

Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system

T. Ziehn,¹ W. Knorr,¹ and M. Scholze¹

Received 11 June 2010; revised 10 December 2010; accepted 22 March 2011; published 18 June 2011.

[1] Better estimates of the net exchange of CO₂ between the atmosphere and the terrestrial biosphere are urgently needed to improve predictions of future CO₂ levels in the atmosphere. The carbon cycle data assimilation system (CCDAS) offers the capability of inversion, while it is at the same time based on a process model that can be used independent of observational data. CCDAS allows the assimilation of atmospheric CO₂ concentrations into the terrestrial biosphere model BETHY, constraining its process parameters via an adjoint approach. Here, we investigate the effect of spatial differentiation of a universal carbon balance parameter of BETHY on posterior net CO₂ fluxes and their uncertainties. The parameter, β , determines the characteristics of the slowly decomposing soil carbon pool and represents processes that are difficult to model explicitly. Two cases are studied with an assimilation period of 1979 to 2003. In the base case, there is a separate β for each plant functional type (PFT). In the regionalization case, β is differentiated not only by PFT, but also according to each of 11 large continental regions as used by the TransCom project. We find that the choice of spatial differentiation has a profound impact not only on the posterior (optimized) fluxes and their uncertainties, but even more so on the spatial covariance of the uncertainties. Differences are most pronounced in tropical regions, where observations are sparse. While regionalization leads to an improved fit to the observations by about 20% compared to the base case, we notice large spatial variations in the posterior net CO₂ flux on a grid cell level. The results illustrate the need for universal process formulations in global-scale atmospheric CO₂ inversion studies, at least as long as the observational network is too sparse to resolve spatial fluctuations at the regional scale.

Citation: Ziehn, T., W. Knorr, and M. Scholze (2011), Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system, *Global Biogeochem. Cycles*, 25, GB2021, doi:10.1029/2010GB003886.

1. Introduction

[2] The quantification of terrestrial CO₂ sinks and sources and identification of the underlying processes are considered a prerequisite for meaningful projections of the future atmospheric CO₂ load [Prentice *et al.*, 2000, 2001]. Usually, sources and sinks are obtained by atmospheric transport inversion. This so-called “top-down” approach allows to obtain important insights into the large-scale patterns of the atmosphere-land flux. Numerical models are used to simulate the atmospheric transport, and the atmosphere-land fluxes are determined from the observational data by inversion [e.g., Bousquet *et al.*, 2000; Rödenbeck *et al.*, 2003]. However, due to the fact that the inverse problem is poorly conditioned (only few observations exist, but many possible flux fields are compatible with these observations [see, e.g., Kaminski and Heimann, 2001]), it is difficult to obtain a detailed flux pattern and regional fluxes are usually calculated instead. Another disadvantage of atmospheric inversions is that

without a process model, they do not allow predictions. In order to compare the many transport models and data used in atmospheric inversion studies, the TransCom experiments were started in the 1990s using a standardized set of 11 land and 11 ocean regions [e.g., Gurney *et al.*, 2008].

[3] Another way of determining spatial patterns of CO₂ sinks and sources is to apply forward runs of ecosystem models that represent the most important processes. A process-based terrestrial model can produce a very detailed flux pattern. Numerous terrestrial biosphere models exist and although they might differ in detail, they tend to have a similar structure and share a common base of process descriptions and global parameters. For example, many of the terrestrial biosphere models use a biochemical model of photosynthesis based on that of Farquhar *et al.* [1980] and they also include processes to model the energy and water balance, carbon balance and phenology. Examples of widely used biosphere models are HYBRID [Friend *et al.*, 1997], the Lund-Potsdam-Jena (LPJ) model [Sitch *et al.*, 2003] or ORCHIDEE [Krinner *et al.*, 2005]. However, this “bottom-up” approach cannot take into account the information contained in CO₂ measurements. Further, all process based models require extensive parameterization. If no reliable estimate exists of a parameter, it will remain highly uncertain and, depending on how

¹Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol, UK.

Carbon Cycle Data Assimilation System (CCDAS)

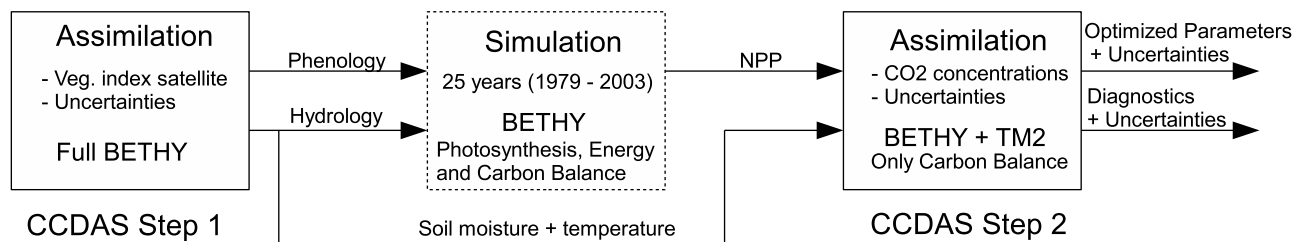


Figure 1. CCDAS structure.

sensitive the results are with respect to this parameter, uncertainty of those parameters might substantially contribute to the overall output uncertainty.

[4] With significant problems associated with both “bottom-up” and “top-down” approaches, one possible way out is the design of carbon cycle data assimilation systems, which use the observed data (for instance atmospheric CO₂ measurements) to systematically constrain ecosystem model parameters. Current approaches [Kaminski *et al.*, 2003; Rayner *et al.*, 2005; Scholze *et al.*, 2007] do not use any regionalization, but rely on globally applicable, universal parameters. That means, the parameters are not regionally differentiated, and if they are differentiated at all, then only by plant type.

[5] This paradigm of universality, however, deserves a closer look, because carbon fluxes might be determined by regional differences that are outside the realm of what the model represents. For example, if the model does not contain a land use change component and no information about the history of a site is available, then these unknown conditions can be subsumed under a common simplified formulation which requires parameterization. In such a case, we are back to the need of geographical differentiation that flux inversions need to follow inevitably. Such geographical differentiation is the subject of the present study.

[6] In this work we use the Carbon Cycle Data Assimilation System (CCDAS) [Rayner *et al.*, 2005] in order to determine a detailed pattern of the atmosphere-land fluxes and its uncertainties. CCDAS provides the possibility to consider both global and regional process parameters. Using the CCDAS framework, current fluxes of CO₂ into the atmosphere can be mapped together with optimal parameter values and their uncertainties. Those parameter uncertainties can also be propagated to any model output quantity to obtain its corresponding uncertainty. In this study, this will be applied to CO₂ fluxes.

2. Methodology

[7] The CCDAS used here is an estimator algorithm for a set of terrestrial biosphere model parameters, which uses automatically generated adjoint code (first derivative) for parameter optimization, and Hessian model code (second derivative) for estimating posterior parameter uncertainties. As its ecosystem model, CCDAS uses the Biosphere Energy Transfer and Hydrology scheme (BETHY) [Knorr, 2000]. This model simulates carbon assimilation and soil respira-

tion within a full energy and water balance and phenology scheme. Calculated fluxes are then mapped to atmospheric concentrations using the atmospheric transport model TM2 [Heimann, 1995].

[8] The CCDAS framework has been previously described in detail by Scholze [2003] and Rayner *et al.* [2005]. Therefore, we provide only a brief summary and highlight differences in our setup. The data assimilation is performed in two steps as outlined in Figure 1. In the first step, the full BETHY model is used to assimilate global monthly fields of the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) for optimizing parameters controlling soil moisture and phenology. In the second step, a reduced version of BETHY is used to assimilate atmospheric CO₂ concentration observations. In contrast to the setup used by Rayner *et al.* [2005], we only optimize the soil carbon part of BETHY in the second step, keeping all parameters controlling net primary productivity (NPP) fixed. In earlier studies with CCDAS [Rayner *et al.*, 2005; Scholze *et al.*, 2007], these parameters were found to be constrained relatively little by the assimilation of CO₂ observations. In practice, fixing NPP parameters is performed via an additional forward simulation over the integration period of 25 years immediately after the first assimilation step. We need to note that the uncertainties estimated in this study are only a lower bound because they do not account for the effect of the uncertainty of NPP related parameters on posterior uncertainties of the remaining parameters, or uncertainties of diagnostics. Instead of aiming for the most realistic uncertainty estimates of parameters and CO₂ fluxes, the present study aims at comparing two cases to test the effect of parameter regionalization on estimated fluxes and their posterior uncertainty covariance.

2.1. Data Assimilation

[9] In this work, we focus on the second assimilation step (see Figure 1), where the reduced BETHY version uses the NPP from the forward simulation and the soil moisture and temperature fields from the first assimilation step as input data. The atmospheric transport model then maps fluxes onto atmospheric concentrations for the atmospheric grid cells representing a list of remote monitoring stations (see Figure 2).

