations (8) and (9)]. Both parameters can therefore not be resolved independently. Hence we removed one of them ($a_{J,T}$) from the list of parameters.

[13] For the purpose of this paper it is useful to recall the formulation of the carbon balance in CCDAS. The net ecosystem productivity (NEP) is defined as

$$\mathrm{NEP} = \mathrm{NPP} - R_\mathrm{S} = \mathrm{NPP} - \left(R_{\mathrm{S,s}} + R_{\mathrm{S,f}}\right), \tag{1}$$

where $R_{\mathrm{S,s}}$ and $R_{\mathrm{S,f}}$ are the respiration fluxes from the slowly and rapidly decomposing soil carbon pools, respectively, and NPP the net primary productivity. Soil respiration R_S is calculated from the size of the soil carbon pool (C_f for the fast and C_s for the slow pool), a constant standard turnover time (τ_f and τ_s, respectively), and a temperature and moisture dependent relative rate (k_f and k_s for the fast and slow pools respectively):

$$R_\mathrm{S} = (1 - f_\mathrm{s})\frac{k_\mathrm{f} C_\mathrm{f}}{\tau_\mathrm{f}} + \frac{k_\mathrm{s} C_\mathrm{s}}{\tau_\mathrm{s}}, \tag{2}$$

where f_s is the fraction of decomposition from the fast pool that goes to the slow soil carbon pool. The size of the slow soil carbon pool, C_s, is held constant through the simulation period as we will discuss below. The value of C_s for each grid cell and PFT is determined by a long-term balance constraint:

$$\overline{\mathrm{NPP}} = \beta\left(\overline{R_{\mathrm{S,s}} + R_{\mathrm{S,f}}}\right), \tag{3}$$

where the overlying bar denotes the average over the calibration period I_c. All of NPP goes into an intermediate living carbon pool, which is transferred to the fast soil carbon pool whenever litterfall occurs. Thus the time averaged flux into the fast soil carbon pool equals time-averaged NPP. Part of the decomposition of the fast carbon pool, i.e., $f_\mathrm{s} \cdot R_{\mathrm{S,f}}$ then goes into the slow carbon pool. However, because the slow carbon pool is large compared to both its decomposition and its production, changes in its size are neglected. Rather, the size of the pool is determined by how far total soil respiration from a particular grid cell is out of equilibrium with total soil carbon production, which is $\overline{\mathrm{NPP}}$ on a long-term average. The PFT-specific scaling parameter β determines whether the corresponding PFT acts as a long-term source ($\beta < 1$) or sink ($\beta > 1$). Via equations (2) and (3), the ratio $C_\mathrm{s}/\tau_\mathrm{s}$ can then be expressed as

$$\frac{C_\mathrm{s}}{\tau_\mathrm{s}} = \frac{\overline{\mathrm{NPP}}/\beta - (1 - f_\mathrm{S})\overline{R_{\mathrm{s,f}}}}{\overline{k_\mathrm{s}}}. \tag{4}$$

As only the ratio, not C_s itself, enters equation (2), there is no need to compute the absolute value of C_s. Note that both C_s and τ_s are constants here.

[14] A constant C_s is a reasonable approximation, because the initial size of C_s is so large that the temporal variations and the overall change in the size over our study period would be relatively small. We use this approach instead of simulating C_s directly for a number of reasons elaborated by *Rayner et al.* [2005b].

[15] This approach also avoids spinning up the slow soil carbon pool, which can take centuries of simulation, thereby dominating the computational cost. Alternatively, as in the adjoint integration, the computational dominance of the spin up could be bypassed in a highly efficient manner as demonstrated by *Kaminski et al.* [2005].

[16] More importantly, the balance constraint circumvents various scientific problems concerning the historical evolution of the terrestrial carbon cycle, such as CO_2 fertilization, climate forcing and, most importantly, disturbance. Current knowledge about these factors is limited and especially the history of disturbance is highly uncertain and difficult to quantify. These unknown processes are then embedded within the balance constraint, in our case the β factor.

3. Methodology

3.1. Formulation of the Problem

[17] *Rayner et al.* [2005b] demonstrate the use of CCDAS in its calibration mode, which is a two-step procedure. First, the model is run over a particular time period, I_c (calibration period), and observations are used to constrain process parameters in the model. Technically, this calibration process solves an inverse or optimization problem, which is formulated as a minimization problem. The function to be minimized quantifies the misfit between

Table 1. Controlling Parameters and Their Initial and Optimized Values and Prior and Posterior Uncertainties[a]

Parameter Number	Parameter	Initial Value	Optimized Value	Prior Uncertainty	Posterior Uncertainty
1	V_{max}^{25}(TrEv)	60	57.6	20	19.4
2	V_{max}^{25}(TrDec)	90	108.5	20	19.3
3	V_{max}^{25}(TmpEv)	41	40.7	20	20.0
4	V_{max}^{25}(TmpDec)	35	51.2	20	19.9
5	V_{max}^{25}(EvCn)	29	26.0	20	18.8
6	V_{max}^{25}(DecCn)	53	119.1	20	20.0
7	V_{max}^{25}(EvShr)	52	130.4	20	19.9
8	V_{max}^{25}(DecShr)	160	137.0	20	18.4
9	V_{max}^{25}(C3Gr)	42	11.4	20	17.4
10	V_{max}^{25}(C4Gr)	8	0.4	20	5.5
11	V_{max}^{25}(Tund)	20	35.6	20	19.7
12	V_{max}^{25}(Wetl)	20	19.2	20	20.0
13	V_{max}^{25}(Crop)	117	95.6	20	17.7
14	$a_{J,V}$(TrEv)	1.96	1.92	5	5.0
15	$a_{J,V}$(TrDec)	1.99	1.99	5	5.0
16	$a_{J,V}$(TmpEv)	2.0	2.0	5	5.0
17	$a_{J,V}$(TmpDec)	2.0	2.0	5	5.0
18	$a_{J,V}$(EvCn)	1.79	1.79	5	5.0
19	$a_{J,V}$(DecCn)	1.79	1.82	5	5.0
20	$a_{J,V}$(EvShr)	1.96	1.97	5	5.0
21	$a_{J,V}$(DecShr)	1.66	1.66	5	5.0
22	$a_{J,V}$(C3Gr)	1.9	1.88	5	5.0
23	$a_{J,V}$(C4Gr)	14.0	9.9	5	2.0
24	$a_{J,V}$(Tund)	1.85	1.86	5	5.0
25	$a_{J,V}$(Wetl)	1.85	1.84	5	5.0
26	$a_{J,V}$(Crop)	1.88	1.92	5	5.0
27	α_q	0.28	0.34	5	4.9
28	α_i	0.04	0.04	5	5.0
29	K_C^{25}	460×10^{-6}	445×10^{-6}	5	4.9
30	K_O^{25}	0.33	0.33	5	5.0
31	$a_{\Gamma,T}$	1.7	1.45	5	4.9
32	E_{K_O}	35948	36218	5	5.0
33	E_{K_C}	59356	58637	5	4.9
34	$E_{V_{max}}$	58520	62045	5	4.8
35	E_k	50967	50592	5	5.0
36	E_{R_d}	45000	42023	5	4.9
37	$f_{R,leaf}$	0.4	0.26	25	7.9
38	$f_{R,growth}$	1.25	1.12	5	4.7
39	f_S	0.2	0.43	−0.1; +0.2	±0.1
40	κ	1.0	0.59	−0.9; +9.0	±0.09
41	$Q_{10,f}$	1.5	2.00	−0.5; +0.75	±0.2
42	$Q_{10,s}$	1.5	1.31	−0.5; +0.75	±0.1
43	τ_f	1.5	6.8	−1.0; +3.0	±1.5
44	β(TrEv)	1	1.33	25	20.3
45	β(TrDec)	1	1.01	25	24.0
46	β(TmpEv)	1	1.23	25	24.9
47	β(TmpDec)	1	2.55	25	24.9
48	β(EvCn)	1	0.74	25	6.4
49	β(DecCn)	1	1.77	25	24.9
50	β(EvShr)	1	1.21	25	24.7
51	β(DecShr)	1	0.20	25	6.1
52	β(C3Gr)	1	0.77	25	19.4
53	β(C4Gr)	1	0.76	25	19.1
54	β(Tund)	1	1.14	25	22.4
55	β(Wetl)	1	0.56	25	24.9
56	β(Crop)	1	3.26	25	24.9
57	offset	338	336.3	0.3	0.19

