SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1029/2006GB002834

Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France — Source link

Diego Santaren, Philippe Peylin, Philippe Peylin, Nicolas Viovy ...+1 more authors

Institutions: French Alternative Energies and Atomic Energy Commission, Centre national de la recherche scientifique

Published on: 01 Jun 2007 - Global Biogeochemical Cycles (John Wiley & Sons, Ltd)

Topics: Eddy covariance, Sensible heat, Heat flux, Latent heat and Diurnal cycle

Related papers:

- A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system
- A Biochemical Model of Photosynthetic CO 2 Assimilation in Leaves of C 3 Species
- Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison
- Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations
- On the Separation of Net Ecosystem Exchange into Assimilation and Ecosystem Respiration: Review and Improved Algorithm

Share this paper:

View more about this paper here: https://typeset.io/papers/optimizing-a-process-based-ecosystem-model-with-eddy-4rnfapyjew

Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France

Diego Santaren, Philippe Peylin, Nicolas Viovy, Philippe Ciais

▶ **To cite this version:**

Diego Santaren, Philippe Peylin, Nicolas Viovy, Philippe Ciais. Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France. Global Biogeochemical Cycles, American Geophysical Union, 2007, 21 (2), pp.n/a-n/a. 10.1029/2006GB002834 . hal-02926764

HAL Id: hal-02926764

https://hal.archives-ouvertes.fr/hal-02926764

Submitted on 17 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France

Diego Santaren,[1] Philippe Peylin,[1,2] Nicolas Viovy,[1] and Philippe Ciais[1]

Received 1 September 2006; revised 18 December 2006; accepted 9 March 2007; published 22 May 2007.

[1] We design a Bayesian inversion method (gradient-based) to optimize the key functioning parameters of a process-driven land surface model (ORganizing Carbon and Hydrology In Dynamic EcosystEms (ORCHIDEE)) against the combination of prior information upon the parameters and eddy covariance fluxes. The model calculates energy, water, and CO_2 fluxes and their interactions on a half-hourly basis, and we carry out the inversion using measurements of CO_2, latent heat, and sensible heat fluxes as well as of net radiation over a pine forest in southern France. The inversion method makes it possible to assess the reduction of uncertainties and error correlations of the parameters. We designed an ensemble of inversions with different set ups using flux data over different time periods, in order to (1) identify well-constrained parameters and loosely constrained ones, (2) highlight some model structural deficiencies, and (3) quantify the overall information gained from assimilating each type of CO_2 or energy fluxes. The sensitivity of the optimal parameter values to the initial carbon pool sizes and prior parameter values is discussed and an analysis of the posterior uncertainties is performed. Assimilating 3 weeks of half-hourly flux data during the summer improves the fit to diurnal variations, but merely improves the fit to seasonal variations. Assimilating a full year of flux data also improves the fit to the diurnal cycle more than to the seasonal cycle. This points out to the key importance of timescales when inverting parameters from high-frequency eddy-covariance data. We show that photosynthetic parameters such as carboxylation rates are well-constrained by the carbon and water fluxes data and get increased from their prior values, a correction that is corroborated by independent measurements at leaf scale. In contrast, the parameters controlling maintenance, microbial and growth respirations, and their temperature dependencies cannot be robustly determined. The CO_2 flux data could not discriminate between the different respiration terms. At face value, all the parameters controlling the surface energy budget can be safely determined, leading to a good model-data fit on different timescales.

Citation: Santaren, D., P. Peylin, N. Viovy, and P. Ciais (2007), Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, *Global Biogeochem. Cycles*, *21*, GB2013, doi:10.1029/2006GB002834.

1. Introduction

[2] Large uncertainties pertain to the distribution and future evolution of terrestrial CO_2 sources and sinks and their controlling mechanisms. In particular, the key processes determining the response of CO_2 fluxes to changes in climate, atmospheric composition and management practice, need to be better quantified when it comes to account for carbon-climate feedbacks in coupled models for future predictions. The question is whether we can exclude any of those model's responses with current observations. At present, there is a global network of nearly 250 eddy covariance flux towers with quasicontinuous in situ observations of CO_2, H_2O and energy fluxes. These data are likely to be the best source of knowledge on terrestrial carbon processes. However, the small spatial footprint of each flux tower requires extrapolation (up scaling) techniques and models. In that process, the spatial heterogeneity of ecosystems or the effects of nonlinear processes on spatial parameter averaging makes it difficult to extrapolate small-scale process formulations to larger scales (regions, continents, globe...). Advanced biogeochemical models which encapsulate the processes determined at the flux tower level and can extrapolate them into regional fluxes appear as logical partners of eddy covariance flux networks.

[3] This paper focuses on a method to optimize processes within a simulation model so as to optimally fit observations of CO_2, H_2O, heat and net radiation fluxes. These four independent fluxes are intimately coupled to each other via the ecosystem functioning (CO_2 and water fluxes are linked

[1] Laboratoire des Sciences du Climat et de l'Environnement, Commissariat à l'Energie Atomique, L'Orme des Merisiers, Gif sur Yvette, France.
[2] Also at Laboratoire de Biogéochimie des milieux continentaux, UMR INRA-CNRS-PARIS6, INRA-INAPG, Thiverval-Grignon, France.

Copyright 2007 by the American Geophysical Union.
0886-6236/07/2006GB002834

for instance via the stomatal conductance and soil moisture budget), so that a multiple constraint approach to invert model parameters must be the ultimate goal. The essence of the multiple constraint approach is to use different kinds of measurements to constrain model parameters to their optimal values, and from these, to infer space-time distribution of carbon/water fluxes and pools or of other sought ecosystem variables. Parameter estimation techniques have been developed to optimize model parameters against CO_2 fluxes [*Vukicevic et al.*, 2001] at small scales, against CO_2 atmospheric concentrations [*Kaminski et al.*, 2002] at large scale, or even against both flux and concentration signals within carbon cycle data assimilation systems [*Rayner et al.*, 2005]. Recently, *Wang et al.* [2001] assimilated eddy-covariance CO_2, H_2O and heat fluxes, and investigated how many parameters could be inferred from these data. Their model structure however did not couple CO_2 fluxes with the dynamics of carbon pools, and thus neglected some of the key interactions within ecosystems.

[4] Here we optimize ecosystem functioning parameters against net radiation and eddy covariance fluxes of CO_2, H_2O and heat using a process-driven ecosystem carbon model, ORCHIDEE (*Krinner et al.* [2005] and http://www.ipsl.jussieu.fr/~ssipsl/), which calculates surface CO_2 and energy fluxes and carbon pools. The study is performed at the Bray forest, a temperate conifer site (Pinus Sylvestris) where continuous eddy covariance fluxes are measured continuously together with net radiation and ancillary information on leaf gas exchange, soil moisture and temperature, vegetation structure and phenology. We first describe briefly the model equations and the set of parameters that are sought for. Second, we describe the Bayesian optimization procedure. We then analyze the results of several inversions where the four types of (half-hourly) flux observations are assimilated either individually or altogether and either for a full year, or only few weeks in summer. In the last section, we analyze the degree of confidence that we can have on the optimization results.

2. Components of the Optimization Procedure

2.1. Eddy-Covariance Flux Data

[5] The Bray forest is located in southwestern France (44°43′N, 0°46′W). It is a nearly homogeneous plantation of 30-year-old *Pinus pinaster*, of mean height 18 m, with a projected Leaf Area Index (LAI) of about 3. The tree canopy is confined to the top 6 m, while the ground is covered by perennial grasses *Molinia Coerula*, with active leaves between April and November and LAI in the range 1.4–2.5. The water table rarely drops below 200 cm, except during periods of intense water stress [*Ogée et al.*, 2003]. Note that the Bray forest is growing and thus acts on an annual basis as a net carbon sink (average NEE equals to -350 gC m^{-2} yr^{-1} over 1996–2001).

[6] Fluxes and meteorology are measured in situ using the standardized CARBO-EUROFLUX protocol [*Aubinet et al.*, 2000] and produced on a 30-min basis. There are about 20% of data gaps in each flux time series but we only optimize parameters against real data. A usual difficulty when applying inverse methods to geophysical problems is to properly define the uncertainties on the data. Eddy-covariance fluxes contain random errors [*Moncrieff et al.*, 1996], which altogether limit their overall accuracy to typically 10–20% [*Wesely and Hart*, 1985]. They also contain biases with a systematic underestimation of fluxes during calm nocturnal period [*Goulden et al.*, 1996] and larger errors during rainy periods [*Wilson et al.*, 2002]. To take such limitations into account, we will assume that the sum of random and systematic errors follows a Gaussian distribution of standard deviation equal to 15% of the maximum datum during daytime (respectively 30% during nighttime) and of null mean. The typical resulting uncertainties on NEE, LE, H fluxes and Rn are of 2 μmol m^{-2} s^{-1}, 30 Wm^{-2}, 30 Wm^{-2}, and 50 Wm^{-2} respectively. Although this is likely not to be the case in reality, we will also assume that half-hourly data errors are independent one from each other (see section 6.4).

2.2. Biogeochemical Flux Model

[7] The ORCHIDEE biogeochemical ecosystem model (“ORganizing Carbon and Hydrology In Dynamic Ecosystems”) was originally developed for global applications, including the coupling with atmospheric models [*Krinner et al.*, 2005]. It is a process-driven model, which calculates fluxes with the atmosphere on a range of timescales, from 30 min to thousands of years. In the following, we applied the model in a “grid point” mode, forced by 30 min gap-filled meteorological measurements made on the top of the tower of air temperature (T_{air}), precipitation, specific humidity (q_{air}), wind speed (V_{wind}), pressure, short wave (R_{SW}) and long wave (R_{LW}) incoming radiation. The model contains a biophysical module dealing with photosynthesis and energy balance calculations each 30 min, a carbon dynamics module dealing with the allocation of assimilates, autotrophic respiration components, onset and senescence of foliar development, mortality and soil organic matter decomposition on a daily time step. The equations of ORCHIDEE are given by *Ducoudré et al.* [1993] and *Krinner et al.* [2005] and in http://www.ipsl.jussieu.fr/~ssipsl/. The optimized parameters are defined in Table 1 and the model equations linking them are given in Appendix A in Figure A1.

[8] As in most global biogeochemical models, the vegetation is classified into Plant Functional Types (PFT), with 13 different PFT over the globe. Distinct PFTs follow the same set of governing equations, but with different parameters values, except for the calculation of the growing season onset and termination, which involves a PFT specific parameterization [*Botta et al.*, 2000]. In order to treat the presence of *Molinia* grasses below the trees at the Bray site, we modeled the fluxes separately for the PFTs temperate needleleaf forest and C3 grasses, and combine grasses (30%) with trees (70%).