[10] CCDAS can be operated in three different modes. In the calibration mode, an optimal set of parameters is derived from atmospheric CO₂ concentration observations using an adjoint approach. The calibrated model can then be used for diagnostic simulations (over the calibration period) using the

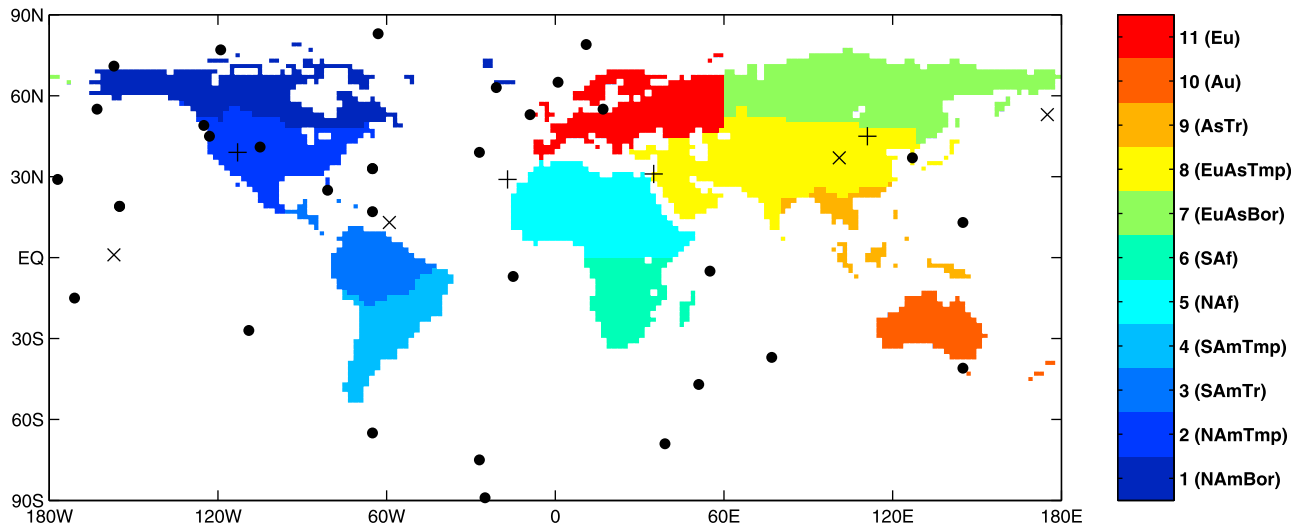


Figure 2. TransCom land regions and the location of the 41 observational sites used in CCDAS. Labels are given in Table 5. The four sites with the greatest improvement in match with observations using the regionalized model are marked with crosses (station codes: SHM, RPB, CHR, WLG) and the four sites where the fit has slightly worsened are marked with pluses (station codes: UUM, WIS, UTA, IZO).

optimal parameter set or for prognostic simulations (period subsequent to the calibration period). Here, we use CCDAS only for the optimization of the control parameters and for diagnostic simulations.

[11] The control parameters are optimized by calculating the mismatch of the observations and prior knowledge of the parameters via the following cost function:

$$J(\mathbf{x}) = \frac{1}{2} \left((M(\mathbf{x}) - \mathbf{c})^T \mathbf{C}_c^{-1} (M(\mathbf{x}) - \mathbf{c}) + (\mathbf{x} - \mathbf{p})^T \mathbf{C}_p^{-1} (\mathbf{x} - \mathbf{p}) \right) \quad (1)$$

where \mathbf{x} is the parameter vector and $M(\mathbf{x})$ the modeled concentrations. The covariance matrices \mathbf{C}_c and \mathbf{C}_p express the uncertainty for the observations \mathbf{c} and for the model priors \mathbf{p} , respectively (see *Rayner et al.* [2005] for further details). The optimization problem is thus formulated using a Bayesian approach [*Tarantola*, 1987, 2005]. A quasi-Newton method, the Davidon-Fletcher-Powell (DFP) formula [*Fletcher and Powell*, 1963; *Press et al.*, 1996], is used for the minimization of the cost function, which requires the calculation of the gradient of J with respect to the control parameters \mathbf{x} in each iteration. All derivative code is directly generated from the model's source code using the tool Transformation of Algorithms in Fortran (TAF) [*Giering and Kaminski*, 1998; *Kaminski et al.*, 2003].

2.2. Biosphere Model

[12] The process-based model of the terrestrial biosphere, BETHY, is run on a $2^\circ \times 2^\circ$ grid with 3462 land grid cells (excl. Antarctica). Global vegetation is mapped onto 13 different PFTs (see Table 1) and each grid cell can contain subareas (subgrid cells) with up to three different PFTs with their amount specified by each PFT's fractional cover. The dominant PFT in each grid cell is presented in Figure 3. In the present study, BETHY is driven by observed climate data over 25 years for the period 1979 to 2003. Following is a brief

outline of the parts of BETHY active during assimilation. A more detailed description can be found in the work of *Knorr* [2000] and *Knorr and Heimann* [2001].

[13] The net ecosystem productivity (NEP) in BETHY is computed as

$$\text{NEP} = \text{NPP} - R_S = \text{NPP} - (R_{S,s} + R_{S,f}) \quad (2)$$

where $R_{S,s}$ and $R_{S,f}$ are the respiration fluxes from the slowly and rapidly decomposing soil carbon pools, respectively, and NPP the net primary productivity. The size of the short-lived litter pool varies with time, whereas the size of the long-lived soil carbon pool is held constant through the simulation

Table 1. Optimal β Parameter for Each of the 13 PFTs for Both Case Studies^a

PFT		β (N = 19)	β (N = 117)	NPP
1	Tropical broadleaved evergreen tree (TrEv)	0.87 (92)	0.91	13 892
2	Tropical broadleaved deciduous tree (TrDec)	1.09 (85)	1.40	6 648
3	Temperate broadleaved evergreen tree (TmpEv)	0.53 (29)	0.79	258
4	Temperate broadleaved deciduous tree (TmpDec)	0.51 (63)	1.83	2 395
5	Evergreen coniferous tree (EvCn)	1.53 (94)	1.37	5 466
6	Deciduous coniferous tree (DecCn)	0.62 (42)	1.35	732
7	Evergreen shrub (EvShr)	0.20 (43)	0.74	2 279
8	Deciduous shrub (DecShr)	11.33 (91)	1.39	934
9	C3 grass (C3Gr)	0.55 (73)	1.03	10 908
10	C4 grass (C4Gr)	0.93 (92)	0.58	18 079
11	Tundra vegetation (Tund)	0.19 (47)	0.75	1 443
12	Swamp vegetation (Wetl)	1.50 (1)	1.03	361
13	Crops (Crop)	0.35 (60)	1.00	4 593

^aThe relative reduction of the parameter uncertainty ($+1\sigma$) is given in % in brackets below (i.e. for the base case $N = 19$). For the regionalization case ($N = 117$), the β values are recalculated via NPP and NEP for comparison. The total mean NPP (TgC yr^{-1}) is also provided for each PFT.

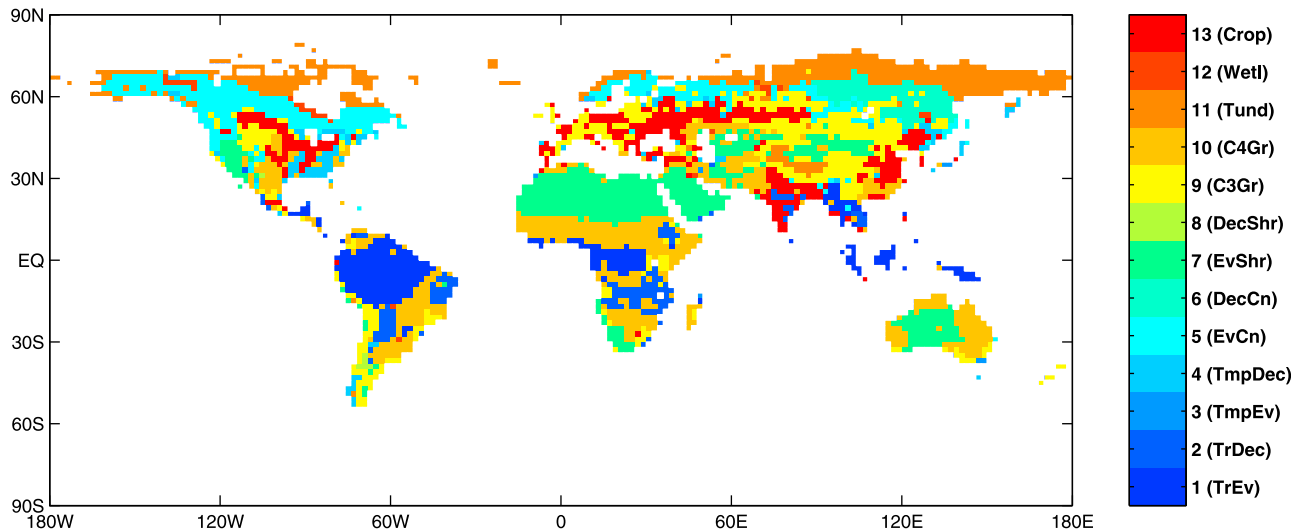


Figure 3. Distribution of the dominant PFT per grid cell. PFT labels are given in Table 1.

period [Knorr, 2000]. Soil respiration is calculated from the following equations:

$$R_{S,f} = (1 - f_s)k_f C_f \quad (3)$$

$$R_{S,s} = k_s C_s \quad (4)$$

with C_f and C_s representing the size of the fast and slow carbon pool, respectively, and f_s the fraction of decomposition from the fast pool that goes into the long-lived (slow) soil carbon pool. The rate constants are

$$k_f = w^\kappa Q_{10,f}^{T_a/10} / \tau_f \quad (5)$$

$$k_s = w^\kappa Q_{10,s}^{T_a/10} / \tau_s \quad (6)$$

where w is plant-available soil moisture divided by field capacity (a value between 0 and 1), T_a air temperature, κ a parameter describing linearity of soil moisture dependence, $Q_{10,f}$ and $Q_{10,s}$ temperature dependence parameters and τ_f and τ_s the pool turn over times at 0°C. Controlling parameters are: f_s , κ , $Q_{10,f}$, $Q_{10,s}$ and τ_f . The turnover time τ_s for the slow carbon pool is determined indirectly via the long-term (25 year in this case) carbon balance at each subgrid cell.