[a]Units are as follows: V_{max}, μmol(CO_2)m^{-2} s^{-1}; $a_{\Gamma,T}$, μmol(CO_2)mol(air)$^{-1}$(°C)$^{-1}$; activation energies E, J/mol; τ_f, years; offset, ppmv; all others unitless. Uncertainties are in percentage except for lognormally distributed parameters for which a range is given. Uncertainties represent one standard deviation.

model simulation and observations plus the deviation of the parameter values from prior information (as specified in Table 1). This minimization is controlled by a gradient algorithm, which searches the parameter space by iterative evaluation of the function and its gradient with respect to the parameters. The gradient information is provided efficiently by the model's adjoint. At the minimum, an uncertainty of the estimated parameter set that is consistent with assumed observational and model uncertainties is approximated by the inverse of the function's Hessian matrix, evaluated for the optimal parameter set. The calibration process, hence, delivers a set of optimized parameters, together with their uncertainties.

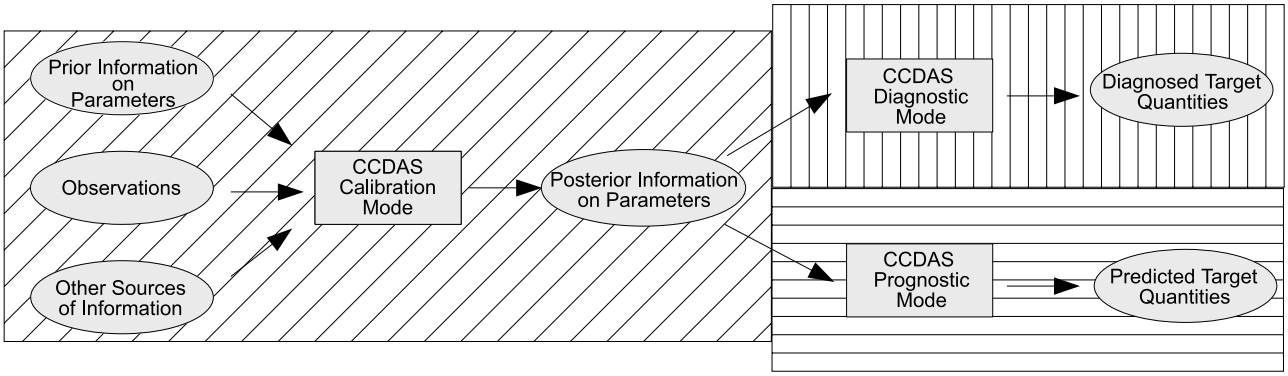

Figure 2. Schematic overview of two-step procedure for inferring diagnostic and prognostic target quantities from CCDAS. Rectangular boxes denote processes, and oval boxes denote data. The diagonally hatched box includes the inversion or calibration step, the vertically hatched box includes the diagnostic step, and the horizontally hatched box includes the prognostic step.

[18] The calibrated model can then be used for both diagnostic and prognostic simulations. By diagnostic we mean simulations over the calibration period; by prognostic simulations we understand simulations for a period subsequent to the calibration period. Figure 2 shows a flow diagram of the two different pathways for calculating diagnostic and prognostic quantities.

[19] *Rayner et al.* [2005b] also demonstrate how to project the parameter uncertainties forward through diagnostic simulations. The uncertainty in a vector of simulated quantities of interest (target quantity) $y(x)$ is approximated to first order by:

$$C(y) = \frac{dy}{dx}(x_{\text{opt}})\,C(x_{\text{opt}})\,\frac{dy}{dx}(x_{\text{opt}})^T + C_{\text{mod}}(y), \tag{5}$$

where $(\cdot)^T$ denotes the transposed, $C_{\text{mod}}(y)$ the uncertainty reflecting the imperfection of the model (model uncertainty) and $\frac{dy}{dx}$ the Jacobian matrix, representing the linearization of $y(x)$ around x_{opt}. If $y(x)$ were linear, equation (5) would hold exactly. In case the target quantity belongs to our set of observations, we use exclusively the first term to quantify the diagnostic uncertainty. In this case, the posterior parameter uncertainty derived by the inversion is consistent with prior uncertainty, observational and model uncertainty and reflects the state of information on the joint parameter $\times$ concentration space. In particular it incorporates the uncertainty in modeling the target quantity. This additional information is missing for target quantities that do not belong to our set of observations, which is reflected by the second term in equation (5).

[20] Often it is instructive to compare this posterior uncertainty with the prior uncertainty, which we compute by projecting the prior parameter uncertainty $C(x_{\text{pr}})$ forward to y_{pr} with the Jacobian that linearizes the model around the prior x_{pr}:

$$C(y_{\text{pr}}) = \frac{dy}{dx}(x_{\text{pr}})\,C(x_{\text{pr}})\,\frac{dy}{dx}(x_{\text{pr}})^T + C_{\text{mod}}(y_{\text{pr}}). \tag{6}$$

The reduction in uncertainty quantifies the gain in information owing to the observations.

[21] CCDAS makes considerable use of derivative code, i.e., the adjoint code to provide the gradient of the misfit function, the Hessian code to approximate parameter uncertainties, and the Jacobian code to propagate these uncertainties forward. All this derivative code is generated directly from the model's source code by the automatic differentiation (AD) tool Transformation of Algorithms in Fortran (TAF [*Giering and Kaminski*, 1998]).

[22] This paper describes and applies a procedure for projecting the parameter uncertainties forward through prognostic simulations. In diagnostic simulations the balance constraint (equation (3)) is implicitly taken care of as equation (4) is part of the diagnostic integration over I_c. C_s is then only an intermediate result on the way to the diagnostic quantities of interest. For prognostic simulations, however, the balance constraint for the slow pool needs some extra care. Depending on the length of the prognostic simulation, there are two cases:

[23] 1. We use the ratio C_s/τ_s computed by the calibration procedure (see equation (4)). We would do this, for instance, to predict a short period directly following the calibration period, e.g., calibrate from 1979 to 1999 and then predict 2000 to 2003. This approximation only introduces a small error, as C_s is large compared to its changes, and τ_s is a constant by definition.

[24] 2. We do not use the constant value for C_s/τ_s from the calibration period but simulate C_s dynamically. This would be the case when it comes to predicting the biospheric response to some prescribed climate forcing for some extended time period such as a century.

[25] In the following we first address case 1 and then generalize to case 2.

3.2. Approach

[26] Let y be the target quantity, e.g., some net flux over a region and (part of) the prognostic period. Let us further define

$$\tilde{C}_s = \frac{C_s}{\tau_s}. \tag{7}$$

[27] As argued above, y depends on the model parameters x and on $\tilde{C}_s$, which in turn also depends on x, i.e.,

$$y = y(\tilde{C}_s(x), x) = g(x), \tag{8}$$

where g is introduced for later usage. Computing y with the calibrated model is, formally, an evaluation of equation (8) with $x = x_{opt}$.

