On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models

T. Ziehn,[1,2] M. Scholze,[1,3] and W. Knorr[1,4]

[1] Terrestrial ecosystem models (TEMs) contain the coupling of many biogeochemical processes with a large number of parameters involved. In many cases those parameters are highly uncertain. In order to reduce those uncertainties, parameter estimation methods can be applied, which allow the model to be constrained against observations. We compare the performance and results of two such parameter estimation techniques - the Metropolis algorithm (MA) which is a Markov Chain Monte Carlo (MCMC) method and the adjoint approach as it is used in the Carbon Cycle Data Assimilation System (CCDAS). Both techniques are applied here to derive the posterior probability density function (PDF) for 19 parameters of the Biosphere Energy Transfer and Hydrology (BETHY) scheme. We also use the MA to sample the posterior parameter distribution from the adjoint inversion. This allows us to assess if the commonly made assumption in variational data assimilation, that everything is normally distributed, holds. The comparison of the posterior parameter PDF derived by both methods shows that in most cases an approximation of the PDF by a normal distribution as used by the adjoint approach is a valid assumption. The results also indicate that the global minimum has been identified by both methods for the given set up. However, the adjoint approach outperforms the MA by several orders of magnitude in terms of computational time. Both methods show good agreement in the PDF of estimated net carbon fluxes for the decades of the 1980s and 1990s.

1. Introduction

[2] Predictions of future climate strongly depend on the concentrations of greenhouse gases in the atmosphere with CO_2 being the most important one. Atmospheric CO_2 concentrations are determined by the size of the global exchange fluxes with the oceans and the land as well as the anthropogenic emissions. Terrestrial ecosystem models (TEMs) can be used to estimate the net exchange flux of CO_2 between the land and the atmosphere and therefore play an important role in the Earth system.

[3] State-of-the-art TEMs such as the Joint UK Land Environment Simulator (JULES) [*Best et al.*, 2011; *Clark et al.*, 2011] or the Biosphere Energy Transfer and Hydrology

(BETHY) scheme [*Knorr*, 2000] contain a large number of biogeochemical processes, which makes them very complex models with a large number of process parameters involved. In most cases, we do not know the exact value of the parameters, and prior parameter values are therefore based on expert knowledge. In some cases this is little more than an informed guess. The large uncertainties associated with prior parameter values also lead to large variations in the predictions of the future land-atmosphere CO_2 fluxes [*Knorr and Heimann*, 2001], which in turn contributes to the uncertainties in future climate projections.

[4] Due to the increasing number of process parameters involved in state-of-the-art TEMs, it becomes more and more important to focus on the reduction of their uncertainties. Parameter estimation methods are very useful in this context, because they provide an objective way of constraining the model against observations and in this way are able to reduce the parameter uncertainties.

[5] Various parameter estimation methods such as adjoint, genetic algorithm, Kalman Filter, Levenberg-Marquardt and Monte Carlo inversion have been compared for example in the OptIC (Optimization InterComparison) project [*Trudinger et al.*, 2007]. The aim here was to estimate four parameters in a highly simplified representation of the carbon dynamics in a TEM with only two state variables. A forward run of the

[1]School of Earth Sciences, University of Bristol, Bristol, UK.
[2]Now at Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, Aspendale, Victoria, Australia.
[3]KlimaCampus, University of Hamburg, Hamburg, Germany.
[4]Now at Physical Geography and Ecosystem Analysis, Lund University, Lund, Sweden.

Corresponding author: T. Ziehn, Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, PMB 1, Aspendale, Vic 3195, Australia. (tilo.ziehn@csiro.au)

model was used to generate artificial data, which were then treated as observations after degradation through added noise, correlations, drifts and gaps. It was found that all methods were equally successful at estimating the parameters. A comparison in terms of computational efficiency was not made, due to the fact that the model was inexpensive to run. Also, the model did not have multiple minima, which therefore did not allow for a comparison in terms of the ability to find the global minimum.

[6] The REFLEX project [*Fox et al.*, 2009] compared methods based on genetic algorithm, Kalman Filter and Monte Carlo inversion using the Data Assimilation Linked Ecosystem Carbon (DALEC) model [*Williams et al.*, 2005]. DALEC is a simple box model of carbon pools used here in two versions, as a model for evergreen and a model for deciduous vegetation. The evergreen version required calibration of 11 parameters related to allocation and turnover of carbon pools, whereas the deciduous version required calibration of 17 parameters. REFLEX used both synthetic (generated from the model with added noise) and real data. It was found that estimates of confidence intervals varied among algorithms. Again, the main focus here was not on comparing the methods in terms of their computational efficiency nor their ability to find the global minimum.

[7] Many parameter estimation methods use the Bayesian approach, which has proven to provide a powerful and convenient framework for combining prior knowledge about parameters with additional information such as observations [*Rayner et al.*, 2005]. The resulting inverse problem described by Bayes' theorem can be solved in different ways. Here we focus on the comparison of two types of methods: Monte Carlo inversion [*Sambridge and Mosegaard*, 2002] and variational data assimilation [*Talagrand and Courtier*, 1987]. Monte Carlo inversion methods such as the Markov Chain Monte Carlo method (MCMC) have a better chance to converge to the global minimum than have gradient-based methods for example. In principle, the MCMC will converge to the global minimum if the number of iterations is large enough. However, the maximum number of iterations may be restricted by the computing time of the model. MCMC methods are easy to implement and they require no assumptions about the model (i.e. continuity) and the posterior probability distribution of parameters may be non-Gaussian, even if the prior distribution is assumed to be Gaussian (normally distributed).

[8] Variational data assimilation, such as the four-dimensional variational (4D-Var) scheme, is one of the most advanced approaches to assimilate observed information into a model. It uses derivative code (i.e. the adjoint of the model) for the optimization of the parameters and therefore requires the model to be differentiable with respect to all parameters. Although the 4D-Var approach is computationally very efficient in most cases, the optimization might only identify a local minimum due to the non-linearity and high dimensionality of the model. Another criticism of the 4D-Var method is that it focuses only on the optimal solution, i.e. the mode of the probability density function (PDF) without considering uncertainties. However, some 4D-Var schemes, such as the Carbon Cycle Data Assimilation System (CCDAS) [*Rayner et al.*, 2005], allow the calculation of posterior parameter uncertainties using the inverse of the Hessian (second order

derivative) of the cost function at the global minimum. Unfortunately, this is only correct for linear problems. If the model is non-linear and a Gaussian distribution is assumed for the prior parameters, the model needs to be linearized around the optimum in parameter space, and the posterior distribution will only be approximated by a Gaussian [*Tarantola*, 1987]. This approximation might not always be reasonable, considering that most TEMs are highly non-linear.