[14] The motivation for this latter procedure is as follows. In a normal forward model run, the soil carbon pools must be spun up until respiration from these pools comes into equilibrium with NPP. Here, this is done for the fast carbon pool using 100 years of spin-up. The spin-up year is the average over the 25 year input meteorology and used continuously over the 100 years with a CO₂ concentration of 338 ppm. Because of its fast depletion, decomposition rates of the fast pool are sensitive to changes in pools size (C_f) in addition to changes in the turnover rate (k_f). This size effect (also called substrate limitation) can be neglected for the slow carbon pool. This makes it possible to estimate the size of the pool on the basis of the long-term carbon balance at each subgrid cell location.

[15] This is important, because with limited knowledge about a particular location's history of soil carbon disturbance, the assumption of a balance between NPP and soil respiration provides a more viable and universal first estimate than any assumption about absolute sizes of the slow carbon pool. Further, changes in the size of organic carbon pools have been linked to the climatic conditions of a site, with colder sites possessing on average larger carbon pools [Bird *et al.*, 2002]. Because these would have a lower k_f , the balance condition $\text{NPP} = R_{S,s} + R_{S,f}$ will automatically reproduce this basic relationship as long as $Q_{10,f}$ and $Q_{10,s}$ are greater than 1.

[16] We generalize the equilibrium condition, introduced by Knorr [2000], to nonequilibrium situations by introducing a carbon balance parameter, β , in the following way [Rayner *et al.*, 2005]:

$$\beta \overline{\text{NPP}} = (\overline{R_{S,s}} + \overline{R_{S,f}}) \quad (7)$$

which is equivalent to

$$\overline{\text{NEP}} = \overline{\text{NPP}}(1 - \beta) \quad (8)$$

The vertical line above signals temporal average over the entire simulation period of 25 years, and the equation applies at each subgrid cell. The carbon balance parameter, β , determines whether a site acts as a long-term source ($\beta > 1$, negative NEP) or a long-term sink ($0 < \beta < 1$) of CO₂. The case $\beta < 0$ is not allowed, as it would entail negative respiration fluxes for positive NPP. The positivity of β is guaranteed by applying a lognormal transformation. With the assumption of a constant C_s , we can introduce equations (4) and (3) into equation (7) and obtain an expression for the size of the slow carbon pool as

$$C_s = (\beta \overline{\text{NPP}} - \overline{R_{S,f}}) / \overline{k_s} \quad (9)$$

Note that this definition of β is the reciprocal of the definition used in previous CCDAS studies. The change was implemented in order to prevent division by small numbers as β approaches 0.

Table 2. Initial and Optimal Control Parameters for the Reduced BETHY Model^a

Parameter	Initial Value	Prior Uncertainty	Base Case (N = 19)		Regionalization (N = 117)	
			Optimal Value	Optimal Uncertainty	Optimal Value	Optimal Uncertainty
$Q_{10,f}$	1.50	-0.500; +0.750	1.20	-0.025; +0.025	1.46	-0.037; +0.038
$Q_{10,s}$	1.50	-0.500; +0.750	1.69	-0.020; +0.020	1.62	-0.024; +0.025
τ_f	1.50	-1.000; +3.000	4.86	-0.218; +0.228	9.63	-0.637; +0.682
κ	1.00	-0.900; +9.000	0.62	-0.010; +0.011	0.57	-0.011; +0.012
f_s	0.20	-0.100; +0.200	0.73	-0.004; +0.004	0.69	-0.010; +0.010
β	1.00	-0.200; +0.250	see Table 3		see Table 1	
offset	338.00	1.000	336.32	0.015	336.53	0.017

^aAll parameters are unitless, but τ_f in years. Uncertainties represent one standard deviation. For parameters with a lognormal distribution, upper and lower percentiles equivalent to one standard deviation are given.

[17] The advantage of this approach is that we can subsume a number of often unknown conditions that lead to changes in the carbon balance, such as removal of organic carbon by land use, under a common simplified formulation. By starting from a first guess of equilibrium, expressed by the same value of $\beta = 1$ everywhere, the optimization of CCDAS can then proceed to providing constraints on β via the atmospheric CO₂ network. The main difficulty with this approach is determining how best to differentiate β spatially. Too fine a spatial resolution would make the inverse problem again underdetermined, returning to the problems faced by atmospheric inversion studies. What resolution will be acceptable before the problem becomes underdetermined will depend on the resolution of the transport model as well as on the biosphere model used. Essentially, a finer spatial differentiation of parameters would have to be tested through optimization and computation of the error covariance matrix of the optimal parameters. We expect that with increasing numbers of parameters, more of these will show high error correlation (error covariance divided by standard deviation of both parameters). This can then be used to judge what resolution and spatial differentiation will be acceptable.

[18] In previous CCDAS studies, β was differentiated by PFT only, assuming that similar plant types would exist in similar ecosystems, and that these would have undergone similar disturbance regimes leading to similar ratios between NPP and total soil respiration. This was essentially a pragmatic approach that avoided differentiation purely by location, something that would have meant defining as many control parameters in CCDAS representing β as there are subgrid cells. Here, we investigate the consequences of this assumption by introducing an additional layer of geographic differentiation of β on top of the PFT dependence in the base case. As a full geographic differentiation to the grid cell levels is not feasible, we have opted for using the established 11 TransCom regions.

[19] Hence, this study investigates two cases: (1) one β parameter per 13 PFTs, and (2) separate β parameters for each PFT within each TransCom region. The total number of β parameters in the second case is 111 (i.e. less than 11 times 13 because not all PFTs exist in all regions). The 11 TransCom regions are shown in Figure 2, the list of PFTs in Table 1 and the remaining control parameters in Table 2. In addition to the β parameters and the global parameters there is one additional parameter representing the global atmospheric CO₂ concentration at the beginning of the optimization period (offset). Consequently, the base case has 19 (13 β s + 5 global

parameters + 1 offset) and the regionalization case has 117 control parameters (111 β s + 5 global parameters + 1 offset).

2.3. Background Fluxes

[20] The focus of this study are the natural land-atmosphere fluxes. Therefore, land use change is not directly included in the BETHY model but specified as an external flux. As described by *Rayner et al.* [2005] we use the estimates of *Houghton* [2008] for the land use flux without seasonality or interannual variability. Background fluxes for fossil fuel emissions are based on the flux magnitudes from *Boden et al.* [2009] as described by *Scholze et al.* [2007]. The flux pattern and magnitude of ocean CO₂ exchange is taken from *Takahashi et al.* [1999] with estimates of inter annual variability taken from *Le Quéré et al.* [2007]. We use the same observational network of 41 stations as described by *Rayner et al.* [2005] (see also Figure 2).

2.4. Parameter Uncertainties

[21] We assume that all input probability density functions follow a Gaussian distribution. Hence, our a priori information consists of the mean and covariance matrix, which represents the uncertainties, for the observations and control parameters. If the model is linear, the posterior probability density will also be Gaussian. If the model is nonlinear (as it is in our case), it can still be linearized around the prior parameter values and the posterior probability function approximated by a Gaussian [*Tarantola*, 1987].

[22] The second order derivative of the cost function is represented by the Hessian, which graphically speaking describes the curvature of the cost function. At the cost function minimum, the Hessian approximates the inverse covariance of the optimal parameters and in this way provides the *a posteriori* parameter uncertainties. The procedure on how the Hessian is calculated and how the *a posteriori* uncertainties are derived is described in detail by *Rayner et al.* [2005].

3. Results

[23] An overview of the optimization results are shown in Table 3. The very small gradient of the cost function indicates success of the cost function minimization. However, running an ensemble of optimizations gives us further indication of the robustness of the solution. For the base case, we performed a set of 20 optimizations by varying the starting point randomly. Half of the optimizations finished in the same minimum we found in the first case. The other half finished in

Table 3. Cost Function Values J and Number of Iterations for Both Case Studies^a

	Base Case ($N = 19$)	Regionalization ($N = 117$)
J	9020	7594
J_o	8874	7014
J_p	146	580
χ^2	1.47	1.25
Iterations	738	1577
Gradient	8.1×10^{-4}	7.7×10^{-4}

^a J_o stands for the mismatch of the observations and J_p for the parameter mismatch. χ^2 is the reduced chi-squared test, where values close to one indicate statistical agreement between model and data.

a minimum where at least one global parameter had a non-physical value (e.g. $f_s > 1$), which is not relevant here. This strengthens our confidence of having found a global minimum within the physical parameter space. After the calibration, the cost function value J was reduced by a factor of about 600 for both cases. Also, with $J = 7594$ the optimal value of the cost function is considerably smaller for the regionalization than for the base case ($J = 9020$). This is expected since we have a larger number of parameters ($N = 117$) in case of the regionalization, which increases the degree of freedom for the optimization. The fit to the observations is also considerably improved for the regionalization case as can be seen by the lower value for J_o . For 37 out of the 41 monitoring stations used in this study we notice an improved fit to the observations. The four sites with the greatest improvement in match with observations and the four sites where the fit has slightly worsened are marked in Figure 2. The reduced χ^2 is also smaller for the regionalization case. Values close to one indicate that model and data are statistically indistinguishable. In none of the cases it is less than one, indicating that there are still some model deficiencies that prevent a statistically complete match to the observations. However, a higher number of degrees lead to values of χ^2 much closer to the threshold of 1.