2.3. Parameters to be Optimized

[9] We optimize 12 parameters controlling directly or indirectly the net CO_2 flux (NEE), the latent heat flux (LE), the sensible heat flux (H), and the net radiation (Rn), according to Table 1. We distinguish, rather arbitrarily between “biophysical” parameters acting on photosynthesis

Table 1. Definition of ORCHIDEE Parameters to be Optimized[a]

Name	Description (units)	Prior Value	Range	GS Optimization		FY Optimization	
				Optimized Value	Error	Optimized Value	Error
			Photosynthesis				
K_{vmax}	Carboxylation maximum rates multiplier	1	0.1–2	1.48	2 (0.02)	1.19	1 (0.0006)
			Photosynthesis and Transpiration				
β	Slope of stomatal conductance	9	0.9–18	6.75	1 (0.01)	8.1	1 (0.0006)
f_{stress}	Soil water stress dependency of the canopy stomatal conductance slope	0.5	0.05–1	0.27	3 (0.03)	0.43	25 (0.015)
			Surface Energy Balance				
K_{alb}	Surface albedo multiplier	1	0.5–2	0.5[b]	1 (0.02)	0.5[b]	10 (0.006)
K_{Csoil}	Soil heat capacity multiplier	1	0.1–2	0.5	1 (0.02)	0.45	8 (0.005)
			Turbulent Transport				
K_{ra}	Aerodynamic resistance multiplier	1	0.1–2	0.47	10 (0.2)	0.43	66 (0.04)
K_{z0}	Roughness length	1	0.1–2	0.97	70 (0.7)	0.58	166 (0.1)
			Respiration				
K_{MR}	Rate of plants maintenance respiration	1	0.1–2	1.11	80 (0.8)	0.91	35 (0.02)
K_{GR}	Rate of plants growth respiration	0.28	0.028–0.56	0.13	20 (0.20)	0.2	35 (0.02)
K_{HR}	Rate of microbial respiration	1	0.1–2	0.77	23 (0.23)	0.7	35 (0.02)
Q_{MR}	Temperature dependency of maintenance respiration (K^{-1})	0.2	0.02–4	0.09	41 (0.41)	0.17	18 (0.01)
Q_{10}	Temperature dependency of microbial respiration (K^{-1})	2	1–4	1[b]	8.2 (0.08)	2.34	57 (0.03)

[a]The term GS stands for an assimilation of 3 weeks of growing season data, and FY denotes 1 year of flux data.
[b]EH denotes edge-hitting parameters whose optimized value reaches one bound of the optimization range. Given errors are normalized by the smallest error. Absolute errors are also provided in parentheses.

and transpiration (K_{vmax}, β, f_{stress}), surface energy budget (K_{alb}, K_{Csoil}) and turbulent transfer of scalars across the top of the canopy (K_{ra}, K_{z0}), and ''biological'' parameters (K_{MR}, K_{GR}, K_{HR}, Q_{MR}, Q_{10}) controlling plant and soil respiration. Those parameters beginning with the letter K do not have a direct biological meaning, but rather are multipliers of ''real'' parameters in the model. Moreover, some ''real'' parameters such as K_{vmax} or β depend on the PFT but we will optimize them in the same ratio for the two PFTs of the Bray site. In other words, we assume that the parameters for understorey grasses are proportional to those of the trees. This assumption is criticized in section 6.3. Hereafter, only the values for the dominant PFT are given. Our choice of parameters gives more emphasis on processes causing rapid variations in fluxes (diurnal to seasonal) rather than long-term changes in water and carbon budgets (e.g., tree growth, soil carbon turnover). The reason for this is that we exploit chiefly the information contained in the diurnal to seasonal variability of the fluxes, using one year of these data.

2.4. Inverse Algorithm

[10] We aim to find a parameter set that minimize the distance between model outputs and the corresponding observations, considering model and data uncertainties, and prior information on parameters. With the assumption of Gaussian errors for both the observations and the prior parameters, the optimal parameter set corresponds to the minimum of the cost function J(**x**) [*Tarantola*, 1987],

$$J(\mathbf{x}) = \frac{1}{2}\left[(\mathbf{y} - \mathbf{H}(\mathbf{x}))^t \mathbf{R}^{-1}(\mathbf{y} - \mathbf{?}(\mathbf{x})) + (\mathbf{x} - \mathbf{x_b})^t \mathbf{P_b}^{-1}(\mathbf{x} - \mathbf{x_b})\right], \tag{1}$$

that contains both the mismatch between modeled and observed fluxes and the mismatch between a priori and optimized parameters. **x** is the vector of unknown parameters, xb the a priori value, **H**() the nonlinear model (ORCHIDEE) and **y** the vector of observations. The covariance matrices $\mathbf{P_b}$ and **R** describe a priori uncertainties on

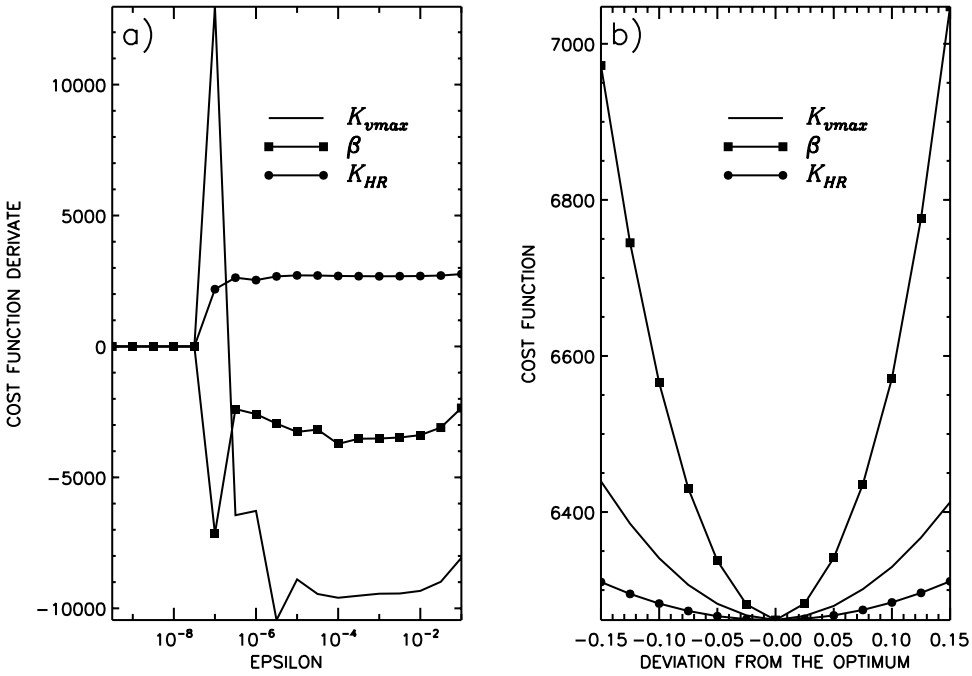

Figure 1. (a) Derivative of the cost function with respect to different parameters ($K_{v\max}$, β, and K_{HR}, Table 1), as a function of epsilon (ε), the step for the finite difference used to calculate the derivative (in normalized unit for all parameters, equation (2)). The derivatives are shown at the prior value of each parameter. Too low value of epsilon gives null derivatives because of computer precision limits (see text). (b) Cross sections of the cost function J(**x**) (equation (1)) for parameters $K_{v\max}$, β, and K_{HR} around their optimal values.

parameters, and on observations. Both matrices are diagonal as we suppose the observation uncertainties and the parameter uncertainties to be independent. This assumption is criticized in section 6.4.

[11] To determine an optimal set of parameters which minimizes J(**x**), we use a gradient-based algorithm which converges far more straightforwardly than Monte-Carlo methods. In our case, the cost running numerous Monte-Carlo simulations would be computationally too prohibitive. We adopt a limited memory quasi-Newton algorithm, called BFGS, that offers the possibility to prescribe an upper/lower limit for each parameter [*Byrd et al.*, 1995]. The BFGS algorithm requires at each step the calculation of the value and of the gradient of J(**x**), the Hessian (second-derivatives matrix) being approximated internally. With ORCHIDEE, we typically converge to a minimum of J(**x**) within $\approx$100 iterations.

[12] An exact estimation of the J(**x**) derivatives to find the minimum would lead in principle to the construction of the Linear Tangent model of ORCHIDEE, a rather sophisticated operation [*Giering and Kaminski*, 1996]. Here we use instead a classic finite difference method, assuming local linearity,

$$\frac{\partial J}{\partial x_i} \approx \frac{J(x_i+\varepsilon)-J(x_i)}{\varepsilon}. \qquad (2)$$

[13] We run twice the model for each parameter x_i, the critical point being to estimate a correct value for ε. With too small ε values, the numerical computation of the derivatives gets adversely affected by rounding errors. In Figure 1a, we show that the derivatives of J(**x**) with respect to the parameters $K_{v\max}$, β and K_{HR} tend to zero if $\varepsilon < 10^{-7}$. Conversely, too large values of ε ($>10^{-2}$) violate the locality requirement. The derivatives are less sensitive to the choice of ε for the range $[10^{-4}, 10^{-2}]$. We tested different ε values in that range. The well constrained parameters were robust to the value of ε. The poorly constrained parameters ($\partial J/\partial x_i$ very small) or the pairs of highly correlated parameters were sensitive to the value ε, which is a limit of our approach but fortunately does not degrade the inference of well-constrained parameters. In the following, the results were obtained with $\varepsilon = 10^{-3}$.

2.5. Uncertainties and Correlations Assessment

[14] The BFGS algorithm does not provide uncertainties nor correlations between parameters. It is yet crucial to determine which parameters are best resolved by the observations and which ones are not. Once the minimum is reached, we quantify the parameters uncertainties from the curvature of the cost function at its minimum (inverse of the Hessian). A large sensitivity of J(**x**) to a given parameter indicates that this parameter is tightly constrained by the measurements, i.e., has a small uncertainty. If the model is linear with Gaussian error distributions, the posterior probability distribution of the optimized parameters is also Gaussian [*Tarantola*, 1987]. Each parameter uncertainties and error correlations can thus directly be estimated by the variance-covariance matrix $\mathbf{P_a}$,

$$\mathbf{P}_a = \left[\mathbf{H}_\infty^t \mathbf{R}^{-1} \mathbf{H}_\infty + \mathbf{P}_b^{-1}\right]^{-1}, \qquad (3)$$

with $\mathbf{H}_\infty$ the derivative of all model output with respect to the parameters at the minimum of J(x). Large absolute values of correlations (close to 1) indicate that the observations do not provide independent information to separate a given pair of parameters.