[28] The uncertainties of y are more difficult to compute. This is because $C(y)$ depends on the uncertainties in the parameters and on the uncertainties in $\tilde{C}_s$. The uncertainty in $\tilde{C}_s$, in turn, depends on the uncertainty in parameters via the optimization procedure, i.e., to first order:

$$C(\tilde{C}_s) = \frac{d\tilde{C}_s}{dx}(x_{opt}) C(x_{opt}) \frac{d\tilde{C}_s}{dx}(x_{opt})^T + C_{mod}(\tilde{C}_s), \tag{9}$$

Hence, the uncertainties in $\tilde{C}_s$ and in the parameters are correlated. Furthermore, the uncertainty in $\tilde{C}_s$ is quantified by a very large matrix, as $\tilde{C}_s$ has the dimension of grid cells times up to three PFTs, and its uncertainty has this dimension squared.

[29] Fortunately, we can avoid computing this uninteresting intermediate uncertainty. As our objective is the computation of the uncertainty in y, all we need is the Jacobian $\frac{dg}{dx}$, with

$$C(y) = \frac{dg}{dx}(x_{opt}) C(x_{opt}) \frac{dg}{dx}(x_{opt})^T + C_{mod}(y). \tag{10}$$

$\frac{dg}{dx}$ can be computed by applying the chain rule to equation (8):

$$\frac{dg}{dx} = \frac{\partial y}{\partial \tilde{C}_s} \frac{\partial \tilde{C}_s}{\partial x} + \frac{\partial y}{\partial x}. \tag{11}$$

[30] Application of first equation (11) and then equation (10) can be implemented with little extra work as described in Appendix A. The required extra derivative code is generated by the AD-tool TAF. Section 5 will also discuss the prior uncertainty, which is computed in a similar fashion via:

$$C(y_{pr}) = \frac{dg}{dx}(x_{pr}) C(x_{pr}) \frac{dg}{dx}(x_{pr})^T + C_{mod}(y_{pr}). \tag{12}$$

[31] Let us now address the case (see section 3.1) in which changes in $\tilde{C}_s$ through the prognostic period cannot be neglected, i.e., the prognostic period is too long to keep $\tilde{C}_s$ constant. In this case, we need to specify a value for τ_s, which is no longer eliminated from the model through the use of $\tilde{C}_s$. Once τ_s has been estimated, however, the value of C_s can be used as an initial size of the slowly decomposing carbon pool in a prognostic dynamic integration as demonstrated by *Rayner et al.* [2005a].

[32] For the uncertainty propagation, it is convenient to write down the functional form of a target quantity y:

$$y = y(C_s(\tilde{C}_s(x), \tau_s), x, \tau_s) = g(x, \tau_s). \tag{13}$$

The Jacobian of g required for the uncertainty propagation is decomposed via the chain rule into sub Jacobians, and evaluated for the posterior parameter vector and a prescribed estimate of τ_s, $\tau_{s,pr}$. The new Jacobians $dC_s/d\tilde{C}_s$ and $dC_s/d\tau_s$ are simple linearizations of equation (7). For propagating the uncertainty, $C(x_{opt})$ is extended by one row and column with $\sigma(\tau_{s,pr})^2$ on its diagonal and zero elsewhere.

4. Experiments

[33] As mentioned in section 2 the formulation of the model has evolved since the study by *Rayner et al.* [2005b]. This reformulation of the model allowed us to relax the prior uncertainties compared to *Rayner et al.* [2005b]. In general, we increased prior uncertainties by a factor of ten, but kept the prior uncertainties for β the same (they already represented realistic values). However, we did not change the prior parameter values themselves.

[34] Hence we repeated the calibration of the model for the same calibration period I_c from 1 January 1979 to 31 December 1999. For this new model calibration, we also updated one of the background fluxes, namely the fossil fuel emissions. We now use the flux magnitudes from *Marland et al.* [2006] for the years 1979 to 2003. We use two different patterns for spatially distributing the fossil fuel emissions corresponding to the years 1990 and 1995. The 1990 pattern is taken from the data of *Andres et al.* [1996] and the 1995 pattern from *Brenkert* [1998]. 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 two remaining background fluxes from land use change and ocean carbon uptake are as in the work by *Rayner et al.* [2005b]; for the prognostic period (2000 to 2003), we use flux pattern and magnitude from *Takahashi et al.* [1999] to describe the ocean flux climatology (both annual mean and seasonal cycle), for the years 2001–2003 we kept the 2000 land use change flux magnitude (2.1 GtC/yr), the last year in the *Houghton* [2003] data set. We use the same observational network as in the work by *Rayner et al.* [2005b], but updated the atmospheric CO_2 concentration data set against which the model is optimized to a newer release of the GLOBALVIEW database [*GLOBALVIEW-CO_2*, 2004]. This allows us to verify prognostic CCDAS results with observations. As in the work by *Rayner et al.* [2005b] our inversion procedure uses an uncorrelated uniform uncertainty of 0.5 ppmv to reflect imperfections of the model. This uncertainty refers to the entire modeling chain, i.e., it quantifies the combined effect of imperfections in both the terrestrial and transport models as well as the background fluxes. For the inversion (model calibration) step, the partitioning between these three does not matter. For the computation of diagnostic and prognostic flux uncertainties we need to specify the uncertainty contribution from imperfection in the terrestrial model. For convenience we specify this uncertainty as zero, i.e., we assume that, for the concentration, the model uncertainty is due to imperfect transport and background. The uncertainty in the background fluxes incorporates the uncertainty in both the flux magnitudes and their spatiotemporal patterns. The uncer-

Table 2. Prior and Posterior Mean Annual Regional Net CO_2 Flux and Posterior Uncertainty Over 1980 to 1999 and Its Relative Uncertainty Reduction in Percentage[a]

Region	Prior Flux	Posterior Flux	Posterior Uncertainty	Uncertainty Reduction
Australia	34.2	−213.0	158.8	12
Brazil	47.4	547.3	232.4	3
China	20.6	469.8	131.3	43
Europe	16.7	392.4	204.4	51
India	8.1	473.3	99.2	38
North America	39.9	293.4	196.5	38
Russia	25.4	−238.8	100.1	44

[a]Unit is MtC y^{-1}.

tainty in the terrestrial model reflects errors in the driving fields, which includes the fields provided to Carbon BETHY by the full model (see section 2). In addition, it reflects the error from modeling the global vegetation with only 13 PFTs. Formally, this aggregation error in PFT space is to be treated similar to the aggregation error in horizontal space described and quantified by *Kaminski et al.* [2001].

[35] We then use the calibrated model to predict the 4-year period I_p from 1 January 2000 to 31 December 2003, including the propagation of uncertainties described in section 3.2. For this first demonstration of the method, we deliberately choose a prognostic period in the past (i.e., we perform a so-called hindcasting experiment), which has two advantages. First, we can drive our model with observed climate data and, second, we can validate the model prediction with observations.

5. Results

5.1. Parameters

[36] The calibrated parameters are shown in Table 1, together with their uncertainties. We see that for parameters pertaining to the photosynthesis (or gross primary productivity, GPP) part of BETHY (parameters 1–36 in Table 1), especially for some of the spatially explicit parameters controlling the maximum carboxylation rate V_{max}, the posterior values have shifted more than two standard deviations from their prior values. For woody PFTs, the optimized V_{max} values generally are larger than the prior values, thus increasing the productivity of these PFTs, whereas for the herbaceous PFTs (C_3 and C_4 grass) the productivity is reduced via smaller posterior V_{max} values. However, for most of these parameters there is no or only little reduction in the uncertainty reflecting the weak constraint of the atmospheric concentration data on the photosynthesis (as explained by *Rayner et al.* [2005b]). For C_4 grass the case is different, as we see both a substantial shift in the posterior parameter values and a large uncertainty reduction. This is because for C_4 plants, V_{max} and $a_{J,V}$ control photosynthesis more directly than for C_3 plants (we refer here to *Rayner et al.* [2005b, equation (12)]).