[24] The parameter mismatch, expressed by J_p , is higher for the regionalization case, mainly due to the larger dimension of the control parameter space. 738 iterations are required in the base case in order to find the minimum of the cost function, while about twice as many iterations are required for the regionalization case. It appears that a larger number of control parameters somewhat increases search time, even though not dramatically given the much higher dimensionality. In contrast to the study by *Rayner et al.* [2005], the Hessian is positive definite at the cost function minimum for both cases, which further strengthens our confidence in having found the exact minimum in the cost function.

[25] The remaining part of this section focuses on the optimized control parameters, their *a posteriori* uncertainties, and the long-term mean CO₂ fluxes together with their uncertainty covariance matrix. Special attention is paid to the carbon balance parameter β . Parameter uncertainties are propagated in order to determine uncertainty ranges for the predicted net CO₂ fluxes for each of the 11 TransCom regions.

3.1. Optimal Parameters

[26] Prior and optimized parameter values for both cases, base case and regionalization, are presented in Table 2. The temperature sensitivity of the slow carbon pool respiration,

$Q_{10,s}$, is somewhat increased in both (1.69 and 1.62, respectively) compared to the prior value of 1.5. This change is, however, within range of the prior parameter uncertainty. The temperature sensitivity of the respiration of the fast carbon pool, $Q_{10,f}$, is reduced from its initial value of 1.5 to 1.2 in the base case, but remains close to its initial value for the regionalization case. The change is again within the range of the prior uncertainties. The posterior uncertainties of the two parameters are reduced by more than one order of magnitude, confirming the result of *Scholze et al.* [2007] that the parameters of soil respiration are well constrained by atmospheric CO₂ data. Here the constraint is even higher because we neglect the uncertainty in NPP.

[27] The soil moisture dependence parameter κ is reduced in both scenarios, from its initial value of 1.0 to 0.62 and 0.57 respectively, meaning a reduced sensitivity to soil moisture when the same is close to field capacity ($w = 1$), but an increased sensitivity when soil moisture is close to the wilting point ($w = 0$). The prior uncertainty of κ , however, is much larger than the change. We find that κ is also well constrained by the data, shown by the small posterior uncertainty. The optimized parameter values for the fast pool turnover time, τ_f , however, are both outside the prior uncertainty range defined by one standard deviation. For the regionalization case, the change is by more than two standard deviations from 1.5 to 9.63 years. (Note that because of the lognormal distribution, two standard deviation from the prior is equivalent to 9 years.) The fraction f_s of the decomposition flux going from the fast to the long-lived soil carbon pool also increases by much more than its prior uncertainty for both cases, but it is very similar between the two. The posterior uncertainty is again very small. Finally, the offset parameter behaves similar to previous studies [*Scholze et al.*, 2007]. In general, posterior uncertainties for all global parameters in Table 2 (all but the β s) are reduced by more than 90% compared to prior uncertainty ranges. This is partly a result of the fairly large prior uncertainty estimates we used, but can also be explained by the fact that parameters that act globally at all subgrid cells are well observed by the global atmospheric CO₂ network.

[28] The optimal values for the soil carbon balance parameter β are given in Table 1 for the base case. As a reminder, β determines whether subgrid cells occupied by the corresponding PFT act as a long-term carbon source ($\beta > 1$) or a long-term carbon sink ($\beta < 1$), independent of the geographic region. As shown in Table 1, most (9 out of 13) PFTs act as a sink. This is simply the result of the fact that the atmospheric CO₂ increase is less than expected from the total anthropogenic CO₂ emissions (the airborne fraction is less than one [*Knorr*, 2009]) and that the oceans take up only approximately half of the excess [*Bopp et al.*, 2002]. Both ocean fluxes and anthropogenic emissions are here implemented as part of the background fluxes. What is interesting in the context of the present study, however, is how this overall sink is distributed spatially.

[29] Here, we find that the optimal parameter β for PFT 8 (deciduous shrub) is extremely large ($\beta_8 = 11.33$), which means that the net flux, NEP, is more than 10 times that of NPP (see equation (8)). This PFT, however, has only a very small total NPP (as seen for all TransCom regions in Figure 7) and occurs only in a few marginal areas as dominant PFT (see Figure 3). The only way the optimization “knows” about the limitation of soil respiration by NPP, however, is via the prior

Table 4. Optimal β Parameter for Each of the 13 PFTs and 11 TransCom Land Regions^a

	PFT	Region										
		NAmBor	NAmTmp	SAmTr	SAmTmp	NAF	SAf	EuAsBor	EuAsTmp	AsTr	Au	Eu
1	TrEv	-	0.85 (8)	1.07 (85)	0.51 (26)	0.44 (31)	2.08 (63)	-	0.45 (28)	0.23 (42)	0.91 (5)	-
2	TrDec	-	0.74 (15)	0.83 (17)	2.89 (74)	0.74 (16)	0.38 (41)	-	13.91 (84)	0.45 (29)	1.00 (0)	-
3	TmpEv	-	0.88 (6)	-	-	1.15 (0)	0.85 (9)	-	0.75 (14)	-	0.70 (21)	1.02 (0)
4	TmpDec	0.89 (6)	0.41 (38)	0.99 (1)	33.11 (79)	0.96 (2)	-	0.78 (13)	0.49 (26)	0.95 (3)	1.08 (1)	1.33 (3)
5	EvCn	1.56 (82)	1.51 (85)	1.61 (0)	1.13 (0)	-	-	1.63 (35)	0.45 (28)	0.66 (18)	1.02 (1)	0.93 (37)
6	DecCn	-	-	-	-	-	-	1.35 (32)	-	-	-	-
7	EvShr	1.41 (1)	0.70 (23)	1.08 (0)	1.06 (0)	1.24 (3)	1.03 (2)	0.77 (15)	0.72 (33)	1.01 (0)	0.34 (34)	0.83 (10)
8	DecShr	1.15 (0)	0.83 (10)	0.99 (1)	2.17 (2)	1.94 (0)	1.02 (0)	0.77 (14)	1.13 (1)	0.97 (2)	0.86 (8)	0.76 (13)
9	C3Gr	0.68 (20)	0.70 (37)	1.06 (3)	0.59 (33)	7.28 (45)	0.53 (26)	0.97 (22)	0.66 (54)	0.51 (25)	0.61 (41)	1.02 (57)
10	C4Gr	0.97 (2)	1.22 (74)	0.56 (31)	0.25 (42)	0.49 (46)	0.49 (38)	0.99 (1)	0.43 (35)	0.28 (37)	1.25 (81)	0.49 (27)
11	Tund	1.17 (32)	1.01 (0)	-	0.98 (1)	-	-	0.57 (42)	0.89 (6)	-	1.01 (0)	0.88 (16)
12	Wetl	1.23 (0)	-	0.97 (1)	1.05 (0)	-	1.14 (0)	0.85 (8)	1.01 (0)	-	-	1.02 (0)
13	Crop	0.95 (3)	0.34 (36)	1.03 (0)	0.91 (6)	0.90 (6)	22.29 (63)	0.54 (24)	0.80 (47)	0.49 (26)	0.88 (8)	0.75 (25)
	Area	8 872	10 799	9 439	8 876	19 512	9 446	13 341	23 910	4 670	7 541	9 332

^aPFT, Plant function type. Area in 10^3 km², for the regionalization, $N=117$. The relative reduction of the parameter uncertainty ($+1\sigma$ from a prior value of 1 and an uncertainty range from 0.8 to 1.25) is given in brackets in %. For an explanation of the abbreviations used for the PFTs and the land regions refer to Tables 5 and 1.

value of β for PFT 8 and its effect on the cost function. Obviously, the penalty for changing β_8 is not large enough to outweigh the benefit from placing a source in the region of this PFT.

[30] Particularly strong sinks relative to their NPP, with soil respiration only between 20 and 35% of their NPP, are evergreen shrubs (PFT 7), tundra (PFT 11) and crops (PFT 13). Temperate trees (PFTs 3 and 4) are also a rather strong sink, while evergreen conifers (PFT 5), situated in the large boreal forests, appear as a source. We also note that the PFT-specific β is generally well constrained by the CO₂ data except for PFT 12 (wetlands). The uncertainty reduction is, however, often much less than 90%, and thus considerably less than for the global parameters.

[31] For the regionalization case, there is no control parameter β that covers all of a given PFT. However, the average value for β can be inferred from equation (8) using the average NEP and NPP of a given PFT. This is also shown in Table 1. The two cases approximately agree for two important PFTs (with large global NPP), namely tropical evergreen trees (PFT 1) shown as a sink, and evergreen conifers (PFT 5), predominantly found in boreal forests, shown as a source. There are, however, large differences between the two cases for a range of PFTs, in particular deciduous shrubs (PFT 8), temperate deciduous trees (PFT 4), and deciduous conifers (PFT 6). As evident from Figure 3, these PFTs cover relatively small areas globally. Crops (PFT 13) and C3 grass (PFT 9), which are more dominant and appear as a significant sink of CO₂ for the base case, appear neutral for the regionalization case. For PFTs 4, 6 and 9 the regionalization suggests a source, while the base case suggests a sink for the same PFTs.