[9] In this contribution we compare the 4D-Var (adjoint) approach as implemented in CCDAS, with the Metropolis algorithm (MA) [*Metropolis et al.*, 1953; *Mosegaard and Tarantola*, 1995], which is one possible MCMC method. We apply both methods in order to estimate the posterior PDF of 19 process parameters in the terrestrial ecosystem model BETHY. BETHY is a complex grid-based model, which simulates carbon assimilation and soil respiration within a full energy and water balance and phenology scheme. The main focus is on the performance of the two methods in terms of their efficiency (i.e. number of required model runs) and their ability to find the global minimum. In addition, the MA will allow us to assess the full shape of the PDF of single model parameters – as part of the evaluation of the full PDF containing the dependence on all model parameters simultaneously – and thus provides an indication of whether or not the assumption of a Gaussian posterior PDF of parameters made in CCDAS is justified.

2. Methodology

[10] Data assimilation can be seen as a way of combining data (i.e. observations) with prior information (i.e. process model formulation and prior process parameter value) to derive posterior parameter estimates. The Bayesian approach [*Tarantola*, 1987, 2005] provides a powerful and convenient framework for data assimilation as it combines the prior probability distribution $p(\mathbf{x})$ of the parameters $\mathbf{x}$ with the (forward) probability distribution $p(\mathbf{c}|\mathbf{x})$ of the observations $\mathbf{c}$ given the parameters $\mathbf{x}$ to obtain the (inverse) posterior probability distribution $p(\mathbf{x}|\mathbf{c})$ of the parameters $\mathbf{x}$ given the observations $\mathbf{c}$:

$$p(\mathbf{x}|\mathbf{c}) = \frac{1}{A}p(\mathbf{c}|\mathbf{x})p(\mathbf{x}). \tag{1}$$

The factor $1/A$ is a normalization constant with $A = p(\mathbf{c})$. It is independent of the parameters $\mathbf{x}$ and used to scale equation (1) so that the integral over the posterior $p(\mathbf{x}|\mathbf{c})$ equals one. The probability distribution $p(\mathbf{c}|\mathbf{x})$ describes the distribution of the observations assuming that we know the parameter values (i.e. given a set of parameters $\mathbf{x}$) and the probability distribution $p(\mathbf{x}|\mathbf{c})$ describes the distribution of the parameters after obtaining information on the observations (i.e. given the observations $\mathbf{c}$). This is known as the inverse problem and Bayes theorem described by equation (1) allows us to obtain $p(\mathbf{x}|\mathbf{c})$ through synthesis of observations and prior information. In principle, equation (1) gives us the exact information about the shape of the posterior PDF of parameters. The difficulty with using the equation directly simply lies in the fact that it is often highly dimensional. However, we are usually only interested in knowing what the most likely set of parameters is, i.e. in the global maximum of $p(\mathbf{x}|\mathbf{c})$, and in the

region around it. This is why solving equation (1) requires optimization.

[11] The forward probability distribution $p(\mathbf{c}|\mathbf{x})$ is determined by the model M at the point $\mathbf{x}$ in parameter space and provides a measure of how good the model is in explaining the observations. The modeled counterpart to the observations are here denoted $\mathbf{c}_M$ with $\mathbf{c}_M = M(\mathbf{x})$. If we assume a Gaussian distribution, we obtain this expression for the PDF of observations for a given set of parameters:

$$p(\mathbf{c}|\mathbf{x}) = \frac{1}{A'} \exp\left(-\frac{1}{2}(\mathbf{c}_M - \mathbf{c})^T \mathbf{C_c}^{-1}(\mathbf{c}_M - \mathbf{c}) \right) \qquad (2)$$

where $\mathbf{C_c}$ is the combined error covariance matrix of the observations and the model. A', and A'' below, are normalization constants. To express prior knowledge about parameters, we can again use a Gaussian formulation:

$$p(\mathbf{x}) = \frac{1}{A''} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C_{x_0}}^{-1}(\mathbf{x} - \mathbf{x}_0) \right). \qquad (3)$$

$\mathbf{x}_0$ is the prior estimate of the model parameters, and $\mathbf{C_{x_0}}$ the corresponding error covariance matrix.

[12] This optimization problem can be solved in different ways, for example through Monte Carlo inversion or variational data assimilation. In many cases, a Gaussian distribution is assumed for the prior parameter values and the observations, as shown in the two equations above. Combining equation (1) with equations (2) and (3) leads to the following full expression requiring inversion:

$$p(\mathbf{x}|\mathbf{c}) = \frac{1}{A} \exp\left(-\frac{1}{2}(\mathbf{c}_M - \mathbf{c})^T \mathbf{C_c}^{-1}(\mathbf{c}_M - \mathbf{c}) \right)$$
$$\cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{C_{x_0}}^{-1}(\mathbf{x} - \mathbf{x}_0) \right). \qquad (4)$$

2.1. Variational Data Assimilation and CCDAS

[13] By taking the negative logarithm of equation (4) (ignoring A), we obtain the so-called cost function $J(\mathbf{x})$, which we can minimize instead of maximizing $p(\mathbf{x}|\mathbf{c})$ directly:

$$J(\mathbf{x}) = \frac{1}{2}\left((\mathbf{c}_M - \mathbf{c})^T \mathbf{C_c}^{-1}(\mathbf{c}_M - \mathbf{c}) + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{C_{x_0}}^{-1}(\mathbf{x} - \mathbf{x}_0) \right).$$
$$(5)$$

$J(\mathbf{x})$ expresses the mismatch between the observations and their modeled equivalents and the mismatch between the parameters and their priors. Variational data assimilation schemes aim to minimize the cost function by making use of its gradient.