[15] One thus need to verify the assumption of both the locality and linearity approximation to compute $\mathbf{P_a}$. Cross sections of the cost function J($\mathbf{x}$) for different parameters near their minimum are useful with that respect. This is illustrated in Figure 1b for $K_{v\max}$, β and K_{HR} parameters. All J(x) curves are of parabolic shape, which confirms a locally linear behavior of the model. Figure 1b also shows that the estimated parameters effectively represent a minimum of the cost function and, that some parameters are better constrained than others. From systematic inspections of the curvature of J($\mathbf{x}$) for each parameter, we checked that ORCHIDEE is only weakly nonlinear with respect to the selected parameters.

3. Optimization Settings

3.1. Time Windows of Inverted Flux Measurements

[16] We perform optimizations, differing by the period during which the flux data are assimilated. First, we use flux data during the summer peak of the growing season (19 July to 9 August), when the diurnal cycle is most pronounced. This is called the GS (Growing Season) inversion. Second, we use alternatively flux data for the entire year 1997 which corresponds to the FY (Full Year) set of inversions. The GS inversions was performed in order to test whether the ORCHIDEE model is generic enough to improve the seasonal flux variations using the parameters derived only from diurnal cycles during the growing season peak.

3.2. Initial Carbon Pools and State Variables

[17] Each model run requires knowledge of the initial biomass and soil carbon pool sizes. Unfortunately, there is no accurate enough pool data (nor site history data), and the model is likely not able to realistically simulate tree growth. To overcome this problem, we initialize biomass and soil carbon pools to their equilibrium values from a 250-year-long spinup driven by cycling 1997–1998 climate inputs. The initialization results in modeled annual NEE equal to zero, whereas in reality, the forest is a net carbon sink of nearly 350 gC m^{-2} yr^{-1}. We let the inversion corrects for this bias, by optimizing parameters K_{MR}, K_{GR}, K_{HR} which scale plant and heterotrophic respirations, and thus indirectly also determine pools sizes that are plainly compatible with the observed annual NEE (Appendix A, equations (A8), (A9), (A11) and (A12)). We further quantify in section 6.2 the sensitivity of optimized parameters to these initial conditions.

3.3. Initial Parameter Uncertainties and Bounds

[18] Initial parameters uncertainties and bounds are crucial in Bayesian inversions to avoid retrieving unrealistic values. Because we want to study the maximum information content of the data, we accordingly assign large prior uncertainties on parameters. The Bayesian term thus has a smaller influence on the retrieved values, but it still ensures the stability of the algorithm toward a proper determination of the minimum. The a priori value of parameters acting as multipliers (K_i) is obviously the unity. For most parameters, the prior uncertainty, lower and upper bounds are set to 100%, 1% and 200% respectively of their prior value (see Table 1). For K_{alb} and Q_{10}, we narrow the lower bound to 50%, because albedo <0.1 is unrealistic for forests, and because $Q_{10} < 1$ would violate the structural assumption that microbial respiration increases with temperature.

4. Fit to the Observations

4.1. Diurnal Cycle During the Growing Season

[19] We discuss the results of the GS and FY inversions together. They are shown in Figure 2. In the a priori model values, the diurnal cycle amplitude of each flux is underestimated, with NEE, LE and Rn being too weak during the day, and H being too large during the night. The GS inversion successfully corrects the diurnal cycle with a cost function reduction equating 41%, 37%, 32% and 84% for respectively NEE, LE, H and Rn data. In the GS and FY inversions, the model-data mismatch is greatly reduced, except for the sensible heat flux at night. However, the FY inversion only partially improves the fit to NEE during the day (Figure 2).

[20] For NEE, the good agreement between a priori model results and data during the night suggests that GPP is the critical gross flux that is underestimated. Shortcomings of the prior model can be seen both for the amplitude of daytime NEE and for the phase at dawn and in the evening (Figure 2). The uptake of CO_2 starts earlier and terminates later in the model than in the data. The optimization nearly doubles the fixation of CO_2 by plants at midday in the GS case, but it does not correct for the phase mismatch. Different explanations are possible for this. In the morning, the turbulent flux measurements can no longer be interpreted as instantaneous net fluxes in presence of a release of respired CO_2 formerly accumulated within the canopy [*Aubinet et al.*, 2000]. The magnitude of this storage effect (not present in the model) is still under investigation for the Bray site and could not be accounted for in this study. However, this effect occurs only at dawn and cannot explain the phase misfit in the evening. We attribute the evening and at least part of the morning misfit to the ''big leaf'' approximation. As ORCHIDEE does not distinguish between shaded and sunlit leaves, this model likely overestimates the amount of radiation intercepted by the canopy at low sun angles. This deficiency or other similar structural deficiencies related to stomatal functioning could not be corrected given the set of ''photosynthetic'' parameters we optimize.

[21] For the latent heat flux (LE), the GS and FY optimizations nicely increase the diurnal cycle amplitude, in a consistent way with the increase in carbon uptake. Contrary to NEE, however, there is no phase mismatch for LE (Figure 2). Given the covariance of NEE and LE, this would logically favor our hypothesis of storage ''contamination'' of instantaneous NEE flux data in the morning when the LE flux is small. For the sensible heat flux (H), the prior model fits the daytime observations well, but the

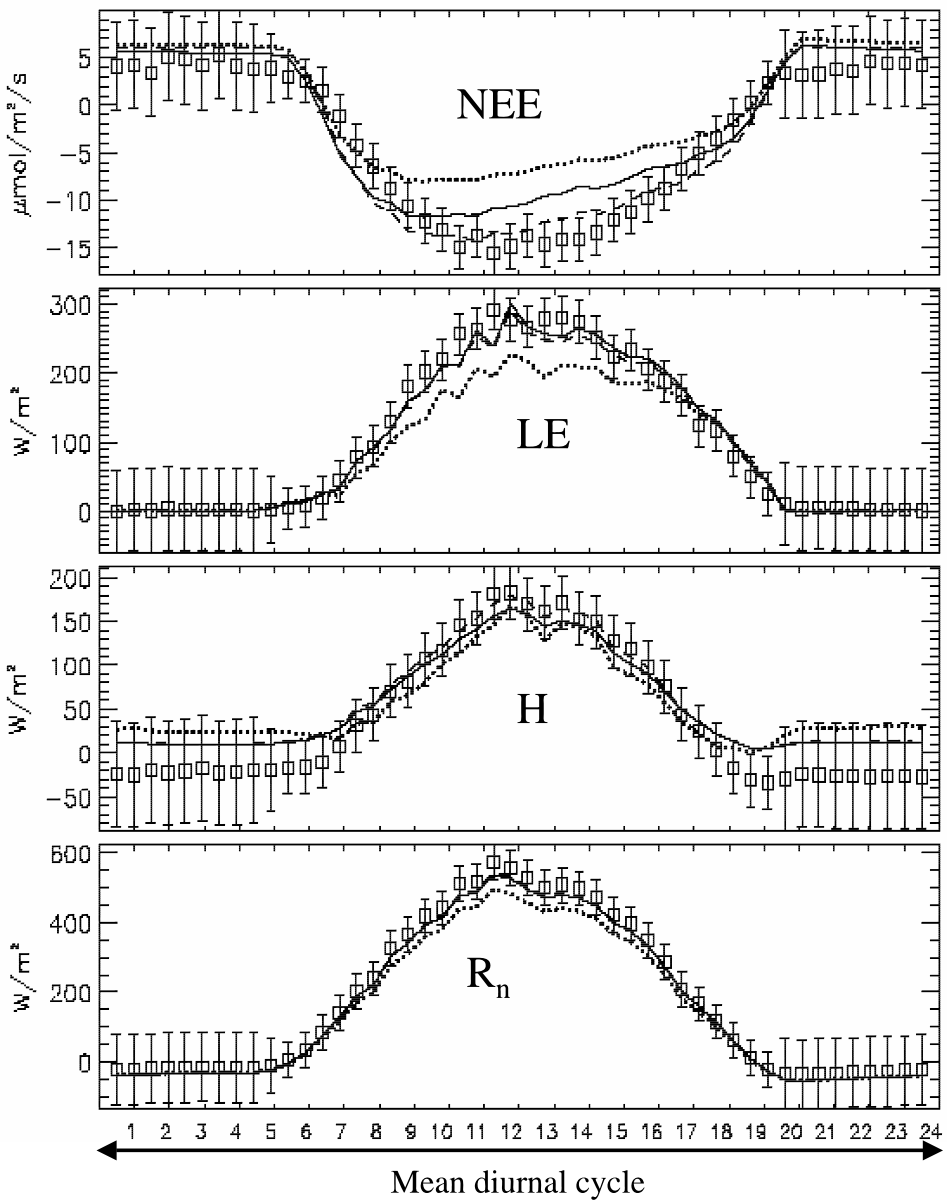

Figure 2. Bin-averaged diurnal cycles for the GS period (days 195–216 of 1997) and for each type of data (NEE, LE, H, and Rn). Diamonds with the errors bars represent the data and their uncertainties. Dotted lines show prior model outputs and solid and dashed lines model outputs with optimized parameters values for the full year (FY) and the growing season (GS) cases, respectively.

nighttime values of H are overestimated, even after the optimization. That misfit shows a structural problem of ORCHIDEE where a single surface temperature is calculated (T_s) ignoring thermal stratification within the canopy. *Ogée et al.* [2003] showed using a multilayer model, that a single temperature for the whole canopy does not allow to accurately model H at the Bray forest. The simulated nocturnal temperature gradient between the surface and the top-of-the-tower reference level (28 m) is probably too large for situations when a temperature inversion develops in the canopy, yielding even a slightly negative sensible heat flux during the night. Finally, we can see that the optimization is always able to successfully modify the overall energy balance, changing either the albedo, LE, H or the ground heat storage to balance the observed net radiation.