[32] Table 4 shows the optimized values of β for the regionalization case. Two main results stand out. First, the reduction in uncertainty of the regionalized parameter is now much less than when the same parameter is distinguished only by PFT. The information gained from the observations is now “spread” over many more individual parameters, where some can be observed better than others. As in the base case, all β parameters associated with PFT 12 show no or hardly any reduction in their uncertainties. Second, there is a considerable spread across regions, where β for one given PFT (apart

from PFT 6 which only occurs in Region 7) takes on widely differing values. This already indicates that the optimal value of β is rather sensitive to the way it is differentiated geographically. In fact, we observe that the same PFT can act as a sink ($\beta < 1$) or a source ($\beta > 1$) depending on the region where they occur. For example, tropical evergreen trees (PFT 1) shown to be a sink overall in Table 1, ranges from an extremely strong sink in Region 9 (tropical Asia), to a strong source in Region 6 (Africa south of the equator). Because PFT 1 is centered around the equator itself, the division of central Africa into two regions leads to an interesting result: north of the equator we have a source ($\beta = 2.08$), and south of equator a sink ($\beta = 0.44$). Similarly, evergreen conifers (PFT 5) appear as a source in boreal Eurasia (Region 7) but as a sink in temperate Eurasia (Region 8). This phenomenon will be revisited later in the present analysis.

[33] We also find some very large values for β in a few cases. For example, the optimal value for PFT 4 (temperate broadleaved deciduous tree) in Region 4 (temperate South America) is 33.11, for PFT 13 (Crops) in Region 6 (Southern Africa) β is 22.29 and for PFT 2 (tropical broadleaved deciduous tree) in Region 8 (Eurasian temperate) we find a value of 13.91. Such extreme values are unlikely from a carbon balance point of view, as discussed before for the base case.

3.2. Fluxes

[34] The previous analysis, which focused on the spatial differentiation of β , has already identified certain trends and patterns, for example a source of CO₂ from boreal conifers and a tropical sink. The same can be found when analyzing NEP for individual regions, as shown in Table 5. Regions 1 and 7 (boreal North America and Eurasia) act as a source (negative NEP), while Regions 3 and 9 (tropical South America and Asia), as well as the sum of Regions 5 and 6 (Africa) consistently appear as a sink. Both cases also find a strong source in Region 4 (temperate South America) and a strong sink in Region 2, which despite its name (temperate North America) contains tropical vegetation (PFTs 1 and 2) in central America (see Table 4). There are, however, a number of noteworthy differences between the two cases. First, Africa is only a slight sink in the base case, but a moderate sink in

Table 5. Mean NEP for Each of the 11 TransCom Land Regions in TgC Per Year and the Optimal (σ_{opt}) and Prior (σ_{prior}) Uncertainties As Well As the Reduction in Uncertainty

TransCom Region	N = 19				N = 117			
	NEP	σ_{opt}	σ_{prior}	Reduction	NEP	σ_{opt}	σ_{prior}	Reduction
1 North American boreal (NAmbor)	-881	39	678	94%	-1 074	93	695	87%
2 North American temperate (NAmTmp)	901	34	527	94%	846	67	578	88%
3 South American tropical (SAmTr)	892	96	1517	94%	304	151	1818	92%
4 South American temperate (SAmTmp)	-1 555	37	1020	96%	-1 637	128	989	87%
5 Northern Africa (NAf)	-97	45	962	95%	250	243	898	73%
6 Southern Africa (SAf)	284	68	917	93%	648	167	817	80%
7 Eurasian boreal (EuAsBor)	-376	52	561	91%	-187	112	490	77%
8 Eurasian temperate (EuAsTmp)	1 154	43	555	92%	-562	142	846	83%
9 Tropical Asia (AsTr)	541	25	517	95%	3032	94	151	38%
10 Australia (Au)	346	33	441	93%	137	94	559	83%
11 Europe (Eu)	991	42	476	91%	436	129	572	78%

the regionalization case. Second, other than temperate North America, temperate Eurasia appears either as a strong sink (base case), or a moderate source (regionalization). In the latter case, the source over boreal Eurasia is much reduced.

[35] An analysis of the net CO₂ flux at the grid cell level of BETHY, shown in Figure 4 for the base case and in Figure 5 for the regionalization, reveals large geographical fluctuations between adjacent grid cells, or between smaller regions (i.e. smaller than the TransCom regions). This phenomenon is evident in both cases, but it is particularly pronounced for the regionalization case in all of the tropics and subtropics, except for northern South America and Africa north of the Sahel. For example, for Region 4 (temperate South America), which is identified as the strongest source with around 1 600 TgC per year within the 11 land regions for both cases, the geographical flux patterns look very different between the two cases. While in the base case, they are either strong sources or only slight sinks, in the regionalization case the area dominated by PFT 10 (C4 grass) turns out as a strong sink and the small area dominated by PFT 4 (temperate broadleaved deciduous tree) turns out as a very large source relative to its

size (see also Figure 3 for the dominant PFT cover). This is the result of a small carbon balance parameter ($\beta = 0.25$) in this region for PFT 10, compensated for by a very large value ($\beta = 33.11$) for PFT 4. In contrast, in the base case, PFT 10 is close to neutral ($\beta = 0.93$) and PFT 4 is a sink ($\beta = 0.51$). As a consequence, the source characteristic of Region 4 is determined by a different PFT, namely PFT 8 (deciduous shrub) which has a β value of 11.33. It appears that the optimization is using its freedom to modify regional fluxes via β , thus creating pronounced alternating source and sink pattern in associated areas below the scale of a TransCom region.

[36] Nearly the same phenomenon as just described is observed for Africa (Regions 5, north, and 6, south). In the base case, all grid cells show comparatively small fluxes, but for the regionalization case there is a large spread in the absolute size of the fluxes between individual grid cells. Differences in the flux direction between cases also exist for Region 8 (temperate Eurasia), where in the regionalization case a strong source region in India turns what is a large sink in the base case into a relatively large overall source (see Table 5). The largest difference in the mean flux between the

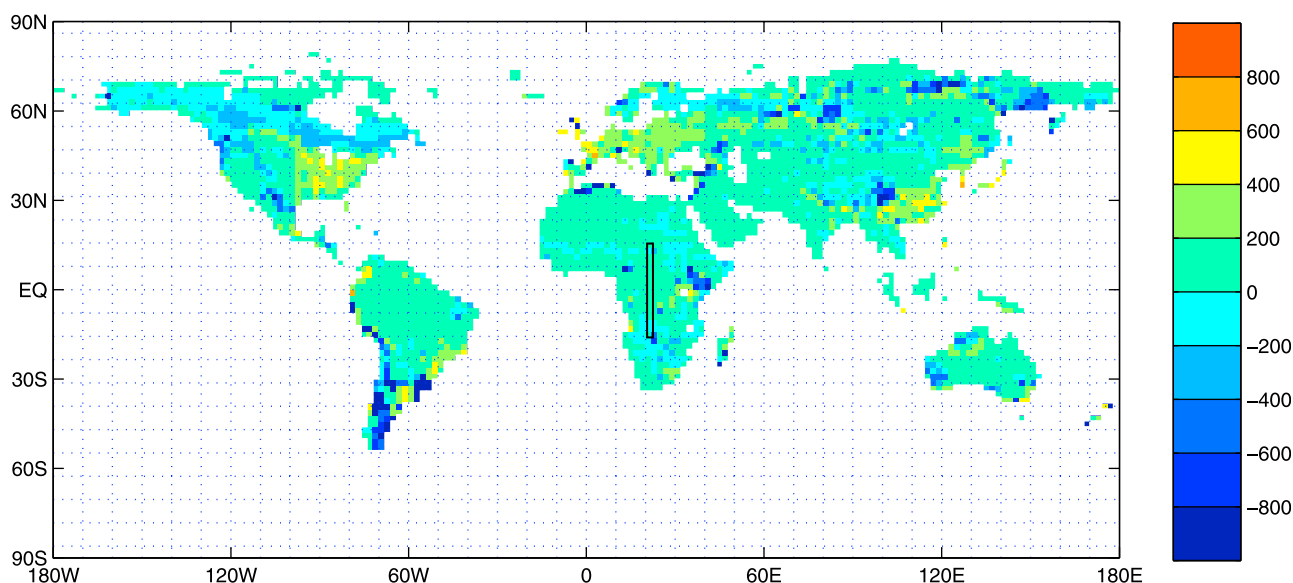


Figure 4. Mean annual net CO₂ flux to the atmosphere for the period 1979–2003 ($\text{gC m}^{-2} \text{yr}^{-1}$) for the base case ($N = 19$ parameters). Negative NEP indicates a source. The black rectangle denotes the transect for central Africa analyzed in Figure 8. The underlying grid represents the resolution of the TM2 transport model.