[14] In this study we use CCDAS, which has been previously described by *Scholze* [2003] and *Rayner et al.* [2005]. Only a short summary is provided here. The data assimilation in CCDAS is performed in two steps: In the first step, the full BETHY model (TEM in CCDAS) is used to assimilate global monthly fields of the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) for optimizing parameters controlling soil moisture and phenology [*Knorr and Schulz*, 2001]. In the second step, soil moisture and leaf area index (LAI) fields are provided as inputs for a reduced version of BETHY. This version is used to assimilate

atmospheric CO_2 concentration observations from a large number of observation stations for optimizing photosynthesis and soil carbon parameters and to derive their posterior uncertainties [*Rayner et al.*, 2005; *Scholze et al.*, 2007].

[15] The focus in this work is on the second data assimilation step. However, in contrast to the set up used by *Rayner et al.* [2005], we only optimize the soil carbon part of BETHY, keeping all parameters controlling net primary productivity (NPP) fixed. Earlier studies [*Rayner et al.*, 2005; *Scholze et al.*, 2007] have shown that these parameters are only relatively weakly constrained by the atmospheric CO_2 data. However, *Ziehn et al.* [2011b] have demonstrated how to constrain the NPP-related parameters (i.e. photosynthesis parameters) using an extensive set of plant traits. We also expect that new satellite observations such as fluorescence will have a strong constraint on these parameters.

[16] CCDAS is operated in a variety of modes. In calibration mode, CCDAS serves as an estimator algorithm for the heterotrophic respiration process parameters, subsequently referred to as soil carbon parameters. A quasi-Newton method, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) variant of the Davidon-Fletcher-Powell (DFP) formula [*Fletcher and Powell*, 1963; *Press et al.*, 1996], is used for the iterative 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 generated from the model's source code using the tool Transformation of Algorithms in Fortran (TAF) [*Giering and Kaminski*, 1998; *Kaminski et al.*, 2003].

[17] If the gradient of the cost function reaches zero, a minimum has been found and the posterior parameter uncertainties can be calculated using the Hessian mode of CCDAS. The Hessian, i.e. the second order derivative of the cost function with respect to the model parameters, describes the curvature of the cost function. At the cost function minimum, the Hessian approximates the inverse covariance of the optimal soil carbon parameters and can thus be used to calculate the posterior parameter uncertainties. We calculate the Hessian by differentiating the gradient vector with respect to all parameters by applying TAF a second time. More details on how the Hessian is calculated and how the posterior parameter uncertainties are derived can be found in *Rayner et al.* [2005]. Due to the fact that the BETHY model is non-linear, the posterior probability density of single parameters will only be approximated by a normal distribution [*Tarantola*, 1987]. In order to confirm if this assumption holds, we compare the posterior parameter PDFs derived by CCDAS with the posterior parameter PDFs derived by the MA.

[18] CCDAS also allows the propagation of uncertainties from process parameters forward through the BETHY model by making use of the model's Jacobian (first order derivative). In this further mode, we obtain the uncertainties and covariances for diagnostic quantities such as the net carbon flux. The Jacobian code is again generated directly from the model's source code using TAF. The propagation of parameter uncertainties using the Jacobian is described in detail by *Scholze et al.* [2007].

2.2. Monte Carlo Inversion

[19] Monte Carlo inversion methods such as the MCMC method are also able to find an optimal solution, in this case

the maximum of $p(\mathbf{x}|\mathbf{c})$ of equation (4), by sampling the posterior parameter PDF directly. Here we use the MA to generate a Markov Chain to derive the posterior PDF of 19 process parameters from the BETHY model. The algorithm is based on an acceptance-rejection strategy and a summary of the required steps is provided in Appendix A.

[20] Although it is not possible to parallelize the MA itself, because the generation of the next point always depends on the previous point, it is possible to run multiple chains or sequences in parallel. The advantage of using multiple parallel chains is that a larger number of samples can be generated within the same amount of time.

[21] Although the MA is easy to implement, it requires tuning of at least one internal parameter, the step length factor s_i. If s_i is too large, the algorithm jumps and the sampled distribution becomes highly irregular because of a low acceptance rate and a low number of sampled points in parameter space. On the other hand, if the proposed steps are too short, the acceptance rate is high, but only a small range of the sampled distribution is covered. For practical applications, according to *Gelman et al.* [2003], the acceptance rate should lie between 0.23 and 0.44, depending on the dimension of the parameter vector.

[22] In this study we perform two experiments based on the MA:

[23] 1. E1 - Sampling the CCDAS posterior minimum

[24] 2. E2 - Optimization starting from the prior parameter values.

[25] In the first experiment (E1), we start the MA in the cost function minimum identified by the adjoint approach of CCDAS, and in the second experiment (E2), we start the MA from the prior parameter values. The first experiment mainly serves the purpose of sampling the minimum to obtain the uncertainties and the shape of the posterior parameter PDFs in order to compare them with the Gaussian approximation we use in CCDAS. The second experiment should confirm whether or not we are able to find the same minimum with the MA as with the adjoint approach. Both experiments will give us an indication of whether the minimum found with CCDAS is in fact the global one.

[26] The MA requires adaptation of the step length Δx_i for each of the parameters in order to obtain an acceptance rate between 0.23 and 0.44 (see above). Here, we calculate the step length as follows:

$$\Delta x_i = s_i \cdot \sigma_{x_{0i}} \cdot N(0, 1) \tag{6}$$

where $N(0, 1)$ is a random number with a normal distribution, s_i is the step length factor and $\sigma_{x_{0i}}$ the prior parameter uncertainty. Due to the fact that we perform the MA in a normalized domain, all parameters have a prior uncertainty of one (i.e. $\sigma_{x_{0i}} = 1$ for all parameters). Therefore, we use the same step length factor for each of the parameters. Tuning s_i requires running the MA for a number of iterations to see the effects of a variable step length on the acceptance rate. Here, we tune the step length factor in a way that we obtain an acceptance rate of 0.35 for both experiments.