4.2. Results for the Full Year

[22] We first tested whether the diurnal variability derived from the GS optimization parameters gets also improved for other periods of the year when no data were assimilated. Figure 3 shows, for each type of data, the bin-averaged monthly diurnal cycles for the months of April, July, October and December as predicted with the GS-optimized parameters (dashed lines). As expected, the model fit to the

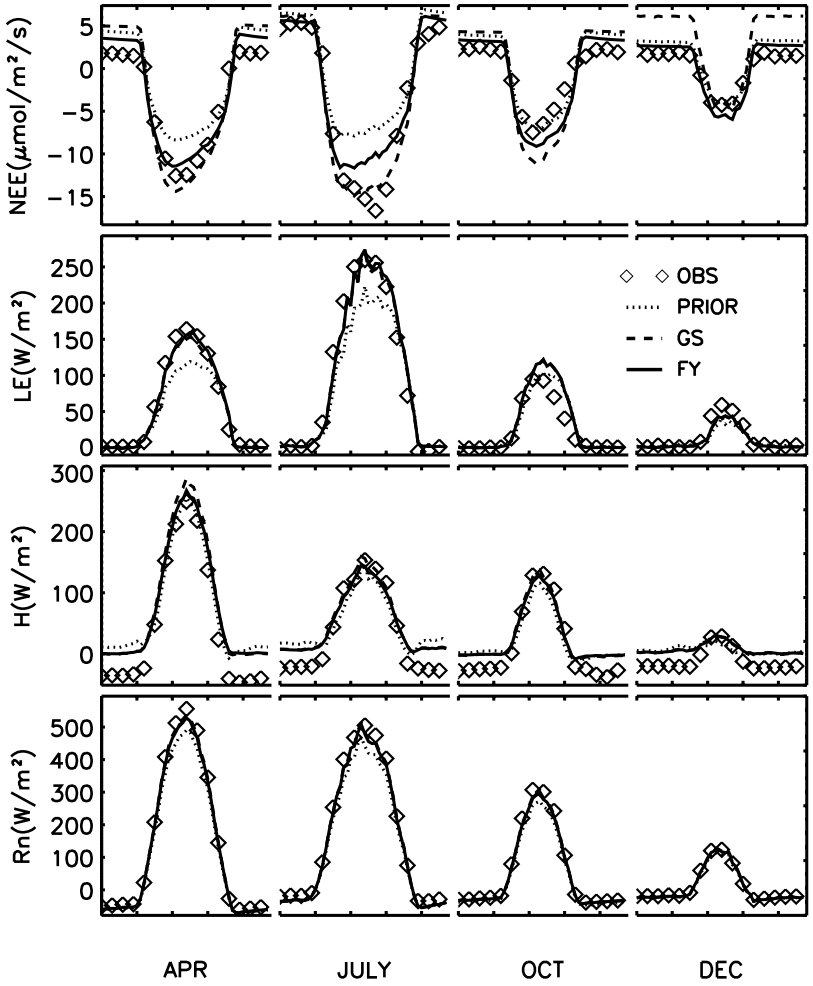

Figure 3. Bin-averaged monthly diurnal cycles for all data type and for the months of April, July, October, and December. Observations (OBS), a priori model simulation (prior), and model simulation with optimized parameters from 3 weeks of data (GS) and from 1 year of data (FY) are shown.

observed diurnal cycles is improved by the GS optimization for July and for each flux. For the rest of the year however, the improvement only concerns LE and H, while daytime NEE remains largely overestimated, both with too large daytime uptake (October) and with too large nighttime release (December). This result may indicate rigidities in ORCHIDEE and/or the need to optimize parameters specifically related to seasonal variability (e.g., phenology). The seasonal variations of the photosynthesis parameters are indeed probably underestimated. In ORCHIDEE, the carboxylation rates are function of temperature and leaf age [*Medlyn et al.*, 2002] and in further work, the parameters of this function should be optimized. If we now compare the optimized NEE between the GS optimization and the FY optimization, Figure 3 shows that the fit to daytime NEE in the FY case (plain line), is underestimated in July but better for the other months than in the GS case. The FY optimization seeks for an overall compromise among all the months with the model being too rigid to find a perfect adjustment each month. There are even few months, March, September, and October, where the a posteriori model-data fit is worse than in the prior (not shown). For other fluxes than NEE, we obtained no clear difference between the GS and the FY fit suggesting that the simulated components of the energy balance are well assessed for "slow" and "fast" timescales in ORCHIDEE, except for the sensible heat flux during the night. We conclude from this that we cannot optimize the ORCHIDEE model using only few weeks of high-quality data in summer (GS case) and that we rather need to use the information from a full year (FY case).

5. Parameter and Uncertainty Estimates

[23] We now examine the estimated parameter values, and their uncertainties, mainly for the FY inversion results. We intend to identify (1) parameters that can be robustly inferred from flux data to improve the model, and (2) parameters that have a nonregular behavior by being set to one of their bounds. These "Edge-Hitting" (EH) parameters may illuminate model structural deficiencies, or show an improper setup of the inverse problem. In Figure 4 we display the prior and posterior value of the

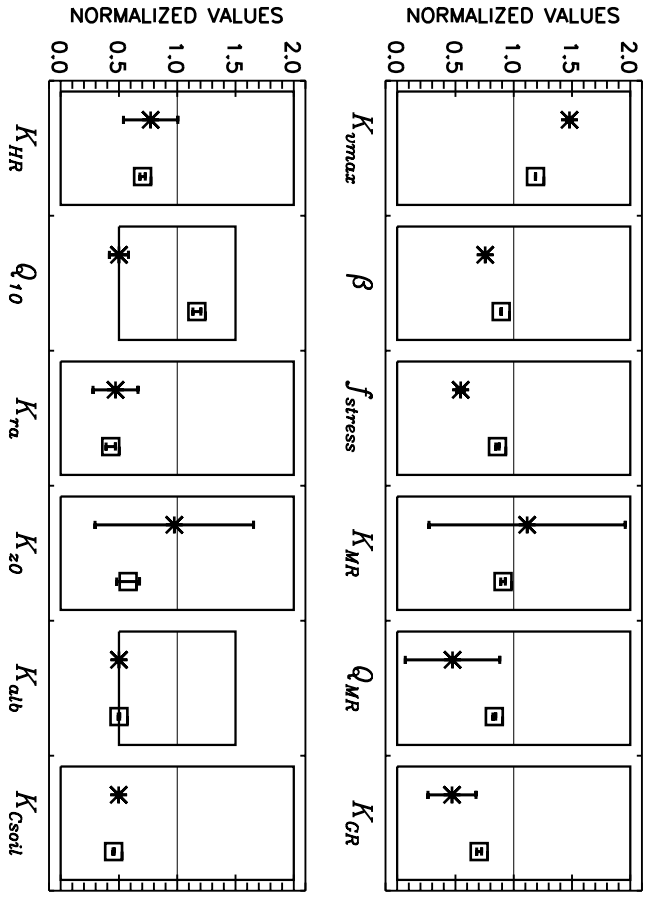

Figure 4. Parameter values and errors estimated by the optimization. Optimized values are normalized by the prior estimate (horizontal line in the boxes centered around one). The boxes' half height equals the normalized a priori uncertainty. Within each box are reported the parameters and errors retrieved from 3 weeks of data in the GS inversion (pluses) and from 1 year of data in the FY inversion (diamonds).

parameters together with the prior and posterior uncertainties. Each parameter is normalized by its prior estimate (value of 1) so that box widths in Figure 4 are normalized prior uncertainties. In complement, Table 1 also reports the optimized parameters values and uncertainties. Three main types of uncertainties are discussed in the following: (1) the 'true' uncertainty of posterior parameter given by the shape of the posterior probability density function, (2) the errors due to the limited precision of the optimization method to detect the best estimate within a given PDF, and (3) the uncertainty due to unknown state variables at the beginning of the observation period (biomass, soil carbon, reserves, soil moisture). The "true" uncertainties computed from equation (3) are globally underestimated. The shapes of the PDF are indeed not strictly coherent with the information content of the data, as detailed in section 6.4. However, the relative "true" uncertainty between the different parameters is informative and Table 1 thus reports each value relative to the best constrained parameter error (β). The uncertainties of type 2 and 3 (illustrated in Figure 6 in section 6.1), being linked to the inverse setup, are discussed in sections 6.1 and 6.2.

5.1. Photosynthesis Parameters

[24] The value of $K_{v\max}$, a multiplier that scales both V_{cmax} and V_{jmax} gets significantly increased from its prior estimate (Table 1) which reflects the increase in daily NEE required to match the measurements (Figure 2). For the dominant temperate needleleaf forest PFT, the FY optimization gives at 25°C V_{cmax} = 42.7 μmol m^2 s^{-1} and V_{jmax} = 83.3 μmol m^2 s^{-1}. Note the larger increase in photosynthetic capacities in the GS optimization (+48% versus +19% in FY), which reflects the larger increase in the NEE diurnal cycle in this case (Figure 3). Overall, both optimizations robustly indicate that the prior V_{cmax} and V_{jmax} values were underestimated for that PFT at the Bray site. The large error reduction in $K_{v\max}$ (Table 1) compared to other parameters enforces that results.

[25] The value of β, the slope of stomatal conductance versus assimilation decreases by nearly 25% in the FY optimization (Table 1). Decreasing β compensates for the increase in $K_{v\max}$. At first glance, we might expect β to increase in order to enhance the modeled LE. However, the $K_{v\max}$ large increase, driven by daytime NEE, propagates to an enhanced stomatal conductance which is probably too strong, and thus forces β to get reduced. However, the small change in the value of β suggests that its prior value was already realistic, and that the so-called Ball-Berry formulation [*Ball et al.*, 1987] is a good predictor of the stomatal conductance at the Bray site. Like for $K_{v\max}$, β is relatively well constrained by the data, indicating that NEE plus LE measurements strongly constrain those two key photosynthetic parameters (see section 5.4).

[26] The stomatal conductance also depends on the soil water content through the parameter f_{stress} (Appendix A, equation (A3)). The value of f_{stress} gets reduced by nearly 15% in the FY optimization (Table 1). Recall that f_{stress} defines a threshold below which the model reduces stomatal conductance in response to soil water limitations. Its initial value, defined to be the same for all PFT, is probably overestimated at the Bray site, knowing that conifers species suffer less than other species from soil water limitations [*White et al.*, 2000]. The *McMurtrie et al.* [1992] relationship we use to account for water effects on photosynthetic activity may be also too simplistic as well as the treatment of the soil hydrology with only two reservoirs ("double bucket model"). The future use of a more physical approach based on diffusion-type equations will be more convenient [*de Rosnay et al.*, 2002].