[37] One of the two parameters controlling autotrophic respiration, $f_{R,leaf}$, shows a very similar behavior as in the work by *Rayner et al.* [2005b], with almost the same posterior value even though its prior uncertainty here is larger by a factor of five. The robustness of this result is further supported by the relatively large uncertainty reduction (posterior uncertainty of 7.9% versus a prior uncertainty of 25%), indicating that the parameter is well observed by our network. Recall that the smaller value of $f_{R,leaf}$ leads to a larger maintenance respiration, and thus changes the relationship between GPP and NPP. The change in $f_{R,growth}$, however, is opposite to the result of *Rayner et al.* [2005b]. Here, the posterior value is slightly smaller than the prior value leading to a smaller growth respiration. This smaller value caused a shift in the relationship between GPP and NPP in the reverse direction compared to the change in $f_{R,leaf}$. However, note that for $f_{R,growth}$, the uncertainty reduction is rather small.

[38] As in the work by *Rayner et al.* [2005b], many of the global parameters controlling the soil respiration model in BETHY show large reductions in uncertainty. Also, posterior parameter values obtained here have changed from their priors mostly in the same direction as the ones in the work by *Rayner et al.* [2005b], but by a somewhat smaller degree. For example, κ, regulating the moisture dependance of soil respiration, is only reduced to a value of 0.59 here as compared to 0.29 in the work by *Rayner et al.* [2005b]. A value less than one reduces the sensitivity of respiration to changes in soil moisture for moisture values near one but increases the sensitivity for values near zero. This nonlinear dependence is less pronounced in this study for the optimized parameter values. A noteworthy exception of the similarities with *Rayner et al.* [2005b] is the calibrated value of τ_f, the turnover time of the fast soil carbon pool: 6.8 years here compared to 1.62 years. However, note also that here, we use a much higher prior uncertainty for τ_f such that the shift relative to its prior uncertainty is not remarkably different from *Rayner et al.* [2005b]. However, this high turnover time together with the low $Q_{10,f}$ value reduces both the seasonality and the interannual variability of the heterotrophic respiration flux. In a further experiment [*Scholze et al.*, 2005], in which we used a set of spatially explicit biomass burning flux patterns as additional basis functions for the optimization, the optimized parameter values controlling the fast soil respiration, especially τ_f, are much closer to their prior values.

[39] The β parameter, as described in section 2, determines whether a PFT acts as a CO_2 source or sink. Together with NPP, it also defines the net carbon flux. In contrast to *Rayner et al.* [2005b], only two of the 13 β parameters show considerable reduction in uncertainty. In total, we can constrain about 12 directions in parameter space which is similar to the 14 constrained directions in our previous study [*Rayner et al.*, 2005b]. The sink patterns are generally similar between the two studies: large β values for tropical PFTs suggesting a substantial tropical sink as well as the midlatitude uptake due to the high values for β(TmpDec) and β(Crop). This is not surprising as the CO_2 sink/source distribution directly controls the net CO_2 exchange with the atmosphere. As in the work by *Rayner et al.* [2005b], we see an extreme shift for β(Crop) along with no reduction in its uncertainty.

5.2. Fluxes and Concentrations

[40] For our regional analysis we focus on seven regions. Table 2 lists the regions and their prior and diagnostic posterior mean annual flux values for the calibration period, together with the posterior uncertainty and reduction in

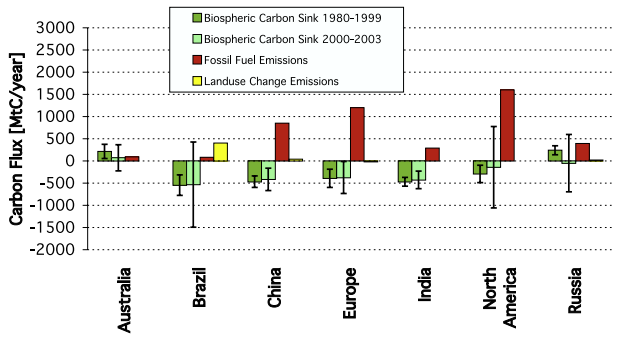

Figure 3. Diagnostic (dark green) and prognostic (light green) mean annual BETHY fluxes to the atmosphere over seven regions. Black bars indicate ±1 sigma uncertainty range. Fluxes from fossil emissions are shown in red, and fluxes from land use change are shown in yellow.

uncertainty through the calibration process. In general, our chosen network (the same as in the works by *Kaminski et al.* [2002] and *Rayner et al.* [2005b]) observes regions in the northern hemisphere better (reduction between 38 and 51%) than regions in the southern hemisphere (reduction between 3 and 12%).

[41] Figure 3 shows diagnostic (dark green) and prognostic (light green) mean annual BETHY fluxes to the atmosphere over the seven regions, together with their uncertainties as quantified by ±one standard deviation. Uncertainties for prognostic quantities are considerably higher.

[42] To understand the mechanisms that determine the uncertainties, it is instructive to study their temporal evolution. Figure 4 shows the evolution of the posterior uncertainties of the global annual net flux for both the diagnostic and the prognostic period. The posterior uncertainties for individual years are all between 0.1 to 0.4 GtC/yr. Compared to corresponding prior uncertainties between 7 and 8 GtC/yr (not shown in Figure 4) this is a large reduction by a factor of between 20 and 100 through the information from the atmospheric network. We can understand this qualitatively: The network was designed to monitor the global trend in the net carbon balance of the atmosphere. Our calibration transforms this into a constraint (quantified by an uncertainty reduction) for that direction in parameter space, which projects well onto this same global trend. It is, therefore, not surprising that this direction also projects well on the set of global annual net fluxes, and our diagnostic and prognostic uncertainty propagation yields a large uncertainty reduction. This is consistent with findings of *Kaminski et al.* [1999b], who analyzed the singular value spectrum for the Jacobian of the same transport model and a similar observational network. They found that the third largest singular value was associated with a singular vector in flux space that projects well on the global net flux.

[43] It is remarkable that years with similar net flux tend to have similar uncertainties. This holds, for instance, for the set of years 1985, 1986, 1996, 1997, and also for the set of years 1999, 2000. For a discussion of this observation we need to recall from section 2 the factors that can change the simulated fluxes from one year to the next. As the parameters are fixed for the entire integration period, the remaining factors are initial (pool sizes) and boundary (climate forcing) conditions. As the slow pool size is constant, the remaining initial condition is the size of the fast pool. Assuming, for a moment, the idealized situation that the climate forcing was perpetual (the same every year), and the fast pool had been spun up (i.e., is in steady state), the model would simulate the same flux every year. For propagating parameter uncertainties forward to fluxes, we use the Jacobian df/dx that expresses the sensitivity of the fluxes with respect to a change of parameters. In our idealized setting, this Jacobian would not change from year to year, and our uncertainty propagation would, thus, yield the same uncertainty for every annual net flux. In our actual setting, a set of years could at least have climate forcings that are similar in the sense that they yield similar functions from parameters to the respective global annual net fluxes. This would then yield similar fluxes and similar uncertainties. Of course, the direction of this argument may not be reversed: There are many climate forcings that yield similar fluxes without yielding similar uncertainties. We also need to keep in mind that the back propagation of the atmospheric uncertainty via the inverse Hessian uses second-order derivative information, while our forward propagation uses only the linear term quantified by the Jacobian.