2.3. Terrestrial Ecosystem Model and Parameters

[27] BETHY is a process-based model of the terrestrial biosphere. It simulates carbon assimilation and soil respiration 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] in order to compare them with the observations provided by the GLOBALVIEW database [*GLOBALVIEW-CO₂*, 2008]. Here we use monthly mean atmospheric CO^2 concentration data from 41 sites globally over 25 years (1979 to 2003). A more detailed description of the data set can be found in *Rayner et al.* [2005].

[28] In this study, we focus only on the soil carbon part of BETHY, and we keep all parameters controlling NPP fixed (for more details see *Ziehn et al.* [2011a]). In order to reduce computational time, we use a coarse grid with a resolution of $8° \times 10°$. Global vegetation is mapped onto 13 different plant functional types (PFTs) (see Table S1 of the auxiliary material) and each grid cell can contain sub-areas (sub-grid cells) with up to three different PFTs with different fractional cover.[1] The dominant PFT in each grid cell is presented in Figure S1 (see auxiliary material). In the present study, BETHY is driven by observed climate data for the period 1979 to 2003. A detailed description of BETHY can be found in *Knorr and Heimann* [2001].

[29] Parameters in BETHY are assigned via a mapping routine and can be either global (i.e. they have the same value in each of the grid cells) or differentiated by certain criteria. Here, all parameters are global, except for the carbon balance parameter β, which is differentiated by PFT j, denoted β_j. This results in a set of 19 parameters (β_1 to β_{13} + 5 global parameters + 1 offset). The five global parameters are the temperature dependence parameters of soil respiration $Q_{10,f}$ and $Q_{10,s}$, the pool turnover time parameter τ_f for the fast soil carbon pool, the fraction f_s of decomposition from the fast pool to the long-lived soil carbon pool, and κ, a parameter describing linearity of soil moisture dependence of soil decomposition. The carbon balance parameter β_j determines whether a PFT acts as a long-term source ($\beta_j > 1$) or long-term sink ($0 < \beta_j < 1$). The offset describes the global atmospheric CO_2 concentration at the beginning of the assimilation period. A more detailed description of the soil carbon part of BETHY is given in the auxiliary material.

[30] We distinguish between model parameters p_i (physical domain) and parameters as used by CCDAS and the MA x_i (normalized domain). For most of the parameters p_i (parameters 1 to 18), we assume a log-normal distribution to guarantee that model parameters are always positive, by applying the following transformation between physical and normalized domain (in the normalized domain, all prior parameters have a Gaussian distribution and an uncertainty of 1):

$$x_i = \frac{\log(p_i)}{\log(\sigma_{p_{0i}} + p_{0i}) - \log(p_{0i})} \tag{7}$$

p_{0i} is the prior value and $\sigma_{p_{0i}}$ the prior uncertainty for the model parameters. For the offset (parameter 19) we assume a normal distribution in the physical domain and apply the following transformation:

$$x_i = \frac{p_i}{\sigma_{p_{0i}}} \tag{8}$$

CCDAS results will be discussed in the physical domain,

---

[1]Auxiliary materials are available in the HTML. doi:10.1029/2011GB004185.

Table 1. Prior ($p_{0,i}$) and Posterior (p_i) Parameter Values Including Uncertainties for the Optimization With CCDAS (Physical Domain)[a]

Parameter		Prior		Posterior		
		p_{0i}	$\sigma_{p_{0i}}$	p_i	σ_{p_i}	Reduction (%)
1	$Q_{10,f}$	1.50	0.75	1.80	0.03	97.02
2	$Q_{10,s}$	1.50	0.75	1.58	0.04	95.33
3	τ_f	1.50	3.00	10.04	0.44	97.82
4	κ	1.00	9.00	0.66	0.01	99.88
5	f_s	0.20	0.20	0.40	0.01	97.71
6	β_1	1.00	0.25	1.05	0.01	95.53
7	β_2	1.00	0.25	0.72	0.03	82.54
8	β_3	1.00	0.25	0.37	0.06	37.59
9	β_4	1.00	0.25	0.55	0.06	56.97
10	β_5	1.00	0.25	1.59	0.03	93.60
11	β_6	1.00	0.25	0.48	0.07	40.76
12	β_7	1.00	0.25	0.17	0.02	45.63
13	β_8	1.00	0.25	11.95	0.45	91.74
14	β_9	1.00	0.25	1.13	0.03	90.98
15	β_{10}	1.00	0.25	0.89	0.01	93.64
16	β_{11}	1.00	0.25	0.37	0.03	63.17
17	β_{12}	1.00	0.25	1.20	0.30	0.86
18	β_{13}	1.00	0.25	0.35	0.03	65.18
19	Offset	336.85	1.00	336.15	0.04	95.59

[a]For parameters with a lognormal distribution (all but offset), the upper percentile equivalent to one standard deviation is given. The relative reduction in uncertainty defined as $(1 - (\sigma_{p_i}/p_i)/(\sigma_{p_{0i}}/p_{0i})) \cdot 100\%$ is also shown.

whereas the comparison of the results from CCDAS and the MA will only be discussed in the normalized domain.

3. Results

[31] We first estimate the soil carbon parameters with CCDAS using the adjoint approach. The prior parameter values and their uncertainties can be found in Table 1. The optimization in CCDAS converges within 361 iterations. The final cost function value is $J = 9642$ (initial cost function value: $3 \cdot 10^6$) and the gradient in the cost function minimum is sufficiently small (reduced from 10^7 to 10^{-3}), which indicates that a minimum has been found. The posterior parameter values are also presented in Table 1. Most of the global parameters are close to their prior values and within the prior uncertainty range. Only the pool turnover time for the fast pool (τ_f) is much larger than its prior value and well outside the prior uncertainty range. The carbon balance parameter (β) for PFT 8 (deciduous shrub) is extremely large and indicates that locations covered by this PFT act as a long term source with a net ecosystem productivity (NEP) more than ten times that of NPP. Although this value is within the allowed physical range, it is unrealistic from a carbon balance point of view. This has already been discussed by *Ziehn et al.* [2011a].