5.2. Respiration Parameters

[27] We found that all respiration parameters are poorly constrained by the flux data and remain strongly sensitive to their prior settings or to the initial carbon pools settings. For the five parameters controlling respiration, K_{MR}, K_{HR}, K_{GR}, Q_{10}, Q_{MR} we determined a smaller uncertainty reduction than for all other parameters, and strong correlations among errors in distinct parameters: we obtain values between 0.5 and 0.8. This points out to the fact that NEE alone cannot properly separate each respiration component, nor the pertaining parameters. Note that the Q_{10} parameter exhibits an "edge-hitting" behavior toward its lower bound (Q_{10} = 1)

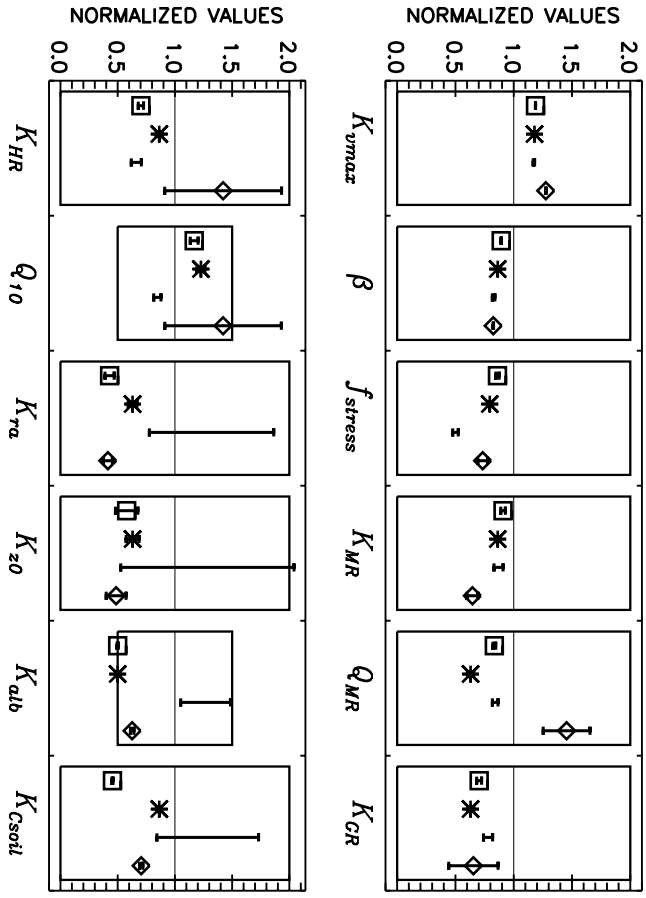

Figure 5. Same as Figure 4 but parameters are inverted from NEE, LE, H, and Rn flux data (pluses), from NEE and LE data only (crosses), from NEE only (dots), and from LE only (diamonds).

in the GS case. The inversion thus flattens the temperature dependency of heterotrophic respiration. This indicates that the nighttime NEE data do not exhibit significant correlations with soil temperature during the short GS assimilation window of 3 weeks. This result is consistent with the fact that the so-called "Q_{10} relationship" has been obtained over seasonal timescales using soil respiration data regressed against air temperature [*Raich and Potter*, 1995] or soil temperature [*Raich and Schlesinger*, 1992]. Therefore this relationship has conceivably no predictive power for matching day-to-day NEE variability. However, the full year inversion (FY) also poorly determines the respiration parameters. We found that the relative error reduction for each respiration parameters is globally the same in both optimizations (Table 1). As further discussed in section 6.1 the information contained in the diurnal cycles of NEE probably overcomes the signal of smoother seasonal variations when assimilating one year of 30 min flux data points.

5.3. Energy Balance and Turbulent Transport Related Parameters

[28] We found that K_{alb}, a multiplier of the surface albedo, is decreased to its lower bound, yielding to unrealistic albedo values ($\approx$0.1) for conifers (Figure 4 and Table 1). This may be due to a model deficiency related to the energy balance calculation. We outlined above in 4.1 a mismatch for nighttime H, which could reflect the use of a unique model surface temperature, T_s, thus overestimated at night. The albedo decrease is consistent with the necessity to increase T_s during the day in order to balance the net radiation flux, but this daytime rise in T_s gets probably too large in the big-leaf approximation, and propagates into an overestimated nighttime H flux (Figure 2). To counteract this effect, the soil thermal capacity, K_{Csoil} gets also reduced by a factor of two (Table 1). Such a change diminishes the soil heat storage during the day (equation (A16)) and helps to maintain cooler surface temperatures during the night. Another possibility is that the adjustment of K_{Csoil} occurs as a "free" degree of freedom in order to close the surface energy balance, which eddy-covariance observations generally fail to do [*Wilson et al.*, 2002].

[29] The parameters K_{ra} and K_{z0} controlling turbulent transport across the top of the canopy are reduced in GS and FY inversions. This implies a reduction in aerodynamic resistance r_a after equation (A7), and in turn, helps to increase H and LE during the day to match the data (Figure 2). The a priori value of r_a was probably overestimated because the roughness height z_0 itself was too high owing to the overestimated surface temperature. The small uncertainty reduction for K_{ra} and K_{z0} as compared to other parameters is however, rather surprising. A closer look at the different resistance terms indicates that r_a is small ($\approx$30 s m^{-1}) compared to the stomatal resistance ($\approx$200 s m^{-1}), explaining the weak sensitivity of NEE and LE to turbulent transport parameters K_{ra} and K_{z0}.

5.4. How Each Flux Constrains the Different Parameters

[30] Figure 5 compares the FY-optimization parameters obtained by assimilating only one type of flux (either NEE, LE, H or Rn) with those obtained by using all fluxes, or the pair NEE plus LE. The main conclusion is that assimilating the couple NEE, LE allows to constrain each parameter almost equally as well as when using the four fluxes, except for K_{Csoil}. The optimization of K_{vmax} and Q_{10} against NEE data alone, leads to the same solution than when using all the fluxes (compare points and crosses in Figure 5). Assimilating only the LE data (diamonds in Figure 5) is also powerful enough to robustly recover all parameters controlling the energy balance, β, f_{stress}, K_{ra}, and K_{alb}. Apart for the soil thermal capacity K_{Csoil}, for which the optimization requires all four different fluxes, one could have used only NEE and LE to optimize the parameters. This result indicates that H and Rn linked to the latent heat flux via the energy balance closure do not bring any significant additional information, at least in our set up, and for the site chosen.

6. Discussion

[31] Discussion is focused on the implications of the inversion set up on the parameter values and their uncertainties. In particular, we address how much information on

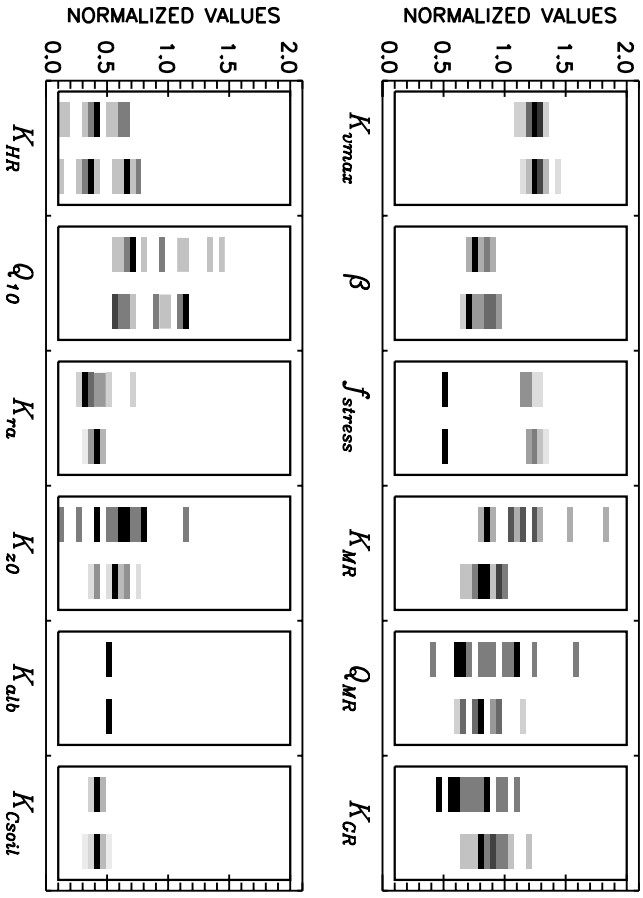

Figure 6. (left) Distribution of the estimated parameter values from a series of 50 different optimizations, where the initial value of the parameters are randomly changed up to 30% for the FY case. (right) Fifty inversions where the initial carbon pool sizes were randomly changed up to 30% for the FY case. Parameter values are normalized with respect to the prior values. The distribution of the 50 estimates is binned into 0.05 intervals. The color table is defined as a linear scale with the maximum values per interval set to black and the zero value to white.

parameters can be retrieved from eddy-covariance data. A major positive outcome of inverse methods is to be able to reject a parameterization without neglecting uncertainties on the data and on other parameterizations. For non-edge-hitting parameters, one can still wonder whether an "optimal" value yet has a useful biogeochemical significance, or rather reflects an impetus to compensate for model deficiencies. Missing processes are always an overlooked source of bias in inversions.

6.1. Number of Optimized Parameters and Sensitivity to Prior Values

[32] A gradient-based minimization algorithm, such as BFGS, has shortcomings when some parameters are equally resolved by the data or are very poorly constrained. This stems from the inversion of the Hessian matrix, which may end up being ill-conditioned and not easily invertible. To overcome this problem, *Wang et al.* [2001] for instance, excluded poorly constrained parameters from their inversion. This approach is not justified from a physical point of view. It adds bias to our effective knowledge of the remaining parameters and possibly yields to overestimate their incremental change from the prior because sensitive parameters are frozen in the beginning. It is difficult to assess beforehand which parameters must be excluded or not. A Monte Carlo sensitivity analysis to browse parameters can be used as an alternative [*Bastidas et al.*, 1999; *Franks et al.*, 1997]. We found that freezing some parameters is unnecessary, provided that the cost function contains a Bayesian term to regularize the Hessian. One should thus not hold any restriction upon the number of parameters to be optimized. In our case, we choose 'only' 12 parameters (i.e., all the parameters related to rapid flux changes) out of nearly 50 in total in the model, because of computing limitations. Calculating each derivative by a finite difference scheme requires one run of ORCHIDEE per parameter and per iteration.