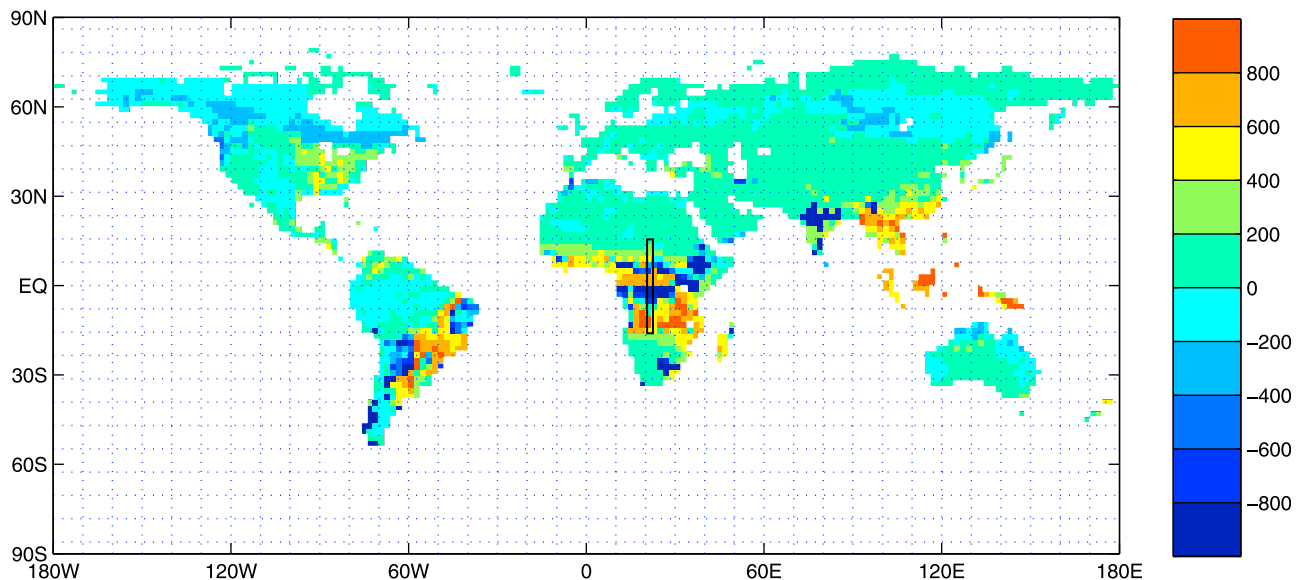


Figure 5. Mean annual net CO₂ flux to the atmosphere for the period 1979–2003 (gC m⁻² yr⁻¹) for the regionalization case ($N = 117$ parameters). Negative NEP indicates a source. The black rectangle denotes the transect for central Africa analyzed in Figure 8. The underlying grid represents the resolution of the TM2 transport model.

two cases exist for Region 9 (Tropical Asia). A very large sink is identified for the regionalization with about 3000 TgC yr⁻¹ and a much smaller sink with about 500 TgC yr⁻¹ for the base case.

3.3. Uncertainty Propagation

[37] In the following analysis, we will make use of CCDAS’ feature of propagating uncertainties from control parameters forward to obtain uncertainties and their covariances for diagnostic quantities, here the net carbon balance. The result is shown in Table 5 for the 11 TransCom regions. It must be noted that the global net flux is well constrained because we neglect any uncertainties in both the anthropogenic emissions and the ocean fluxes. As a result, the uncertainty of the posterior global NEP is only determined by the uncertainty in the atmospheric CO₂ concentrations, and the reduction in uncertainty becomes close to 100%. Also note again that the posterior uncertainties neglect the impact of uncertainties in NPP. The analysis will focus more on a comparison between the two cases rather than putting undue emphasis on absolute values.

[38] By region, the posterior uncertainty varies considerably between the two cases, and is approximately three times larger on average for the regionalization case. The additional freedom the optimization has with the regionalized β obviously translates into less constrained net fluxes.

3.3.1. Uncertainty Covariance for the 11 Land Regions

[39] In addition to comparing the reduction in the uncertainty of the net flux, we also consider the covariance between flux uncertainties. This is expressed via the uncertainty correlation matrix of diagnostics, \mathbf{R}_d , which is defined as follows:

$$R_d^{i,j} = \frac{C_d^{i,j}}{\sigma_i \sigma_j} \quad (10)$$

where $C_d^{i,j}$ is element i, j of the uncertainty covariance matrix of the diagnostics (NEP), and σ_i the posterior uncertainty of parameter i derived from the diagonal elements $C_d^{i,i}$ of the matrix \mathbf{C}_d .

[40] We illustrate two extreme but typical cases. In one, NEP of two adjacent grid cells have exactly the same impact on modeled atmospheric concentrations. Further, both NEP values are controlled by completely separate sets of parameters. Here, a change in NEP of grid cell 1 by any amount a can be compensated by a change of NEP in grid cell 2 without a change in the data part of the cost function, J_o (disregarding J_p). In this case, NEP uncertainties of the two cells are anticorrelated and $R_d^{1,2} < 0$. In the other case, NEP of cells 1 and 2 are controlled by the same parameters in the same way. Here, changes in these parameters in any direction will result in the same change in NEP in both cells and as a result the NEP uncertainties of both cells are positively correlated ($R_d^{1,2} > 0$). Other cases are also possible: a change in NEP in one cell is compensated by a change in the same direction in NEP of the other (usually not adjacent) grid cell, or the same parameters have an impact on NEP that is of opposite sign between the two grid cells. Such cases, however, are less typical of the CCDAS setup. For this analysis this usually means that positive values of $R_d^{i,j}$ approaching +1 indicate that the NEP of the two cells is modeled concurrently by BETHY, and large negative values approaching -1 that atmospheric transport cannot distinguish between the two fluxes.

[41] Figure 6 shows the correlation matrix for the NEP of the 11 TransCom regions for the base case and the regionalization case. It is evident that in the base case, which “binds” together distant regions by common values of β , the NEP of some pairs of regions are highly positively correlated, in particular Regions 3 and 9 (tropical South America and tropical Asia) and Regions 5 and 10 (North Africa and Australia). As Figure 7 shows, Regions 3 and 9 stand out as having the largest NPP contribution from PFT 1 (tropical

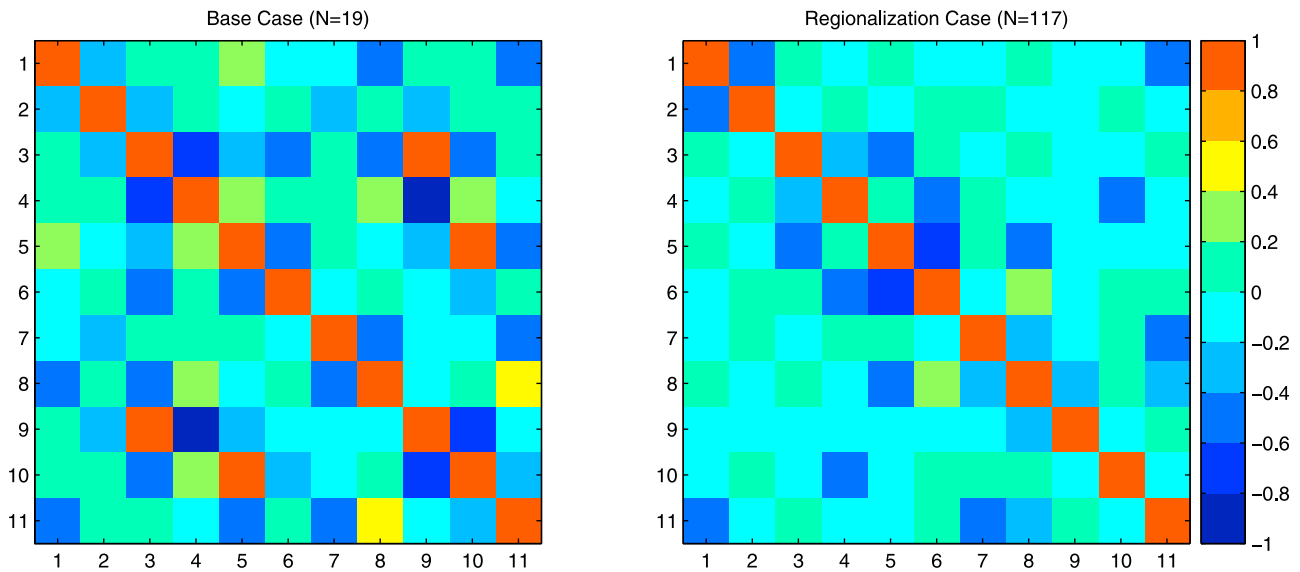


Figure 6. Uncertainty covariance matrix of annual mean NEP per TransCom region for the base case ($N = 19$ parameters) and the regionalization case ($N = 117$ parameters).

evergreen trees), whereas Regions 5 and 10 have both a large contribution from PFT 10 (C4 grass). Region 4 also has a large contribution from PFT 10, and indeed its uncertainty in NEP is positively correlated with that of Region 5 and 10. A

large positive correlation is also found for Regions 8 and 11, with both strong NPP contributions from PFTs 9, 10 and 13. [42] The reason we assume these correlations are due to common value of β is that the same correlations are all negative in