[32] After the calibration mode of CCDAS, we compute the Hessian of the cost function to obtain posterior parameter uncertainties, as mentioned above. The uncertainties for most parameters can be reduced by more than 90% (see Table 1), which indicates that the parameters are well constrained by the atmospheric CO_2 data. Only parameter β_{12} cannot be constrained by the data. However, NPP associated with this PFT (swamp vegetation) is very small and a simple sensitivity analysis performed by changing the parameters away from the minimum by $\pm 10\%$ has revealed that this parameter

has only little effect on the overall cost function value. The same is true for some of the other β parameters, where the uncertainty reduction is relatively small, for example β_3 and β_6.

3.1. Ensemble Runs With CCDAS

[33] One of the shortcomings of the adjoint approach is that we may identify only a local minimum. We therefore perform a large number of optimizations (ensemble runs), starting each run in a different point in parameter space [*Kaminski et al.*, 2010]. Ideally we would like to see all ensemble runs converging to the same minimum. However, in practice this is rather unlikely due to the high dimensionality of the parameter space and the non-linearity of the BETHY model. Here, we investigate the outcome of three sets of ensemble runs, where we randomly select the starting points by taking the prior parameter values and varying them by a maximum of 1%, 10% and 25%. Each set of ensemble runs consists of 25 optimizations. All ensemble runs are performed in parallel on a computer cluster so that no additional computational run time is required.

[34] Within the 75 (3 × 25) optimizations, we identify five different minima. However, only two of them are within the physical parameter space (minimum M_1 and M_2, see Table 2). The other three minima contain non-physical parameter values (i.e. $Q_{10,f} < 1$, $Q_{10,s} < 1$ or $f_s > 1$) and are therefore not relevant here. For nine runs, the optimization did not converge, two runs within the set of the 10% variation and seven runs within the set of the 25% variation. All of the 25 runs for the 1% variation finished in the same minimum M_1 we found using the prior parameter values, which shows that the optimization is robust for small parameter changes. However, the further we move away from the prior parameter values (10% and 25% variation), the more likely it is that we end up in a different minimum. Out of the 75 ensemble runs, M_1 has the smallest cost function value and therefore appears to be the global minimum. Although it is not possible to prove that M_1 really is the global minimum in the whole physical parameter space, we refer to it as global minimum in the following discussion.

[35] The performance of the optimization for the case where we varied the starting point by a maximum of 10% is shown in Figure 1. In addition to the 25 ensemble runs we also present the run where the starting point was set to the prior parameter values. This is the standard set up in CCDAS. We obtain the fastest convergence rate by starting in the prior parameter values (361 iterations). All runs finishing in minimum M_1 require less than 1000 iterations. Within the first 100 iterations we obtain the largest reduction in the cost function value (Figure 1a) by two to three orders of magnitude. After that convergence is slower (Figure 1b) and the

Table 2. Results From the Ensemble Runs With CCDAS[a]

Minima	J_o	J_p	J	25 Ensemble Runs		
				1%	10%	25%
M_1	9500.82	141.33	9642.15	25	16	3
M_2	12531.41	67.69	12599.10	-	1	3

[a]Cost function value J, mismatch of the observations J_o, and mismatch of the parameters J_p are given for two identified minima (M_1 and M_2). The starting point was randomly disturbed from the prior by 1%, 10% and 25% for 25 realizations each.
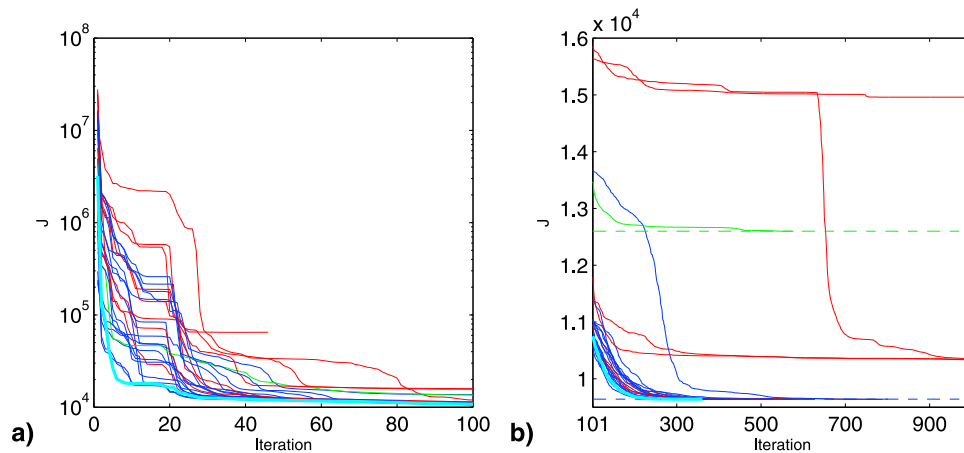
Figure 1. Cost function J for 25 ensemble runs (random starting points at up to 10% variation from prior) and for the optimization with start in the prior parameter values for CCDAS in (a) log scale for the first 100 iterations and (b) in linear scale for the following iterations. Blue: ensemble runs finishing in minimum M_1, green: ensemble run finishing in minimum M_2, light blue: optimization with start in prior parameter values, red: no convergence or ensemble runs finishing in minimum with non-physical parameter values. Cost function minima are marked with dashed lines.

cost function value changes only by a small amount in between iterations. Although parameter values are not substantially changing any more, we need to perform these iterations to reduce the value of the gradient until it reaches zero.

3.2. Metropolis Algorithm (MA)

[36] For the first experiment, E1 (sampling the minimum with the MA), we use five chains, all starting in the minimum M_1 with a maximum number of four million iterations. In order to avoid correlations between subsequent samplings, we use only every 10th iteration for the estimation of the posterior parameter distribution, which leaves us with a total of two million samples ($5 \times 400,000$).

[37] The results for all parameters are presented in Table 3 together with the results from the CCDAS optimization. Note that in order to compare the results of both methods we only present the normalized domain.

[38] There is a good agreement between CCDAS and the MA sampling for the five global parameters, the offset and the majority of the beta parameters. However, there is disagreement in either the mean value or standard deviation or both for some β parameters, namely for β_3, β_4, β_6, β_7, β_{11} and β_{12}. This discrepancy will be discussed in detail in section 4.