[33] If we increase the number of parameters, we face the problem that different physically reasonable parameter sets optimally fit the data. This problem of equifinality [*Franks and Beven*, 1997; *Schulz et al.*, 2001] is present in our inversions and makes it more difficult to assess the best set of parameters. Equifinality mainly results in this study from the approximate derivatives of J(x) in the finite difference calculation, which hampers a robust convergence of the algorithm for ill-constrained parameters. Equifinality does not result from J(x) having multiple local minima, as verified by inspecting the cross sections of J(x) at the last iteration (see examples in Figure 1b). We also show that, among the different parameter sets leading to a similar fit to the data, the parameters retrieved with the greater uncertainty reduction (equation (3)) have the narrower interval of optimal values. To investigate this, we made 50 sensitivity tests (FY optimization) by changing randomly the prior values. We show in Figure 6 (first column) the frequency distribution of the parameters best estimates. One can see that different parameter sets lead to similar model-data fits, with differences in $J(\mathbf{x}) < 5\%$, but the well-constrained parameters are much less sensitive to their initial value than the poorly constrained ones (compare Figure 4 and Figure 6). In the end, we conclude that five parameters, K_{vmax}, β, K_{ra}, K_{alb} and K_{Csoil} can be robustly estimated with a standard deviation for the 50 estimates lower than 0.1. Table 1 indicates that these parameters have indeed the greatest uncertainty reduction as estimated from equation (3) (except for K_{ra}).

6.2. Sensitivity to Initial Carbon Pools Values

[34] We have shown in section 5.1 that photosynthetic parameters are well determined by the optimization. We investigate here to what extent the inversion solutions can be affected by the fact that the prior modeled carbon pool sizes correspond to equilibrium (annual NEE = 0), whereas in reality the Bray forest is a net carbon sink (annual NEE = -350 gC m^{-2} yr^{-1}). This is an additional source of uncertainty (related to the inversion set up) which might hamper the determination of the optimal parameter set.

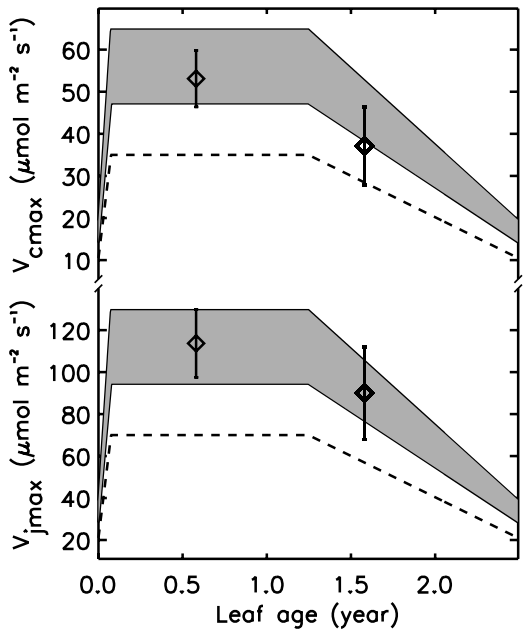

Figure 7. Dependency of carboxylation rates $V_{c\max}$ and $V_{j\max}$ (at 25°C) on leaf age in ORCHIDEE for the prior (dashed line) and for the optimized model (grey area). The grey area corresponds to the range of $K_{v\max}$ given by the sensitivity study of the GS retrieved parameters to perturbation of the initial carbon pool sizes. Independent leaf-scale cuvette determinations at Bray for ''young'' and ''old'' needles with their uncertainties (diamonds) are used to verify the inverse results.

[35] We verified that optimizing respiration parameters K_{MR}, K_{GR}, K_{HR} which also scale the soil pools, allows to fit the observed annual NEE well. We first performed several sensitivity tests perturbing the half-hourly flux data in order to scale the annual average NEE between −750 and 0 gC m^{-2} yr^{-1}. We found strong correlations between the annual NEE magnitude and the optimized values of K_{GR}, K_{MR} and K_{HR} but no correlations with all other parameters. We then tested the sensitivity of optimized parameters to the initial carbon pools size, through 50 sensitivity inversions where the initial leaf, sapwood, litter and soil pools are randomly changed within a 30% range. The results (Figure 6, second column) indicate that the photosynthetic parameters ($K_{v\max}$, β and partially f_{stress}) and the energy budget parameters (K_{alb}, K_{Csoil}, K_{ra} and to a less extent K_{z0}) are only weakly sensitive to the assumed initial carbon pools value. In contrast, all the poorly constrained respiration parameters critically depend on initial conditions.

[36] Note, however, that we still have an inconsistency in the model because the initial pool values are determined from the prior parameter values, after an equilibrium run. To improve this, one should solve independently for the initial carbon pool values. In this study, we have implicitly optimized in that manner the initial pools through the respiration parameters, K_{MR}, K_{GR}, and K_{HR}. These three respiration parameters cannot be meaningfully related to respiration rates because their value depends on state variables at the beginning of the observation period (biomass, reserves). However, they can be regarded as meaningful for adjusting these unknown initial variables and bringing them to compatible values with observed NEE. One may finally wonder if the initial soil moisture content also impacts the parameters retrieval. For the year 1997, owing to abundant rainfalls, the initial soil moisture has only a minor influence on the optimized parameters, but this might not be true for drier conditions.

6.3. Validation of Optimized Parameters Against Independent Observations

[37] We compare now the optimized carboxylation rates $V_{c\max}$ and $V_{j\max}$, obtained via the multiplier $K_{v\max}$ against independent leaf-scale cuvette measurements from [*Porté and Loustau*, 1998]. Leaf-scale carboxylation rates are determined at 25°C from the fitting process of assimilation versus leaf internal CO_2 concentration and PAR data, using a photosynthetic model. Cuvette measurements were taken both for young needles (0.1 to 1 years) and for old cohorts (1 to 2 years). In ORCHIDEE, $V_{c\max}$ and $V_{j\max}$ are function of the leaf age [*Ishida et al.*, 1999] and of leaf temperature [*McMurtrie et al.*, 1992]. In Figure 7, we compare the model prior and optimized age dependency of $V_{c\max}$ and $V_{j\max}$ with the leaf measurements. The grey area is the range of optimal values, defined by the 50 sensitivity tests to initial carbon pools. Note that we consider here the GS optimization for which air temperature was close to 25°C as in the leaf-scale observations. The optimized values are closer to the leaf observations than were the prior, suggesting that the prior values of $V_{c\max}$ and $V_{j\max}$ for the Bray site were underestimated. A cause for this could be that we use carboxylation rates measurements on *Pinus Pinaster,* whereas in the model this parameter combines trees and understorey grasses with different individual values. *Wang et al.* [2001] showed that nonlinearity makes it impossible to separate carboxylation rates from different PFTs spatially distributed in the footprint of the same tower. We performed a sensitivity optimization assuming 100% of trees, and found that the optimized carboxylation rates remained close to those of the mixed ecosystem case. This is because the trees have a larger light interception all year round, and grasses are usually fairly dry in summer: 70% of the total NEE come from the pine trees [*Ogée*, 2000]. The inversion proves here to be quite useful to update the value of $V_{c\max}$ and $V_{j\max}$ for the evergreen needleleaf forest at the Bray site. In general, inversion performances should be verified if possible by using independent validation data.

6.4. Overall Uncertainty Estimates

[38] We now discuss the estimated Bayesian ''true'' uncertainty, in comparison with other sources of error. The inverted parameter errors reported in Table 1 (or ''true'' uncertainties) are unrealistically low, as shown in section 5. In addition, these errors are smaller in the FY case than in the GS case. This behavior reflects an overestimation of the information content of the eddy-covariance flux data in our approach. We suppose that all flux data errors are independent (**R** matrix is diagonal in equation (3)) and assimilate half-hourly LE, H and Rn observations altogether, whereas these fluxes remain

linked by the energy balance closure. This latter "hidden redundancy" was evidenced in section 5.4, with the result that most parameters could be determined with only the NEE, and LE fluxes, whereas the error reduction was yet much larger when matching the four types of fluxes rather than NEE plus LE alone (Figure 5). We prefer however to keep using the entire set of observations because the energy balance is not closed for most of the flux towers [*Falge et al.*, 2001; *Wilson et al.*, 2002], and we ignore which flux(es) is biased versus the unknown truth.

[39] Determining adequate error correlations to account for the right amount of information in the data is not an easy task. Data error covariance could be estimated from the data autocorrelation (data correlations taken as a proxy of error correlations), given that those errors mainly arise from model uncertainties. For the NEE, we found a significant autocorrelation for at least 8 days (not shown) which reflects that nighttimes values hardly change days after days. This suggests that we must account for error correlations in order to obtain better parameter uncertainties. However, the resulting underestimation of the uncertainties is likely to affect all parameters. As discussed in section 6.1, the relative error reduction becomes then useful to distinguish well-constrained from poorly constrained parameters. Note that the residuals (standard deviation between data and model outputs) only slightly differ from Gaussian distribution (with larger tails), which is a critical hypothesis for the error estimates in equation (3).

[40] As shown in section 5, sensitivity tests upon prior parameter values and initial pool values give additional hints on parameter uncertainties. The use of a finite difference scheme to approximate the derivative of the cost function increases the uncertainty of poorly constrained parameters. The standard deviations of parameters across the different sensitivity tests illustrated in Figure 6 are much larger than the "true" uncertainties from the Bayesian inversion (equation (3)) and they give a more realistic error estimate. In this approach, the respiration parameters are found to have uncertainties of 30–50%. In contrast, the photosynthesis parameters have much smaller errors, on the order of 5 to 10%.

7. Conclusion and Perspectives

[41] We designed a simple and generic inversion method to optimize many parameters of an ecosystem model using high-frequency eddy covariance flux measurements of NEE, latent heat, sensible heat, and net radiation. The method is tested at a pine forest site in France, for which we inverted 12 parameters controlling the carbon, water, and energy fluxes in a detailed process oriented model. The model is nonlinear and deals with multiple timescales and their interactions. We focus the analysis on the robustness of inversion results, and on the numerous biases and uncertainties which hinders the retrieval of parameters.

[42] We calculate explicitly the uncertainty reduction on each parameter, using a Bayesian formalism. However, we use a range of sensitivity studies to estimate an overall uncertainty range and the robustness of the solutions. Direct use of Bayesian posterior uncertainties would need to take into account the error correlations between half-hourly data. We found that only few parameters, mostly related to photosynthesis and energy balance can be robustly inferred from the flux data, while other "edge-hitting" parameters are still useful to point out to structural deficiencies of the model. Carboxylation rates could be verified successfully with independent leaf-scale data, but respiration parameters show spurious variations, and are not robust through the different sensitivity inversions. However, instead of freezing the ill-constrained parameters (as was done in some earlier studies), we recommend to adjust all parameters to avoid aliasing effects.

[43] The inversion successfully matches the observed diurnal cycle of the diverse fluxes, and corrects for the prior misfit to daytime NEE and LE. Some discrepancies remain for sensible heat emissions at night. Our inversion procedure tends to favor the fit to the diurnal variations over the fit to the seasonal variations. This has been quantified with a Fourier Transformation of the difference between the posterior and the prior model outputs (not shown) and it primarily results from the fact that we consider all diurnal cycles to be independent throughout the year.