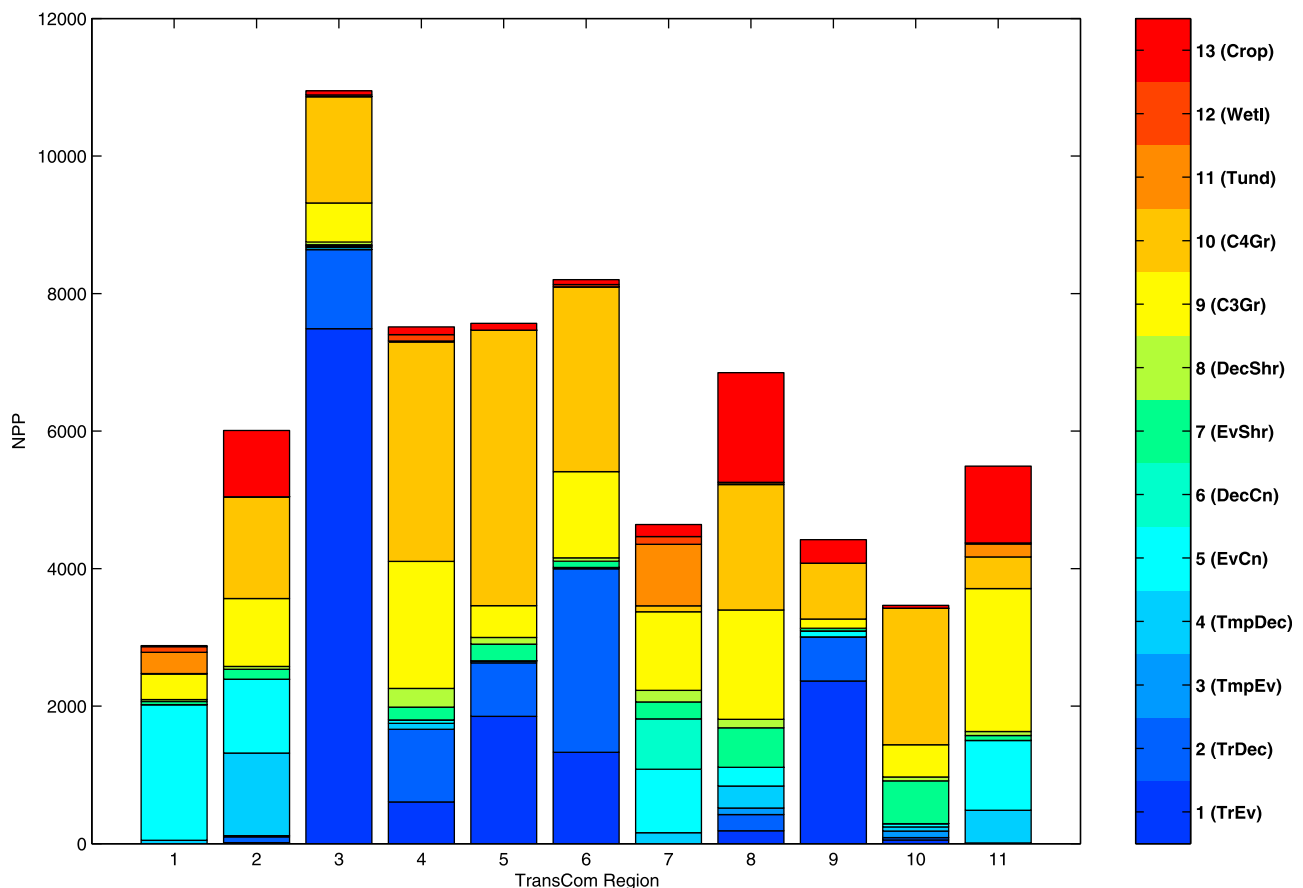


Figure 7. NPP per TransCom land region and PFT (TgC yr^{-1}).

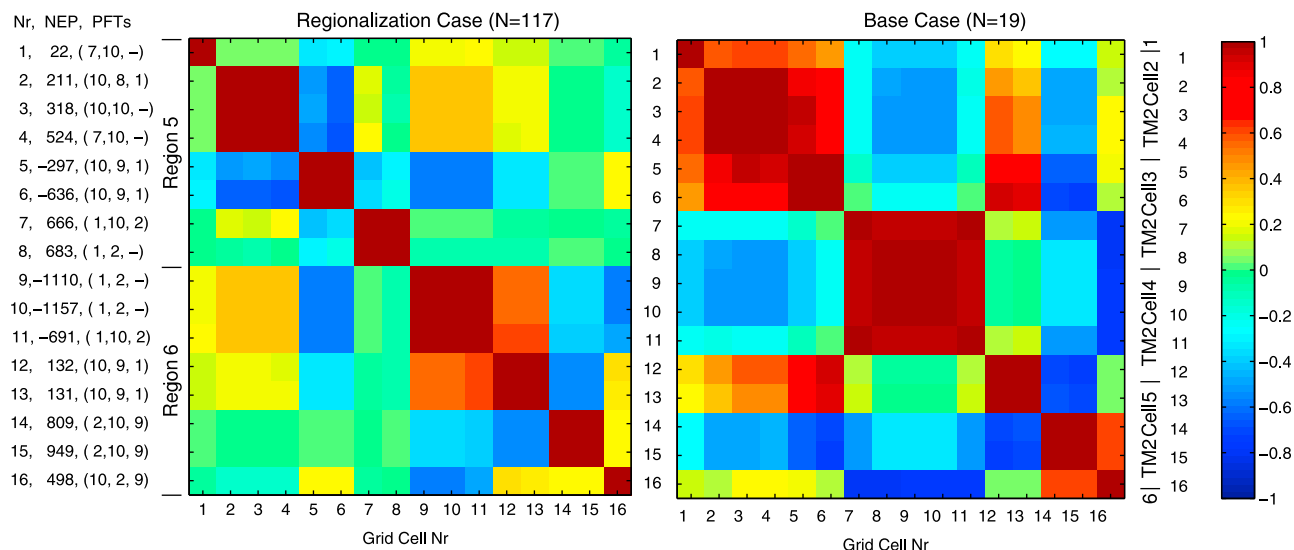


Figure 8. Uncertainty covariance matrix of NEP per grid cell for a transect through Africa (21°E , 15°N (grid cell 1) to 15°S (grid cell 16), every 2°). NEP for the regionalization case is also shown in $\text{gC m}^{-2} \text{yr}^{-1}$, as well as the PFTs occurring at each grid cell from most to least abundant.

the regionalization case (see Figure 6). There is only one larger positive correlation between the two Regions 6 and 8 (southern Africa and temperate Asia). Because these two regions only have the universal soil carbon parameters in common, which is the case for all pairings, we suspect that this must be a feature resulting from the atmospheric transport. Overall, correlations between regions are either small in absolute terms, or large and negative, such as Regions 5 and 6 (northern and southern Africa), 1 and 11 (boreal North America and Europe), 7 and 11 (boreal Eurasia and Europe), 3 and 5 (tropical South America and northern Africa), and 5 and 8 (northern Africa and temperate Eurasia). We expect a negative correlation if an increase in NEP in one region is compensated for by a decrease in NEP in the other, or vice versa, via atmospheric transport and constrained by atmospheric CO_2 data. The fact that the examples just mentioned are geographically close suggest exactly that. The question is, however, how atmospheric transport can lead to a positive uncertainty correlation, as found for Regions 6 and 8 above. We observe that the two regions are in opposite hemispheres and suspect that this might have to do with the constraint given by the hemispherical CO_2 gradient. If NEP increases in one hemisphere and the predicted gradient exceeds observations, then NEP must increase in the other hemisphere to compensate for the discrepancy. Further analysis will be needed to ascertain this.

[43] The global inversion study of Rödenbeck *et al.* [2003] also analyzed the *a posteriori* covariance structure of the uncertainties of the long-term fluxes for the TransCom land regions. The covariance matrix as presented in Figure 13 of their paper agrees in many ways with the pattern we obtain for our regionalization case. They also find that correlations are either very small or land regions predominantly have negative correlations between each other (for instance Region 1 and 11 or Region 7 and 11). Although the overall pattern of our covariance matrix shows large similarities, the study of Rödenbeck *et al.* [2003] identifies positive correlations between the neighboring Regions 3/4, 5/6 and 9/10, whereas

our study identifies a strong negative correlation for Regions 5/6 and weaker negative correlations for Regions 3/4 and Regions 9/10.

[44] The overall correlation structure for the 11 land regions appears to be more realistic for the regionalization case confirmed by the small or predominantly negative correlations among the land regions. In the base case distant regions are bound together by common β values which results in strongly positively correlated NEP uncertainties.

3.3.2. Covariance Structure for a Transect in Africa

[45] As a final analysis of the way the optimization interacts with the regionalization, we present the uncertainty correlation matrix for NEP for a north-south transect either side of the equator in central Africa, as shown in Figure 5. The transect is characterized by large fluctuations in the net CO_2 flux within a small area in the regionalization case, but much less fluctuation in the base case. Since both solutions are consistent with the atmospheric constraints, we can assume that the network cannot resolve such regional differences and suspect that the uncertainties between the fluxes must be highly correlated. We also observe that the transect cuts through two of the TransCom regions (5 and 6) and that a large swing in net flux coincides with this boundary. The areas just north and south of the boundary are dominated by PFT 1, followed by PFT 10 further north and south. In the southern region, there is also an area dominated by PFT 2 further to the south. The uncertainty correlation for the NEP of this transect is shown in Figure 8 for both the base case and the regionalization case.

[46] The first thing to notice for the regionalization case is that of the 16 grid cells concerned, the optimization effectively distinguishes 8 groups of one to three grid cells each. If two grid cells are dominated by the same PFTs and are in the same TransCom region, they effectively act as one and the uncertainty correlation between the cells approaches +1 (if there is more than one in the group). An example is the group of cells 9–11, which share PFT 1 in Region 6, or cells 12 and 13, which have the same PFTs and are also in

Region 6. We also find that the NEP uncertainties of neighboring groups of cells are often anticorrelated, for example the zone within Region 5 dominated by PFT 10 (cells 5 and 6) and the one to the south dominated by PFT 1 (cells 7 and 8). Correlations between grid cells of different regions are generally small. No pronounced impact of the position of the transport model (TM2) grid cells is detectable in the results, except possibly for one example: the two groupings cells 2–4 and cells 5–6 are anticorrelated even though both are in the same zone and dominated by PFT 10.

[47] We find that the pattern of uncertainty covariance is very different in the base case compared to the regionalization case. The artificial boundary at the equator introduced by the regionalization (between cells 8 and 9) is no more present. Instead, the grid cells can be roughly divided into four groups: cells 1–6 for the grass and shrub land region to the north, the tropical rain forest (cells 7–11), the grassland zone to the south (cells 12–13), and the deciduous forest and savanna zone furthest south (cells 14–16). There is some similarity between the results within Region 5, for the pairing 12–13 with 14–16, which is highly correlated within and highly anticorrelated across. But further north, for cells 1–11, the structures are completely different and more dominated by ecosystem type in the base case.