[39] The PDFs of the five global parameters and β_1 are shown in Figure 2. Not only is there a good agreement between the uncertainties derived by CCDAS and the MA, the shape of the PDFs is also close to a Gaussian as indicated by skewness and kurtosis (see Table 3). Skewness is a measure of the asymmetry of the PDF. For example, a negative

Table 3. Posterior Parameter Values and Uncertainties (One Standard Deviation) Obtained From CCDAS and the MA[a]

		CCDAS		MA Sampling				MA Optimization			
	Parameter	x_i	σ_{x_i}	x_i	σ_{x_i}	Skewness	Kurtosis	x_i	σ_{x_i}	Skewness	Kurtosis
1	$Q_{10,f}$	1.44	0.04	1.45	0.04	0.31	−0.07	1.48	0.04	−0.04	−0.18
2	$Q_{10,s}$	1.13	0.06	1.11	0.07	−0.36	0.08	1.07	0.06	0.17	−0.35
3	τ_f	2.10	0.04	2.11	0.04	0.39	−0.00	2.13	0.04	−0.23	−0.25
4	κ	−0.18	0.00	−0.18	0.00	−0.03	0.00	−0.18	0.00	−0.03	−0.03
5	f_s	−1.34	0.03	−1.36	0.04	−0.35	0.62	−1.33	0.03	−0.12	−0.21
6	β_1	0.22	0.05	0.17	0.05	0.04	0.17	0.18	0.05	0.18	−0.14
7	β_2	−1.50	0.19	−1.62	0.18	0.10	−0.29	−1.75	0.15	0.04	−0.35
8	β_3	−4.43	0.65	−4.83	0.36	−0.34	−0.76	−3.04	0.65	−0.07	−1.01
9	β_4	−2.67	0.46	−3.06	0.37	−0.82	0.48	−2.98	0.29	−0.14	−0.89
10	β_5	2.07	0.07	2.03	0.09	−0.13	−0.23	2.10	0.07	−0.37	0.18
11	β_6	−3.31	0.62	−2.69	0.62	1.38	1.44	−1.54	0.39	−0.26	−1.30
12	β_7	−8.05	0.57	−8.89	0.63	−0.53	−0.76	−6.37	0.73	−0.37	−0.43
13	β_8	11.12	0.09	11.21	0.08	0.34	−0.16	11.09	0.10	−0.07	−0.33
14	β_9	0.55	0.10	0.69	0.09	−0.01	−0.37	0.54	0.09	0.19	0.32
15	β_{10}	−0.54	0.07	−0.56	0.07	−0.33	0.53	−0.47	0.06	0.29	−0.10
16	β_{11}	−4.50	0.39	−5.42	0.68	−0.07	−1.13	−4.93	0.48	0.98	0.53
17	β_{12}	0.81	0.99	−0.38	0.75	−0.06	−1.20	−3.96	0.97	−0.01	−0.82
18	β_{13}	−4.73	0.37	−4.61	0.24	0.10	−0.29	−5.06	0.37	−1.16	0.16
19	Offset	336.62	0.04	336.61	0.05	0.01	−0.23	336.59	0.05	−0.18	0.28

[a]$N = 400,000$ samples for both experiments E1 and E2. Parameter values are in the normalized domain.
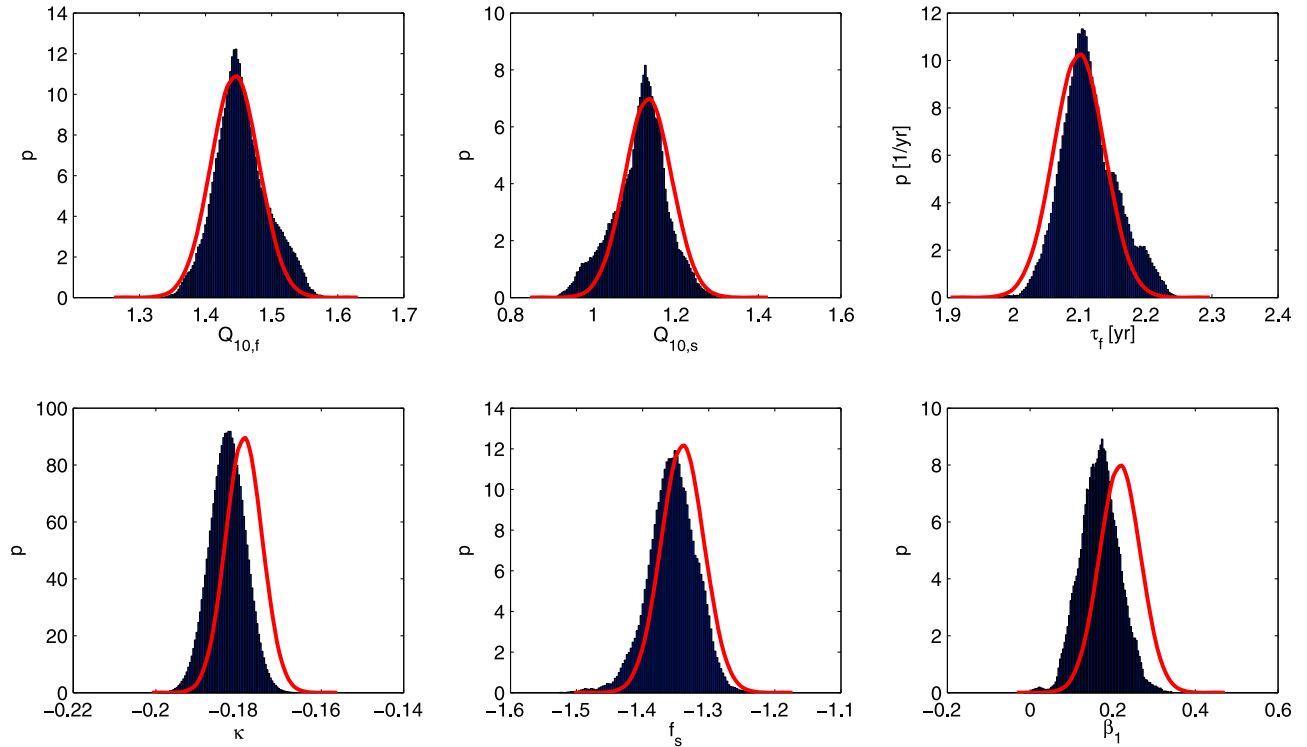
Figure 2. Posterior parameter PDF obtained from optimization with CCDAS (red) and using the MA experiment E1 with $N = 5 \times 400{,}000$ samples (black) for the five global parameters and one β parameter. Parameter values are in normalized space.

skew indicates that the left tail of the PDF is longer than the right tail. Kurtosis is a measure of how peaked a distribution is. Positive kurtosis indicates a peaked distribution whereas negative kurtosis indicates a flat distribution [*von Storch and Zwiers*, 1999]. A Gaussian distribution has a value of zero for both skewness and kurtosis. Further, we can see that the MA samples the same parameter space as suggested by CCDAS without diverging into a different minima.