[44] In this attempt to optimize a nonlinear complex ecosystem water-carbon-energy model we estimated the potential of the different eddy covariance fluxes. Further improvements should include (1) the optimization of additional parameters especially those related to the seasonal processes (e.g., the temperature dependence of carboxylation), (2) the optimization of initial carbon pool sizes, and (3) the modification of the cost function to account more evenly for the information from different timescales.

[45] A logical next step would be to invert parameters for different vegetation types, using data of the FLUXNET program (http://daac.ornl.gov/FLUXNET/). Whether the parameters are invariant or site specific, and whether they take or not different values for different PFTs, are important research questions that can be investigated. In a further development we also must invert parameters linked to the "slow" carbon processes like biomass allocation, and turnover times of soil carbon. We could then use "slow" biometric measurements jointly with "fast" flux measurements.

Appendix A: ORCHIDEE Model

[46] The components of the ORCHIDEE model together with the number of parameters optimized in each submodule are describe in Figure A1.

A1. NEE Component Fluxes

[47] Processes by which parameters are related to fluxes and other prognostic variables of the model are summarized below. The parameters we optimized are written in bold. The Net Ecosystem Exchange flux (*NEE*) is calculated as the sum of four terms,

$$NEE = R_m + R_g + R_h - A. \tag{A1}$$

R_m is the maintenance respiration, R_g the growth respiration, R_h the heterotrophic respiration and A the net carbon

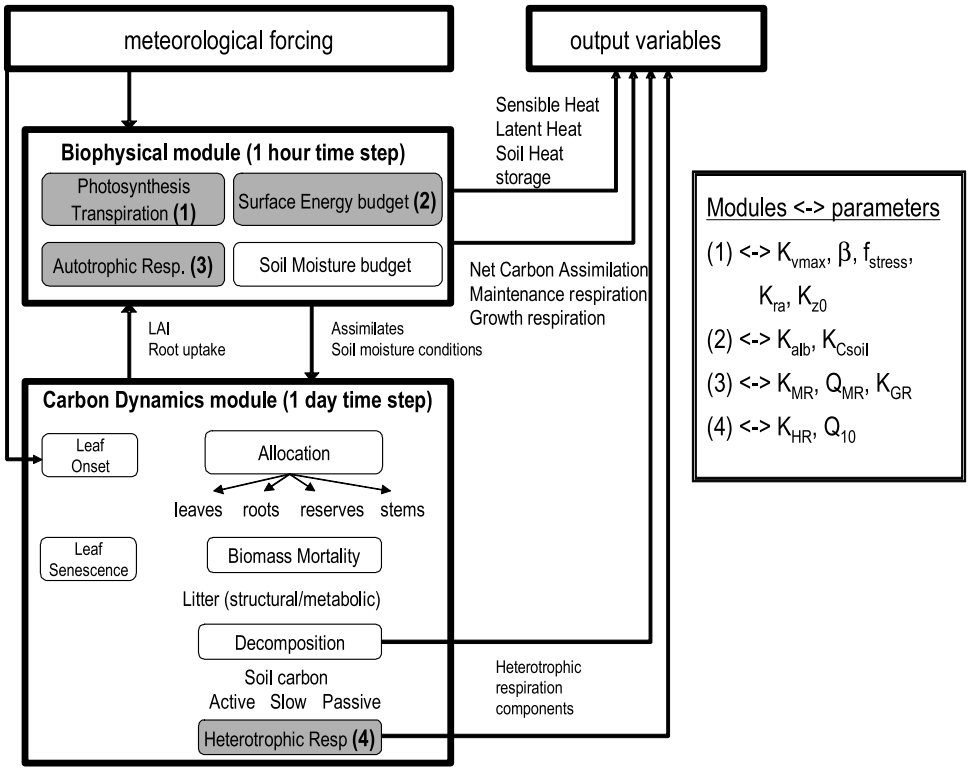

Figure A1. Flow diagram of the ORCHIDEE model structure with the different modules for which the parameters are optimized. Number in parentheses in the diagram refers to the parameters listed in the right-hand box.

assimilation rate (A = photosynthesis minus leaf respiration in light). ORCHIDEE is a "big leaf" model where vegetation is treated as a single equivalent surface for the carbon cycle. Net assimilation (A), stomatal conductance (g_s) and the CO_2 concentration in the chloroplast (C_i) are solutions of the following system of three equations:

$$g_s = \beta . w_l . A . h_r / C_a + gsoffset, \tag{A2a}$$

$$A = \boldsymbol{K_{vmax}} \,.\, \min\left(V_c, V_j\right).\left(1 - \frac{\Gamma^*}{C_i}\right) - R_d, \tag{A2b}$$

$$A = g_s(C_a - C_i). \tag{A2c}$$

Equation (A2a) gives the stomatal conductance $g_s(A, C_i)$ following the experimental data of *Ball et al.* [1987] obtained for plants under no-stress conditions. Here β is the slope of the stomatal conductance versus A linear relationship, h_r the relative air humidity (%) and C_a the CO_2 atmospheric concentration. To account for the effect of soil water stress on the stomata aperture, g_s is modulated by a function w_l of the water fraction available for the plant in the root zone f_w,

$$w_l = \begin{cases} 1 \;\; if \;\; f_w > \boldsymbol{f_{stress}} \\ \dfrac{f_w}{\boldsymbol{f_{stress}}} \;\; if \;\; 0 < f_w < \boldsymbol{f_{stress}}. \end{cases} \tag{A3}$$

Parameter $\boldsymbol{f_{stress}}$ defines the soil water fraction above which maximum opening of the stomata occurs ($w_l = 1$).

[48] Equation (A2b) describes $A(C_i)$ with distinct rates of carboxylation for the Rubisco (RuBP) limited regime (V_c) and the electron transfer limited regime (V_j), following *Farquhar et al.* [1980] for C3 photosynthesis and *Collatz et al.* [1992] for C4 photosynthesis. The Rubisco-limited carboxylation rate is given by

$$V_c = \frac{V_{c\max} . C_i}{C_i + K_c(1 + O_i/K_o)}. \tag{A4}$$

$V_{c\max}$ (μmol m^{-2} s^{-1}) is the maximum rate of RuBP carboxylation, K_c and K_o are the Michaelis-Menten constants for enzyme catalytic activity for CO_2 and O_2 respectively, and O_i is the intercellular concentration of Oxygen. The electron transfer limited regime is defined by a nonrectangular hyperbola function of the incident photon flux I,

$$V_j = \frac{\dfrac{1}{2\Theta}\left[\alpha_j I + V_{j\max} - \sqrt{\left(\alpha_j I + V_{j\max}\right)^2 - 4\Theta\alpha_j I V_{j\max}}\right]}{1 + 2\Gamma^*/C_i}. \tag{A5}$$

$V_{j\max}$ defines the maximum potential rate of RuBP regeneration at quantum saturation, α_j the quantum yield of RuBP regeneration, Θ the curvature of the quantum response and Γ^* the CO_2 compensation point. Both $V_{j\max}$

and $V_{c\max}$ are optimized simultaneously via the parameter, $\boldsymbol{K_{v}}_{\max}$, which scales both $V_{j\max}$ and $V_{c\max}$ (equation (A2b)), assuming that they are linearly related. Several studies have shown that $V_{c\max}$ and $V_{j\max}$ are linked by the nitrogen cycle in a factor of ≈ 2 [*Leuning*, 2002],

$$V_{j\max} = 2.V_{c\max}. \tag{A6}$$

Equation (A2c) calculates the gas phase molecular diffusion of CO_2 from canopy air to chloroplast. Altogether, the system of equations (A2a), (A2b), and (A2c) is solved iteratively to update at each time step the values of A, g_s and C_i at the leaf level.

[49] To scale up to the canopy level, we integrate the value of A and g_s over the canopy depth, that is over the leaf area index (LAI) assuming an exponential decrease of $V_{c\max}$ and $V_{j\max}$ [*Johnson and Thornley*, 1984]. By doing so, we introduce the canopy aerodynamic resistance, r_a, which embodies the resistance to the transfer of CO_2 between the canopy and the measurement plane,

$$r_a = \frac{1}{\boldsymbol{K_{ra}}.V_{wind}.C_d(\boldsymbol{K_{z0}})}. \tag{A7}$$

The value of r_a depends on canopy turbulence and on surface roughness via the surface drag coefficient C_d. V_{wind} is the wind speed norm. We optimize two parameters: $\boldsymbol{K_{ra}}$ a multiplier of the overall aerodynamic conductance ($1/r_a$) and $\boldsymbol{K_{z0}}$ modifying the roughness height z_0 which intervenes in the calculation of C_d.

[50] The release of CO_2 to the atmosphere by maintenance respiration (R_m) is function of each living biomass pool B_i, and has a linear temperature dependency $c(Ti)$, where T_i is the temperature of the pool i [*Ruimy et al.*, 1996],

$$R_m^i = \boldsymbol{K_{MR}}.c(T_i).B_i. \tag{A8}$$

For leaves maintenance respiration, a function of the leaf area index also enters the calculation,

$$R_m^{leaf} = \boldsymbol{K_{MR}}.c\left(T_{leaf}\right).B_{leaf}.f(LAI), \tag{A9}$$

where $\boldsymbol{K_{MR}}$ is a parameter to be optimized and acting as a multiplier of the whole maintenance respiration flux. The slopes and intercepts of the temperature dependency functions $c(T_i)$ are different for each pool and the slopes will be optimized via parameter $\boldsymbol{Q_{MR}}$ in the same proportion for each pool.

$$c(T_i) = \max\left(c_0^i.(1 + Q_{MR}.T_i), 0\right). \tag{A10}$$

Growth respiration R_g is computed as a fraction of the difference between assimilation inputs and maintenance respiration outputs to plant biomass. This fraction is scaled by the parameter $\boldsymbol{K_{GR}}$,

$$R_g = \boldsymbol{K_{GR}}.(A - R_m). \tag{A11}$$

[51] Processes controlling the decomposition of litter, soil organic matter, and subsequent heterotrophic respiration (R_h) losses of CO_2 to the atmosphere are similar to those described by *Parton et al.* [1988] and popular among biosphere modelers. Soil litter laid off to the forest floor distinguishes a structural and a metabolic pool of distinct turnover times. Soil organic matter is distributed among three soil carbon pools of increasing turnover with carbon flowing among them and CO_2 emitted to the atmosphere by heterotrophic processes. The evolution of each pool is governed by a first-order linear differential equation, where pool-specific turnovers have soil moisture and soil temperature dependencies (A12). We define a multiplier $\boldsymbol{K_{HR}}$ of the total R_h flux that we optimize. The moisture dependency $g(swc)$ of R_h is not optimized. The temperature dependency of R_h, currently being parameterized with a Q_{1O} function of soil temperature T_{soil} is optimized by adjusting $\boldsymbol{Q_{1O}}$,

$$R_h = \boldsymbol{K_{HR}} \sum_s \alpha_s.B_s.g(swc).\boldsymbol{Q_{10}}^{\frac{T_{soil}-30°C}{10°C}}. \tag{A12}$$

In equation (A12), B_s is the size of each soil carbon pool and α_s is a pool specific coefficient partitioning heterotrophic respiration into pools.