[44] Figure 5 displays the Jacobians that quantify the sensitivity of global annual net fluxes to parameters, where parameters are measured in multiples of the prior sigmas (see Table 1). Prior (red) and posterior (black) Jacobians are shown. The two panels condense the information content in this 25 × 57 Jacobian matrix differently. Figure 5a displays the average over rows, i.e., average over the annual global net fluxes of the entire period. The mean annual net flux is most sensitive to changes in parameter 39 (f_s in equation (4)) in both the prior and the posterior Jacobians: An increase of the parameter value by one sigma reduces the prior and posterior net uptakes by 10 GtC and 15 GtC/yr, respectively. Figure 5b shows all rows but only the 39th column; that is, it shows the sensitivity of the annual net flux for each of the 25 years with respect to parameter 39. For both the prior and posterior Jacobians, this sensitivity is almost constant from year to year.

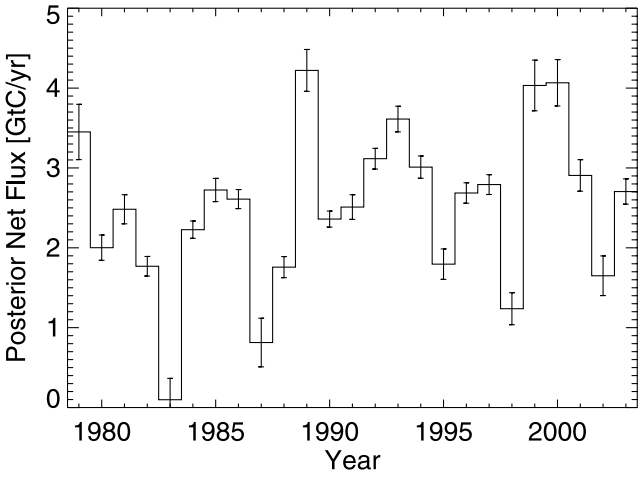

Figure 4. Posterior flux estimates for the global annual net flux to the biosphere, together with ±one sigma uncertainty ranges.

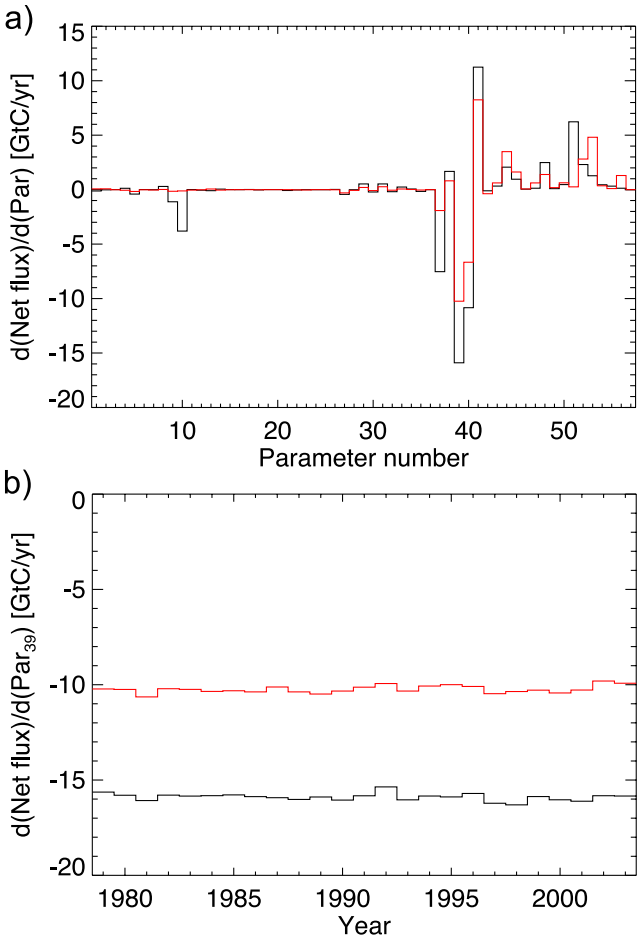

Figure 5. (a) Prior (red) and posterior (black) Jacobians averaged over rows, i.e., expressing the average global annual net flux sensitivity with respect to parameters in GtC/yr, as parameter units are multiples of prior sigmas. (b) Prior (red) and posterior (black) Jacobians expressing the sensitivity of the 25 global annual net fluxes from 1999 to 2003 with respect to parameter 39.

[45] A familiar effect in transport inversions is negatively correlated posterior uncertainties for fluxes in neighboring grid cells and fluxes in subsequent time periods over the same grid cell. The atmospheric network constrains regional and temporal sums much better than differences. For example, underestimating the flux at a particular time and grid cell can be compensated by overestimating the flux from a neighboring grid cell or the flux in the subsequent time period. In transport inversions, often uncorrelated prior uncertainties are chosen. Figure 6 shows the correlations for both prior (Figure 6a) and posterior (Figure 6c) estimates of global annual net fluxes. Unlike in the flux inversions, here we obtain positive flux correlations, all of which are above 0.997. To understand why this should be the case, let us return to our idealized setting of the spun up pools and perpetual climate forcing. For this idealized case, all rows in the Jacobian $df/dx(x_{pr})$ are the same, and evaluation of equation (6) yields a correlation of 1 for any two components. The fact that we selected zero correlation for prior parameter uncertainties simplifies the evaluation, but any

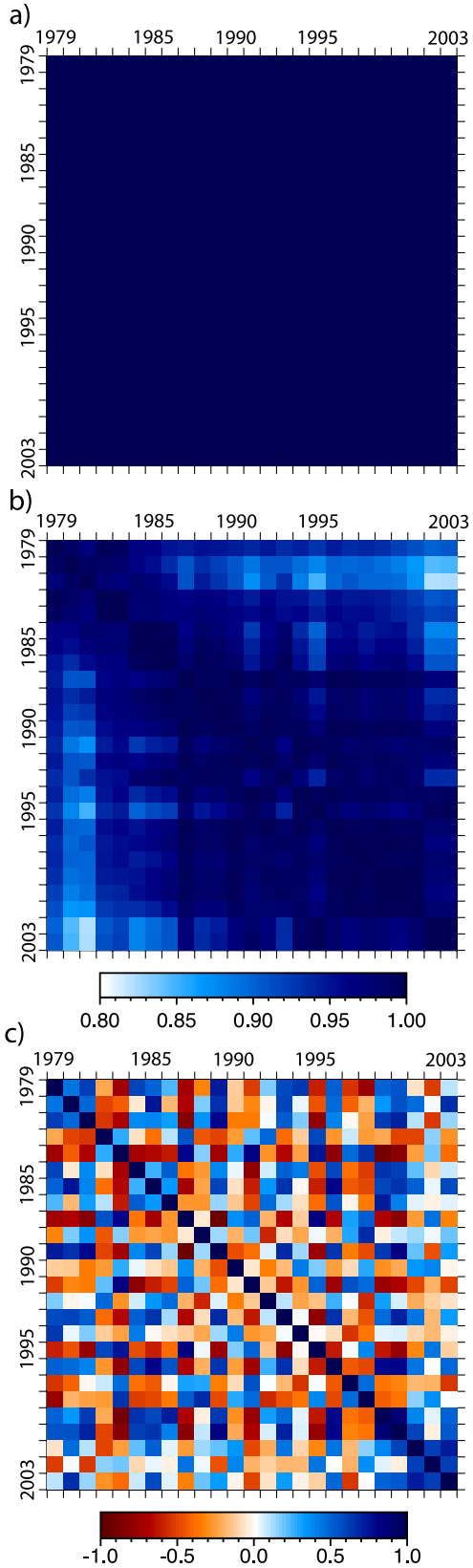

Figure 6. Correlations of uncertainties in global annual net fluxes based on (a) prior parameter values with prior uncertainties, (b) posterior parameter values with prior uncertainties, and (c) posterior parameter values with posterior uncertainties.