[40] For the second experiment E2 (optimization with the MA) we use four chains, starting each of them in the prior parameter values ($\mathbf{x}_0$). Here we use a maximum number of eight million iterations, because we also have to consider a burn-in time (cut-off before convergence to the PDF maximum). We choose a burn-in time of four million iterations to be on the safe side and sample the remaining four million iterations by choosing every 10th parameter set, which leaves us with a total of 1.6 million samples (4 × 400,000).

[41] The convergence rate of the MA is presented in Figure 3. About 10,000 iterations are required to reduce the cost function value by two orders of magnitude (Figure 3a). However, in order to get into the vicinity of the global PDF maximum (cost function minimum M_1), about three million iterations are required (Figure 3b). We can also see that we do not obtain cost function values which are smaller than M_1. This is a further indication that minimum M_1 might indeed be a global minimum.

[42] Calculated mean and standard deviation for all parameters are shown in Table 3. For the global parameters, the offset and the majority of the β parameters the results are very similar to the ones obtained from experiment E1 and agree well with the results derived by CCDAS including the

shape of the PDFs. There is disagreement again in the results for β_3, β_4, β_6, β_7, β_{11} and β_{12} for mostly both, mean and standard deviation, not only in comparison to CCDAS, but also in comparison to the previous experiment. Reasons behind this will be discussed in section 4.

3.3. Parameter Uncertainty Covariances

[43] In addition to the mean and uncertainties for single posterior parameters derived by CCDAS and the MA, we also compare the posterior uncertainty covariance between parameters. The covariance between the parameters can be expressed via the uncertainty correlation matrix $\mathbf{R}$, which is defined as follows:

$$R^{i,j} = \frac{C_x^{i,j}}{\sigma_i \sigma_j} \qquad (9)$$

where $C_x^{i,j}$ is element i, j of the posterior uncertainty covariance matrix of the parameters, and σ_i the posterior uncertainty of parameter i derived from the diagonal elements $C_x^{i,j}$ of the matrix $\mathbf{C_x}$.

[44] Figure 4 presents the correlation matrix for the three cases: parameter uncertainties derived by CCDAS and by the MA for both experiments E1 and E2. In CCDAS the uncertainties are derived using the inverse of the Hessian in the cost function minimum M_1, for the MA the uncertainties are calculated using 2 million samples (5 × 400,000 for start in minimum) and 1.6 million samples (4 × 400,000 for start in prior, after burn-in) respectively.

[45] There is good agreement in the correlation for the five global parameters between all three cases. Parameters $Q_{10,f}$
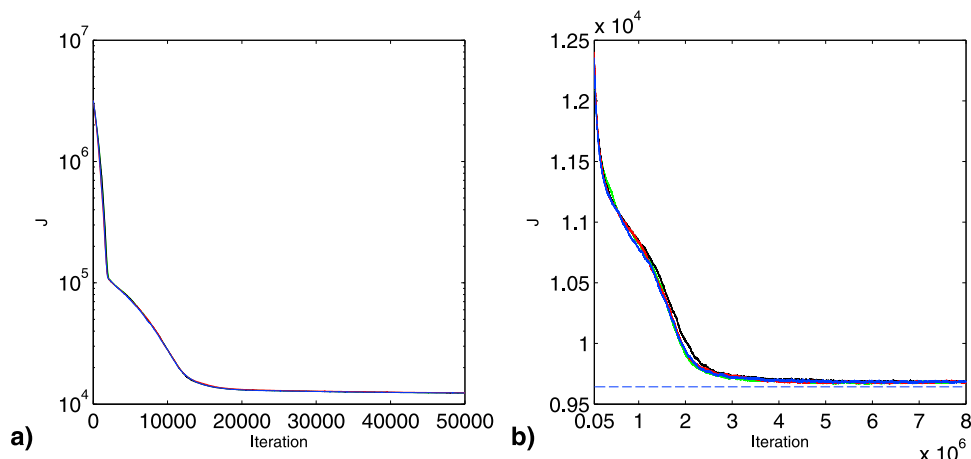
Figure 3. Cost function J for four chains using the MA (all four chains start from the prior parameter values) in (a) log scale for the first 50,000 iterations and (b) linear scale for the following iterations. Cost function minimum M_1 is marked with a dashed line.

and $Q_{10,s}$ show a strong negative correlation, whereas $Q_{10,f}$ and τ_f are positively correlated. We believe that these correlations are caused by our specific set up to optimize only soil carbon parameters. As we keep NPP fixed the optimization can change the seasonality of the net flux only by changing the seasonality of the heterotrophic respiration which is controlled by these three parameters. However, there is disagreement between the three cases for some of the β parameters. With CCDAS we obtain relatively weak, mainly negative correlation between the parameters β_3, β_4, β_7, β_{11} and β_{12}, whereas with the MA we obtain mainly a strong positive correlation. These are the same parameters we had difficulties to recover with the MA in both experiments. If we focus on the course of those parameter values over their iterations (i.e. β_3 and β_7 as shown in Figure 5 for both MA experiments), we can see that all parameters follow a negative trend, which explains why they are all positively correlated. This is further investigated in section 4.

3.4. Diagnostics

[46] We consider two diagnostic quantities here, global NEP for the decade of the 1980s and the 1990s. With CCDAS we propagate the posterior parameter uncertainties forward through the BETHY model in order to obtain the

uncertainties for NEP. The computational effort is negligible in this case.