4. Discussion

[48] The outcome of the study presented here has shown that the spatial differentiation of the carbon balance parameter β leads to an improved fit to the observations. The uncertainty covariance matrix for annual mean NEP per TransCom region shows mainly small correlations in absolute terms or large and negative ones. In the base case, where β is applied globally, some regions are highly positively correlated due to a common value of β . Therefore, the optimization is unable to effectively differentiate between these two regions as far as mean simulated NEP is concerned.

[49] On a grid cell level however the flux pattern seems to be more realistic for the base case. The regionalization leads to very large fluxes of opposite sign especially in South America, Africa and India. Even though this is a subjective judgment, we would like to state that both the magnitude of NEP and the uncertainty correlation structure for the transect appear more realistic and in line with ecophysiological understanding in the base case. While regionalization allows a better fit of the model to the data, this happens at the cost of creating flux patterns which seem unrealistic. On the other hand, unrealistic flux patterns in the African region might be intensified by the artificial region boundary at the equator which enables the optimization to produce large opposing fluxes within the same ecosystem.

[50] A further result is that the optimization finds posterior values for β which are within the allowed range, but appear unrealistic from a carbon balance standpoint. The prior value for β , contained in J_p , does not constrain the optimization sufficiently to ensure that posterior values are within realistic bounds. The contribution of the observations, J_o to the total cost function is too large.

[51] Finally, we find values of χ^2 that are still somewhat above 1, indicating some missing processes. These results raise a number of questions, as discussed in the following:

[52] 1. Possibly, the current Bayesian framework puts sufficient emphasis on parameter priors when it treats one parameter value and one observation equally. Possibly, observations entering the cost function are not independent as assumed here [e.g., Ricciuto *et al.*, 2008], and prior parameter values represent more than just one measurement. Both arguments would favor a stronger weighting of prior parameter values. This problem is not restricted to the regionalization case (although it is most pronounced here), but also applies in the base case where for example β for PFT 8 is extremely large ($\beta_8 = 11.33$). Previous studies using CCDAS [e.g., Scholze *et al.*, 2007] also found high net fluxes for this PFT, even though the reciprocal of the current definition of β was used. An additional parameter constraint restricting β to a range (i.e. $0 \leq \beta \leq \beta_{\max}$), would most effectively prevent extremely large net fluxes. Such a solution, however, would imply a deviation from the premise of Gaussian distributed prior PDFs of parameters, and it would have to be investigated how CCDAS could be adjusted to account for the additional constraint.

[53] 2. What would be a sensible criterion to better spatially differentiate the β parameter? The TransCom land regions as applied here do not seem to be the best choice. In particular the “artificial boundary” at the equator which divides Africa into Regions 5 and 6 has created questionable results.

[54] 3. What is the best way of representing biomass decomposition fluxes in a CCDAS? This question is still open, but it appears that the regionalization based on the TransCom regions alone does not lead to enough improvement in the match with observations. Insufficient observations in the tropical regions and potential missing processes in the terrestrial biosphere model BETHY, such as fire, might “encourage” the extremely large β values. What is still needed is a scheme that takes into account all possible factors influencing the slow and fast decomposing of soil and litter carbon stocks.

5. Summary and Conclusions

[55] In this study we have used the Carbon Cycle Data Assimilation System (CCDAS) to investigate the effects of geographical differentiation of one control parameter for the long-term soil carbon balance on predicted net CO₂ fluxes. In one case, this parameter was differentiated only by vegetation type, while in the other case it was additionally differentiated geographically following the 11 standard TransCom land regions.

[56] A restructuring of CCDAS in which NPP related parameters are kept fixed led to a marked improvement of the performance of the optimization itself. We were thus able to find a cost function minimum in both cases for which the gradient of the cost function is close to zero with respect to all parameters and all eigenvalues of the Hessian are positive definite. This is an important outcome, since we derive posterior parameter uncertainties from the Hessian, thus approximating the inverse uncertainty covariance of the posterior parameters. We also show that the minimum is robust and we find good indications that we are dealing with a global minimum in physical parameter space.

[57] The regionalization of the soil carbon balance parameter led to a significantly improved fit to the observations.

However, analysis of the net CO₂ fluxes and their uncertainty covariance revealed widely diverging patterns between the two cases. The search for an appropriate spatial differentiation of β is therefore still open. Future work is required to show how sensitive β is to various regionalization patterns.

[58] In the regionalization case, the tropics and subtropics are dominated by widely diverging net fluxes of opposite sign. In some cases, the net carbon flux far exceeds NPP, which is extremely unlikely from an ecophysiological standpoint. We find that we require a better method of incorporating prior information into the Bayesian framework to constrain β to fall into its ecophysiological probable range as well as a denser observational network, in particular in the tropics and a complete representation of the processes governing long-term soil carbon balance that relies entirely on input data and universal parameters.

[59] **Acknowledgments.** This work was supported by the QUEST programme of the Natural Environment Research Council, U.K.

References

- Bird, M., H. Santruckova, J. Lloyd, and E. Lawson (2002), The isotopic composition of soil organic carbon on a north-south transect in western Canada, *Eur. J. Soil Sci.*, *53*, 393–403.
- Boden, T. A., G. Marland, and R. J. Andres (2009), Global, regional, and national fossil-fuel CO₂ emissions, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn., doi:10.3334/CDIAC/00001.
- Bopp, L., C. Le Quéré, M. Heimann, and A. C. Manning (2002), Climate-induced oceanic oxygen fluxes: Implications for the contemporary carbon budget, *Global Biogeochem. Cycles*, *16*(2), 1022, doi:10.1029/2001GB001445.
- Bousquet, P., P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein, and P. P. Tans (2000), Regional changes in carbon dioxide fluxes of land and oceans since 1980, *Science*, *17*, 1342–1346.
- Farquhar, G. D., S. von Caemmerer, and J. A. Berry (1980), A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species, *Planta*, *149*, 78–90.
- Fletcher, R., and M. J. D. Powell (1963), A rapidly convergent descent method for minimization, *Comput. J.*, *6*, 163–168.
- Friend, A. D., A. K. Stevens, R. G. Knox, and M. G. R. Cannell (1997), A process-based, terrestrial biosphere model of ecosystem dynamics (Hybrid v3.0), *Ecol. Modell.*, *95*, 249–287.
- Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Software*, *24*, 437–474.
- Gurney, K. R., D. Baker, P. Rayner, and S. Denning (2008), Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO₂ inversions for the period 1980 to 2005, *Global Biogeochem. Cycles*, *22*, GB3025, doi:10.1029/2007GB003082.
- Heimann, M. (1995), The global atmospheric tracer model TM2, *Tech. Rep.*, *10*, Dtsch. Klim., Hamburg, Germany.
- Houghton, R. A. (2008), Carbon flux to the atmosphere from land-use changes: 1850–2005, in *TRENDS: A Compendium of Data on Global Change*, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Kaminski, T., and M. Heimann (2001), Inverse modeling of atmospheric carbon dioxide fluxes, *Science*, *294*, 259.
- 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, Lecture Notes Comput. Sci.*, vol. 2668, edited by V. Kumar et al., pp. 95–104, Springer, New York.
- Knorr, W. (2000), Annual and interannual CO₂ exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, *Global Ecol. Biogeogr.*, *9*, 225–252.
- Knorr, W. (2009), Is the airborne fraction of anthropogenic CO₂ emissions increasing?, *Geophys. Res. Lett.*, *36*, L21710, doi:10.1029/2009GL040613.
- 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*, 207–225.
- Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogeé, 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.
- Le Quéré, C., et al. (2007), Saturation of the Southern Ocean CO₂ sink due to recent climate change, *Science*, *316*, 1735–1738, doi:10.1126/science.1136188.
- Prentice, I. C., M. Heimann, and S. Sitch (2000), The carbon balance of the terrestrial biosphere: Ecosystem models and atmospheric observations, *Ecol. Appl.*, *10*, 1553–1573.
- Prentice, I. C., et al. (2001), The carbon cycle and atmospheric carbon dioxide, in *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., chap. 3, pp. 183–237, Cambridge Univ. Press, New York.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1996), *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge Univ. Press, New York.
- 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.
- Riccio, D. M., K. J. Davis, and K. Keller (2008), A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty, *Global Biogeochem. Cycles*, *22*, GB2030, doi:10.1029/2006GB002908.
- Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964.
- Scholze, M. (2003), Model studies on the response of the terrestrial carbon cycle on climate change and variability, Ph.D. thesis, Max Planck Inst. für Meteorol., Hamburg, Germany.
- Scholze, M., T. Kaminski, P. Rayner, W. Knorr, and R. Giering (2007), Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *J. Geophys. Res.*, *112*, D17305, doi:10.1029/2007JD008642.
- Sitch, S., et al. (2003), Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model, *Global Change Biol.*, *9*, 161–185.
- 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 CO₂ flux over the global oceans: An improved estimate based on the sea-air pCO₂ difference, paper presented at 2nd International CO₂ in the Oceans Symposium, Cent. for Global and Environ. Res., Natl. Inst. for Environ. Stud., Tsukuba, Japan.
- Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, New York.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.
- W. Knorr, M. Scholze, and T. Ziehn, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Rd., Bristol BS8 1RJ, UK. (wolfgang.knorr@bristol.ac.uk; marko.scholze@bristol.ac.uk; tilo.ziehn@bristol.ac.uk)