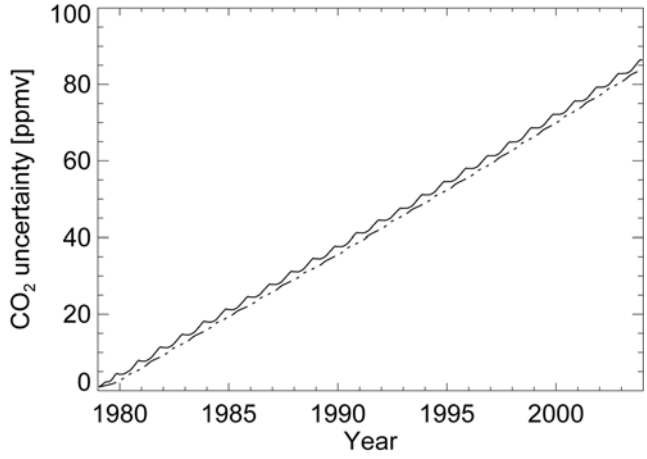

Figure 7. Uncertainties at Mauna Loa (solid line) and South Pole (dash-dotted line) corresponding to prior parameter uncertainties.

other choice would yield the same result. This is particularly relevant for the posterior case, which clearly has nonzero correlations for parameter uncertainties.

[46] The high correlation of prior flux uncertainties reflects the previously discussed similarity between the rows in the prior Jacobian (corresponding to a different year each). The slight deviations from the perfect correlation reflect the differences in the Jacobian. In the posterior case shown in Figure 6c, the observational information has, obviously, imposed a considerable change in the correlation structure of the parameter uncertainties. Some of this can also be attributed to the change in the linearization point, as is illustrated by Figure 6b, which, as a test, shows the result of propagating prior parameter uncertainties with the posterior Jacobian. We defer a further discussion of posterior flux uncertainties until we have looked at uncertainties of simulated concentrations.

[47] As prominent examples, we have selected time series covering both calibration and prognostic periods at two observational sites, namely Mauna Loa and South Pole. To propagate uncertainties from parameters to concentrations, we apply equation (6) and equation (12), for the prior case with concentrations c_{pr} taking the role of the target quantity y. Figure 7 shows the resulting uncertainties at both sites. To understand the steady increase in uncertainty by roughly 3.5 ppmv/yr, we build a conceptual one box model with an annual time step into which annual net fluxes described by a vector $f = (f_1, f_2, \ldots, f_N)$ with uncertainties $C(f)$ are emitted over N years. Obviously, the response of the concentration is then an increase by kf_1 in the first year, kf_2 in the second and so on, where k is the constant that converts emissions into concentrations. The concentration change after N years of emissions, Δc_N, is then given by

$$\Delta c_N = (k, \ldots, k) \cdot \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix}; \tag{14}$$

that is, the Jacobian of this transport model is $d(\Delta c_N)/df = (k, \ldots, k)$. As shown in Figure 6 prior uncertainties in global annual net fluxes are almost perfectly correlated in time, which yields

$$\sigma(\Delta c_N)^2 = k^2 \sum_{i,j=1,N} \sigma(f_{\mathrm{pr},i})\sigma(f_{\mathrm{pr},j}) \tag{15}$$

One sigma ranges for prior flux uncertainties are 7–8 GtC/yr, and TM2 uses a conversion of k = 0,476 ppmv/GtC. This yields a slope between 3.33 and 3.81 ppmv/yr, which is consistent with Figure 7.

[48] Figure 8 shows the posterior uncertainties at both sites. Over the diagnostic period, they remain in a one-sigma range of about 0.06–0.21 ppmv. At the start of the prognostic period, the uncertainty increases by our assumed model uncertainty of 0.5 ppmv, which is to be added when switching to nonobserved target quantities (see section 4). During the prognostic period, the uncertainty increases by about 0.1 ppmv per year. The action of an atmospheric data item on the parameter space is such that it narrows the uncertainty in the parameter direction that projects onto the observed direction in concentration space. We note also that, in a well-behaved system, the posterior uncertainty on an observation should not exceed the prior uncertainty for the same observation. For the fluxes over the diagnostic period, we see the two effects we mentioned above: First, annual net flux uncertainties are reduced to 0.1–0.4 GtC/yr (see Figure 4), much less than the corresponding prior values between 7–8 GtC/yr. The parameter directions constrained by the observations obviously project well on the 21 flux space directions corresponding to the sequence of global annual net fluxes in the diagnostic period. Second, strong negative correlations among uncertainties of many flux pairs are obtained (see upper left 21 by 21 submatrix in Figure 6c. Without these negative correlations in flux space, posterior uncertainty in concentration space would accumulate and observational uncertainties could not be reduced as required.

[49] Figure 4 also shows that prognostic annual net flux uncertainties are in the same range as their diagnostic counterparts. This means that the projections of constrained

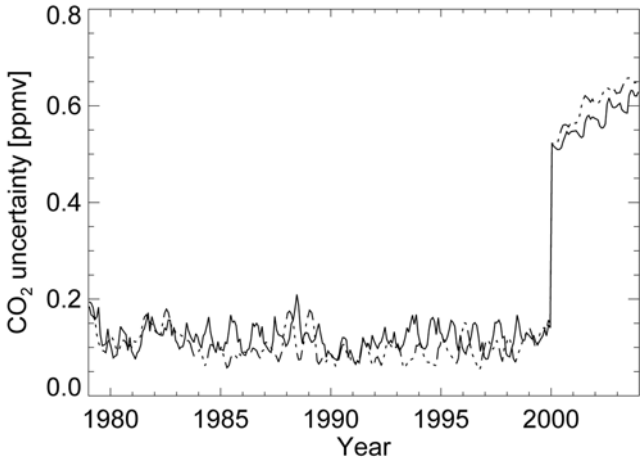

Figure 8. Uncertainties at Mauna Loa (solid line) and South Pole (dash-dotted line) corresponding to posterior parameter uncertainties.

Table 3. Comparison of Predicted and Observed Concentrations for the Prognostic Time Period (2000–2003, a Total of 48 Data Points)[a]

Station Code	Descriptive Name	RMS	Bias	Bias Corrected RMS
BRW	Barrow, Alaska	1.7	−0.6	1.59
MLO	Mauna Loa, Hawaii	0.6	−0.32	0.51
ASC	Ascension Island	0.82	−0.57	0.59
EIC	Easter Island	0.71	0.22	0.68
SPO	South Pole	0.9	−0.87	0.21

[a]Unit for columns 3–5 is ppmv.

parameter directions onto prognostic annual net fluxes are similar to projections onto diagnostic annual net fluxes. At the same time, prognostic concentration uncertainties increase from year to year (see Figure 8), reflecting the positive correlations for the prognostic flux uncertainties shown in Figure 6c. Again we can use the simplified transport model to check the plausibility of the slope for the prognostic uncertainty. With an average correlation coefficient of $\alpha_{i,j} = 0.7$ (estimated from Figure 6c) the equation

$$\sigma(\Delta c_N)^2 = k^2 \sum_{i,j=1,N} \alpha_{i,j}\sigma(f_i)\sigma(f_j) \tag{16}$$

yields a slope between 0.03 and 0.13 ppmv/yr, which is consistent with the slope for the prognostic period in Figure 8 of about 0.1 ppmv year.