A2. Net Energy Balance Component Fluxes

[52] The latent heat flux (LE) is computed as the sum of snow sublimation, soil evaporation, plant transpiration and evaporation of water intercepted by foliage. Each of these fluxes is linearly related to the gradient of specific humidity between the evaporating surface (q_i) and the air overlying the canopy (q_{air}), the latter being an input data. The aerodynamic resistance r_a intervenes in the calculation of all LE components, as illustrated for plant transpiration ET in (equation (A13)). The value of r_a mediates the transfer of all scalars from their emitting surface up to the top of the canopy. One can see the importance of the couple of parameters $\boldsymbol{K_{ra}}$ and $\boldsymbol{K_{z0}}$ controlling r_a in equation (A7), since it directly impacts the triad of fluxes NEE, LE and H. Moreover, LE also depends on the stomatal conductance g_s ($g_s = 1/r_s$) via its transpiration component ET and thus constrains the parameter $\boldsymbol{\beta}$, according to (A13) below, where K_i is a coefficient specific to each evaporating surface.

$$ET = \sum_i \frac{K_i}{r_a + r_s}(q_i - q_{air}). \tag{A13}$$

[53] The sensible heat flux H is entirely determined by the aerodynamic resistance r_a, the surface temperature T_s and the input data of air temperature T_{air}, according to

$$H = \frac{\rho_{air}.C_{air}}{r_a}(T_s - T_{air}), \tag{A14}$$

where ρ_{air} and C_{air} are the air density (m^{-3}) and the air thermal capacity (J m^{-3} K^{-1}) respectively. The surface temperature in ORCHIDEE is derived from the energy budget calculation, considering that soil and vegetation form a single medium assigned with a single surface temperature T_s. The energy balance is expressed by

$$(1 - \boldsymbol{K_{alb}}.albedo).R_{SW} + R_{LW} - \varepsilon\sigma T_s^4 = R_n = H + LE + G. \tag{A15}$$

[54] We optimize a multiplier of the *albedo*, $\boldsymbol{K_{alb}}$ in equation (A14) directly against net radiation flux data R_n

and indirectly via changes in T_s, against energy fluxes H and LE. In equation (A14) the incoming short and long wave radiation fluxes R_{SW} and R_{LW}, are input data. The average albedo is a linear combination of dry and wet soil albedos and leaf albedo. Only leaf albedo is a prognostic variable in ORCHIDEE, being a function of LAI. Heat storage in the soil, G, is optimized by adjusting a multiplier $\boldsymbol{K_{Csoil}}$, of the storage capacity, according to equation (A16) where T_{soil}, ρ_{soil} and C_{soil} are calculated by a seven layers model of thermal diffusion within the soil [*Warrilow et al.*, 1986].

$$G = \rho_{soil} . \boldsymbol{K_{Csoil}} . C_{soil} . (T_s - T_{soil}). \quad \text{(A16)}$$

[55] **Acknowledgments.** We thank J. Ogee, P. Friedlingstein, and N. de Noblet for helpful comments and discussions when interpreting the results, and B. Sportisse for help with the minimization problems. The Commissariat à l'Energie Atomique (CEA) funded the Ph.D. programme of D. Santaren, and provided the necessary computing resources. This work has been financed and supported by the EU project CAMELS, contract EVK2-CT-2002-00261, within the EU's 5th Framework Program for Research and Development.

References

Aubinet, M., et al. (2000), Estimates of the annual net carbon and water exchange of forests: The EUROFLUX methodology, *Adv. Ecol. Res.*, *30*, 113–175.

Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A model predicting stomatal conductance and its application to the control of photosynthesis under different environmental conditions, in *Progress in Photosynthesis*, edited by I. Biggins, pp. 221–224, Martinus Nijhoff, Zoetermeer, Netherlands.

Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Sensitivity analysis of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,481–19,490.

Botta, A., N. Viovy, P. Ciais, P. Friedlingstein, and P. Monfray (2000), A global prognostic scheme of leaf onset using satellite data, *Global Change Biol.*, *6*, 709–725.

Byrd, R. H., P. Lu, and J. Nocedal (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.*, *16*(5), 1190–1208.

Collatz, G. J., M. Ribas-Carbo, and J. A. Berry (1992), Coupled photosynthesis-stomatal conductance model for leaves of C_4 plants, *Aust. J. Plant Physiol.*, *19*, 519–538.

de Rosnay, P., J. Polcher, M. Bruen, and K. Laval (2002), Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes, *J. Geophys. Res.*, *107*(D11), 4118, doi:10.1029/2001JD000634.

Ducoudré, N. I., K. Laval, and A. Perrier (1993), SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model, *J. Clim.*, *6*, 248–273.

Falge, E., et al. (2001), Gap filling strategies for long term energy flux data sets, *Agric. For. Meteorol.*, *107*, 71–77.

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

Franks, S. W., and K. J. Beven (1997), Bayesian estimation of uncertainty in land surface-atmosphere flux predictions, *J. Geophys. Res.*, *102*(D20), 23,991–23,999.

Franks, S. W., K. J. Beven, P. F. Quinn, and I. R. Wright (1997), On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: Equifinality and the problem of robust calibration, *Agric. For. Meteorol.*, *86*, 63–75.

Giering, R., and T. Kaminski (1996), Recipes for adjoint code construction, *Tech. Rep. 212*, Max-Planck-Inst. für Meteorol., Hamburg, Germany.

Goulden, M. L., J. W. Munger, S.-M. Fan, B. C. Daube, and S. C. Wofsy (1996), Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability, *Science*, *271*, 1576–1578.

Ishida, A., A. Uemura, N. Koike, Y. Matsumoto, and A. L. Hoe (1999), Interactive effects of leaf age and self-shading on leaf structure, photosynthetic capacity and chlorophyll fluorescence in the rain forest tree, *dryobalanops aromatica*, *Tree Physiol.*, *19*, 741–747.

Johnson, I. R., and J. H. M. Thornley (1984), A model of instantaneous and daily canopy photosynthesis, *J. Theor. Biol.*, *107*, 531–545.

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.

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

Leuning, R. (2002), Temperature dependence of two parameters in a photosynthesis model, *Plant Cell Environ.*, *25*, 1205–1210.

McMurtrie, R. E., R. Leuning, W. A. Thompson, and A. M. Wheeler (1992), A model of canopy photosynthesis and water use incorporating a mechanistic formulation of leaf CO_2 exchange, *For. Ecol. Manage.*, *52*, 261–278.

Medlyn, B. E., D. Loustau, and S. Delzon (2002), Temperature response of parameters of a biochemically based model of photosynthesis, 1. seasonal changes in mature maritime pine, *Plant Cell Environ.*, *25*, 1155–1165.

Moncrieff, J. B., Y. Malhi, and R. Leuning (1996), The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water, *Global Change Biol.*, *2*, 231–240.

Ogée, J. (2000), Développement et applications du modèle MuSICA: étude des échanges gazeux d'eau et de carbone entre une pinède landaise et l'atmosphère, Ph.D. thesis, Univ. Paul Sabatier–Toulouse III, Toulouse, France.

Ogée, J., Y. Brunet, D. Loustau, P. Berbigier, and S. Delzon (2003), MuSICA, a CO_2, water and energy multilayer, multileaf pine forest model: Evaluation from hourly to yearly time scales and sensitivity analysis, *Global Change Biol.*, *9*, 697–717.

Parton, W., J. Stewart, and C. Cole (1988), Dynamics of C, N, P, and S in grassland soil: A model, *Biogeochemistry*, *5*, 109–131.

Porté, A., and D. Loustau (1998), Variability of the photosynthetic characteristics of mature needles within the crown of a 25-year old *Pinus pinaster*, *Tree Physiol.*, *18*, 223–232.

Raich, J. W., and C. S. Potter (1995), Global patterns of carbon dioxide emissions from soils, *Global Biogeochem. Cycles*, *9*(1), 23–36.

Raich, J. W., and W. H. Schlesinger (1992), The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate, *Tellus, Ser. B*, *44*, 81–99.

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.

Ruimy, A., L. Kergoat, C. B. Field, and B. Saugier (1996), The use of CO_2 flux measurements in models of the global terrestrial carbon budget, *Global Change Biol.*, *2*, 287–296.

Schulz, K., A. Jarvis, K. Beven, and H. Soegaard (2001), The predictive uncertainty of land surface fluxes in response to increasing ambient carbon dioxide, *J. Clim.*, *14*, 2551–2562.

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

Vukicevic, T. B., B. H. Braswell, and D. Schimel (2001), A diagnostic study of temperature controls on global terrestrial carbon exchange, *Tellus, Ser. B*, *53*, 150–170.

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

Warrilow, D., A. Sangster, and A. Slingo (1986), Modelling of land surface processes and their influence on European climate, report, Met Office, Bracknell, U.K.

Wesely, M., and R. Hart (1985), Variability of short term eddy-correlation estimates of mass exchange, in *The Forest-Atmosphere Interaction*, edited by B. A. Hutchinson and B. B. Hicks, pp. 591–612, Springer, New York.

White, M. A., P. E. Thornton, S. W. Running, and R. R. Nemani (2000), Parameterization and sensitivity analysis of the BIOME-BGC Terrestrial Ecosystem Model: Net primary production controls, *Earth Interact.*, *4*(3), 1–85.

Wilson, K., et al. (2002), Energy balance closure at FluxNet sites, *Agric. For. Meteorol.*, *113*, 223–243.

P. Ciais, P. Peylin, D. Santaren, and N. Viovy, Laboratoire des Sciences du Climat et de l'Environnement, Commissariat à l'Energie Atomique, Bat. 712, F-91191 Gif sur Yvette, France. (Philippe.Peylin@cea.fr)