[50] The mechanism discussed above, which keeps the uncertainty within the bounds of the prognostic annual net fluxes (and, hence, the concentration increments from one year to the next low) does not work for the concentration itself. We can understand this by decomposing the concentration Jacobian with respect to parameters:

$$\frac{dc_i}{dx} = k\frac{df_i}{dx} + \frac{dc_{I-1}}{dx}. \tag{17}$$

While the Jacobian for the increment is similar with those in the diagnostic period, the Jacobian for last year's concentration does change from year to year, simply because the atmosphere integrates the effect of all past fluxes.

[51] To validate our hindcast, we use again the GLOBALVIEW database [*GLOBALVIEW-CO_2*, 2004]. Table 3 compares model predicted monthly mean concentration from 2000 to 2003 with observations at five sites. We compute root mean squared (RMS) difference between model prediction and observations (column three) for each station separately. Part of this RMS difference is due to a bias, i.e., the mean of the model-observation difference (column four) is not zero. The RMS difference with this bias removed is shown in column five. The magnitude of the bias is not surprising given that we ran our prognostic simulation with climatological background fluxes. For instance, an additional uniform annual emission of only 0.2 GtC would be enough to compensate for the bias at Mauna Loa. At the stations in the Northern hemisphere, Barrow and Mauna Loa, most of the RMS difference comes from slight deviations in the seasonal cycle; removing the bias has only a small effect in the RMS difference. By contrast, at South Pole with its small seasonal cycle most of the RMS difference is due to the bias.

[52] One can use the prognostic uncertainties of Figure 4, together with the observational uncertainty in the GLOBALVIEW record, to put the above values in context. Adding the squares of these two uncertainties yields the square of the uncertainty in the model-data difference for each observation. This uncertainty in the model-data difference is 0.74 ppmv at Mauna Loa and 0.61 ppmv at South Pole on average over the 48 monthly mean observations. This is not to be confused with the uncertainty in the bias, which is smaller by a factor of $1/\sqrt{47}$.

[53] Overall, this comparison indicates that our combination of prior parameter uncertainty and model uncertainty, together with the assumption of uncorrelated observational uncertainties is slightly optimistic.

6. Conclusions

[54] We have reported on updates on CCDAS, and the calibration of the updated model with more conservative prior parameter uncertainties than those used by *Rayner et al.* [2005b]. For some of the parameters pertaining to the photosynthesis scheme, the higher prior uncertainty allowed a considerable shift from their prior values; however, they are usually not very well constrained by the atmospheric observations. The turnover time of the fast soil carbon pool is also strongly increased from 1.5 years to 6.8 years. We believe that this rather large value may be due to an aliasing effect that enhances soil respiration because the model lacks the process of fire, since the value of the parameter effects the seasonality of the respiration. We will elaborate on this in a subsequent paper.

[55] Further, we have presented a methodological framework for propagating parameter uncertainties as inferred by the calibration process through prognostic CCDAS simulations that relies on the Jacobian that linearizes the calibrated model. We have demonstrated the concept in a hindcast experiment spanning the years 2000 to 2003 with terrestrial net CO_2 fluxes and atmospheric CO_2 concentrations as target quantities. Besides the uncertainties in parameters that are explicitly calibrated, our formulation also covers the uncertainty in quantities which are implicitly calibrated through a long-term balance constraint. In our case, this is the size of the slowly decomposing soil organic carbon pool, which is grid cell and PFT-specific and hence a high-dimensional field. This object is not a field of process parameters; in a dynamical formulation of the slow pool, it would be a state variable. Over the prognostic period, our demonstration yields posterior uncertainties on global annual net fluxes that remain in the same range as over the diagnostic period. Posterior uncertainties in prognostic concentrations increase from year to year.

[56] The demonstrated concept is useful in quantitative network design as described by *Kaminski and Rayner* [2007]. It allows one to optimize observational networks such that the uncertainty in prognostic target quantities is minimized.

[57] The methodological framework for uncertainty propagation is applicable to other components of the Earth

system or an entire Earth system model. Examples for further candidates for implicit calibration via balance constraints could be the state of slowly varying components such as the state of the deep ocean or the ice sheets.

Appendix A: Implementation

[58] For the uncertainty propagation to prognostic target quantities via equation (10) we need the Jacobian $\frac{dg}{dx}$. This Jacobian is composed of the three sub-Jacobians on the right hand side of equation (11). TAF generates efficient derivative code from the model's source code for evaluating the respective sub-Jacobians. The model code, as we have previously used it in calibration and diagnostic modes and presented by *Kaminski et al.* [2003] and *Rayner et al.* [2005b], requires only a few slight extensions and new TAF calls. For the computation of the posterior Jacobian $\frac{dg}{dx}(x_{\mathrm{opt}})$ we perform the following steps:

[59] 1. The calibration code is modified such that the C_{s} field is recorded.

[60] 2. The diagnostic/prognostic code is modified such that C_{s} is read in. We write diagnostic/prognostic because the same code is underlying both modes. For later convenience we also modify the code such that the remaining variables in equation (2) as well as NPP (in equation (1)) are recorded.

[61] 3. TAF is applied in vector tangent mode to generate $\frac{dC_{\mathrm{s}}}{dx}$ from the modified calibration code, and this Jacobian is evaluated for $x = x_{\mathrm{opt}}$.

[62] 4. TAF is applied in vector tangent or vector adjoint mode to generate $\frac{\partial y}{\partial x}$ from the modified diagnostic/prognostic code and this Jacobian is evaluated for $x = x_{\mathrm{opt}}$.

[63] 5. TAF is applied in vector adjoint mode to generate $\frac{\partial y}{\partial C_{\mathrm{s}}}$. Because of the previously recorded variables the underlying model code can by largely reduced, i.e., BETHY can be bypassed, and NEP can be directly computed from equation (2) and then equation (1). This Jacobian is evaluated for $C_{\mathrm{s}} = C_{\mathrm{s}}(x_{\mathrm{opt}})$.

[64] To compute the prior form of the Jacobian, i.e., $\frac{dg}{dx}(x_{\mathrm{pr}})$, we replace the linearization point x_{opt} by x_{pr} in the above instruction list.

[65] We could have computed the Jacobians for steps 4 and 5 simultaneously in a single run using code that simultaneously differentiates with respect to x and C_{s}. We have opted not to do so in order to benefit from the computational saving in step 5 thanks to the reduced model.

Notation

We use a notation very similar to that of *Rayner et al.* [2005b].

- x vector of parameter values.
- x_{pr} prior parameter values.
- x_{opt} optimal (posterior) parameter values.
- $C(x_{\mathrm{opt}})$ covariance of the associated uncertainty.
- y diagnostic/prognostic quantity of interest (target quantity), e.g., a net flux.
- $C(y)$ covariance of the associated uncertainty.
- $C_{\mathrm{mod}}(y)$ covariance of the uncertainty reflecting imperfection of the model.
- β beta factor, see section 2.
- NPP net primary productivity.
- NEP net ecosystem productivity: NEP = NPP $- R_{\mathrm{S,s}} - R_{\mathrm{S,f}}$, where $R_{\mathrm{S,s}}$ and $R_{\mathrm{S,f}}$ are respiration from slow and fast soil carbon pool, respectively.
- k_{s}, k_{f} temperature and moisture dependent rate multiplier for the slow and fast soil carbon pool, respectively.
- τ_{s}, τ_{f} constant standard turnover time for the slow and fast soil carbon pool, respectively.
- C_{s}, C_{f} slow and fast soil carbon pool size, respectively.
- I_{c} calibration period.
- I_{p} prognostic period.

[66] **Acknowledgments.** Many thanks to Thomas Lavergne and Bernard Pinty for their quick help with graphics software. FastOpt were supported in part by the European Community within the 6th Framework Programme for Research and Technological Development under contract 511176 and contract 026188. This work was partly supported by the QUEST programme of the Natural Environment Research Council, U.K.

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*(3), 419–429.

Barrett, D. J. (2002), Steady state turnover time of carbon in the Australian terrestrial biosphere, *Global Biogeochem. Cycles*, *16*(4), 1108, doi:10.1029/2002GB001860.

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 basis, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn. (Available at http://cdiac.esd.ornl.gov/ndps/ndp058a.html).

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell (2000), Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, *Nature*, *408*, 184–187.

Cox, P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford, and C. D. Jones (2004), Amazonian forest diaback under climate-carbon projections for the 21st century, *Theor. Appl. Climatol.*, *78*, 137–156.

Dufresne, J.-L., P. Friedlingstein, M. Berthelot, L. Bopp, P. Ciais, L. Fairhead, H. Le Treut, and P. Monfray (2002), On the magnitude of positive feedback between future climate change and the carbon cycle, *Geophys. Res. Lett.*, *29*(10), 1405, doi:10.1029/2001GL013777.

Friedlingstein, P., J.-L. Dufresne, P. M. Cox, and P. Rayner (2003), How positive is the feedback between climate change and the carbon cycle?, *Tellus, Ser. B*, *55*(2), 692–700, doi:10.1034/j.1600-0560.2003.01461.x.

Friedlingstein, P., et al. (2006), Climate-carbon cycle feedback analysis: results from the C^{4}MIP Model Intercomparison, *J. Clim.*, *19*, 3337–3353.

Fung, I., P. Rayner, and P. Friedlingstein (2000), Full-form earth system models: Coupled carbon-climate interaction experiment (the flying leap), *IGBP Global Change Newsl.*, *41*, 7–8.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Software*, *24*(4), 437–474, doi:http://doi.acm.org/10.1145/293686.293695.

GLOBALVIEW-CO_2 (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).

Heimann, M. (1995), The global atmospheric tracer model TM2, *Tech. Rep. 10*, Max-Planck-Inst. für Meteorol., Hamburg, Germany.

Houghton, R. A. (2003), Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, *Tellus, Ser. B*, *55*(2), 378–390.

Jones, C. D., and P. M. Cox (2005), On the significance of atmospheric CO_2 growth rate anomalies in 2002–2003, *Geophys. Res. Lett.*, *32*, L14816, doi:10.1029/2005GL023027.

Kaminski, T., and P. J. Rayner (2007), Assimilation and network design, in *Observing the Continental Scale Greenhouse Gas Balance of Europe*, edited by H. Dolman, A. Freibauer, and R. Valentini, Springer, New York, in press.

Kaminski, T., M. Heimann, and R. Giering (1999a), A coarse grid three dimensional global inverse model of the atmospheric transport: 1. Adjoint model and Jacobian matrix, *J. Geophys. Res.*, *104*(D15), 18,535–18,553.

Kaminski, T., M. Heimann, and R. Giering (1999b), 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.

Kaminski, T., P. Rayner, M. Heimann, and I. Enting (2001), On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*(D5), 4703–4716.

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—ICCSA 2003, International Conference Montreal, Canada, May 2003, Proceedings, Part II, Lecture Notes in Computer Science*, vol. 2668, edited by V. Kumar et al., pp. 95–104, Springer, Berlin.

Kaminski, T., R. Giering, and M. Voßbeck (2005), Efficient sensitivities for the spin-up phase, in *Automatic Differentiation: Applications, Theory, and Implementations*, edited by H. M. Bücker et al., pp. 283–291, Springer, New York.

Keeling, C. D., T. P. Whorf, M. Wahlen, and J. van der Plicht (1995), Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980, *Nature*, *375*, 666–670.

Knorr, W. (2000), Annual and interannual CO_2 exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, *Global Ecol. Biogeogr.*, *9*(3), 225–252.

Knorr, W., and M. Heimann (2001), Uncertainties in global terrestrial biosphere modeling: 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, *Global Biogeochem. Cycles*, *15*(1), 207–225.

Knorr, W., and J. Kattge (2005), Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using Monte Carlo sampling, *Global Change Biol.*, *11*, 1333–1351.

Marland, G., T. A. Boden, and R. J. Andres (2006), Global, regional, and national CO_2 emissions, in *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, *430*, 768–772.

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.

Prentice, I. C., et al. (2001), The carbon cycle and atmospheric carbon dioxide, in *Climate Change 2001: The Scientific basis*, edited by J. T. Houghton et al., chap. 3, pp. 183–237, Cambridge Univ. Press, Cambridge, U. K.

Raddatz, T. J., C. Reick, W. Knorr, J. Kattge, E. Roeckner, R. Schnur, K.-G. Schnitzler, P. Wetzel, and J. Jungclaus (2007), Will the tropical land biosphere dominate the climate carbon cycle feedback during the 21st century?, *Clim. Dyn.*, doi:10.1007/s00382-007-0247-8, in press.

Randerson, J. T., et al. (2002), Carbon isotope discrimination of arctic and boreal biomes inferred from remote atmospheric measurements and a biosphere-atmosphere model, *Global Biogeochem. Cycles*, *16*(3), 1028, doi:10.1029/2001GB001435.

Rayner, P., M. Scholze, P. Friedlingstein, J.-L. Dufresne, T. Kaminski, W. Knorr, R. Giering, and H. Widmann (2005a), The fate of terrestrial carbon under climate change: Results from a CCDAS, poster presented at 7th Carbon Dioxide Conference, Clim. Monit. and Diagn. Lab., NOAA, Broomfield, Colo.

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

Scholze, M. (2003), Model studies on the response of the terrestrial carbon cycle on climate change and variability, Examensarbeit, Max-Planck-Inst. für Meteorol., Hamburg, Germany.

Scholze, M., P. Rayner, J. Kattge, W. Knorr, C. Rödenbeck, T. Kaminski, R. Giering, and H. Widmann (2005), Inferring fluxes of biomass buring from a global carbon cycle data assimilation system (CCDAS), poster presented at 7th Carbon Dioxide Conference, Clim. Monit. and Diagn. Lab., NOAA, Broomfield, Colo.

Smith, S. J., and J. A. Edwards (2006), The economic implications of carbon cycle uncertainty, *Tellus, Ser. B*, *58*, 586–590.

Stainforth, D. A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*, 403–406.

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 pCO_2 difference, paper presented at 2nd International CO_2 in the Oceans Symposium, Cent. for Global and Environ. Res., Natl. Inst. for Environ. Stud., Tsukuba, Japan, 18 –22 Jan.

Vukicevic, T., M. Steyskal, and M. Hecht (2001), Properties of advection algorithms in the context of variational data assimilation, *Mon. Weather Rev.*, *129*(5), 1221–1231.

Wang, Y. P., R. Leuning, H. Cleugh, and P. A. Coppin (2001), Parameter estimation in surface exchange models using non-linear inversion: How many parameters can w e estimate and which measurements are most useful?, *Global Change Biol.*, *7*, 495–510.

Wilson, M. F., and A. Henderson-Sellers (1985), A global archive of land cover and soils data for use in general-circulation climate models, *J. Climatol.*, *5*(2), 119–143.

Zaehle, S., S. Sitch, B. Smith, and F. Hatterman (2005), Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics, *Global Biogeochem. Cycles*, *19*, GB3020, doi:10.1029/2004GB002395.

R. Giering and T. Kaminski, FastOpt, Schanzenstr. 36, D-20357 Hamburg, Germany. (ralf.giering@fastopt.com; thomas.kaminski@fastopt.com)

W. Knorr and M. Scholze, QUEST, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK. (wolfgang.knorr@bristol.ac.uk; marko.scholze@bristol.ac.uk)

P. Rayner, LSCE, CEA de Saclay, Orme des Merisiers, F-91191 Gif-sur-Yvette, France. (peter.rayner@cea.fr)