[47] For the MA we estimate the PDF for NEP directly from sampling the results. Here we use every 100th iteration from the MA experiment E1 which provides us with a total of 200,000 samples (40,000 samples per chain). The results for CCDAS and the MA are shown in Figure 6. There is good agreement in the results from both methods and the PDFs obtained by the MA are close to a Gaussian distribution as indicated by skewness and kurtosis (see Figure 6). The simulated values by CCDAS are for the 1980s 1.66 PgC/yr and for the 1990s 2.4 PgC/yr. We notice a negligible shift in the mean of NEP to smaller values with the MA. For both decades we obtain slightly larger uncertainties with the MA than in comparison with CCDAS. However, for both methods the uncertainties are very small for both decades. Uncertainties in NEP for individual years are much larger (e.g. a factor of three between NEP for the year 1990 and NEP for the decade of the 1990s). This is due to a large number of negative correlations between individual years [see also *Ziehn et al.*, 2011a]). Also note that the uncertainties derived here for the diagnostic quantities only reflect the uncertainties in the soil carbon parameters. The small difference in the NEP
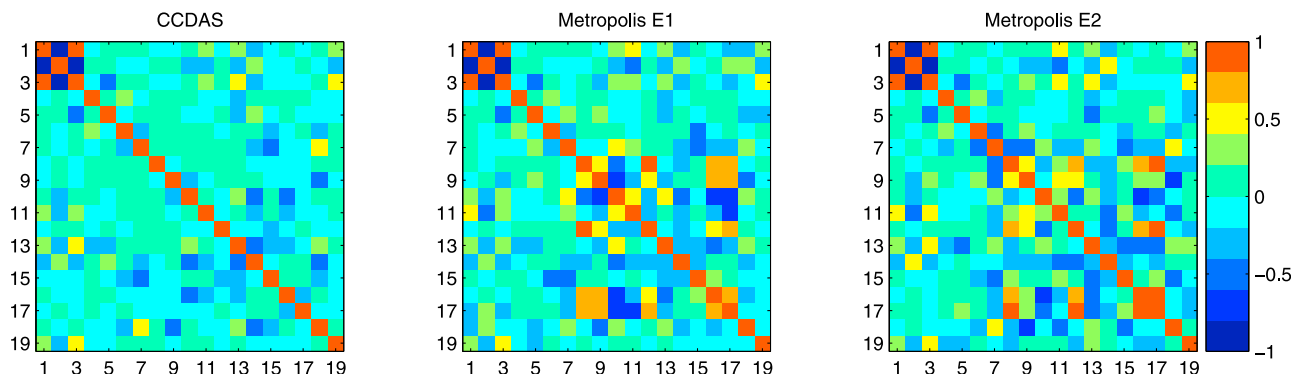

Figure 4. Uncertainty covariance matrix for the 19 process parameters derived by CCDAS, the MA experiment E1 using 4,000,000 iterations and the MA experiment E2 using 4,000,000 iterations after the burn-in.
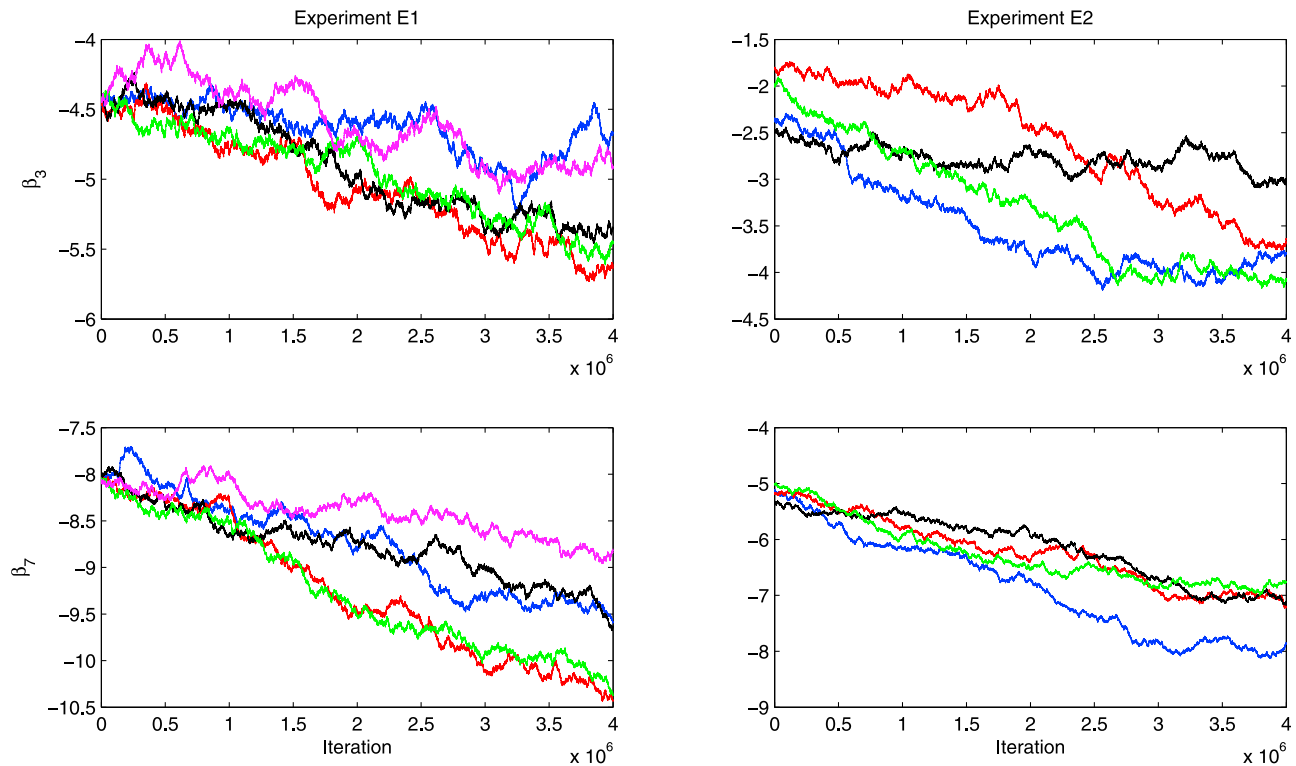
Figure 5. Sampled parameter values for (top) β_3 and (bottom) β_7 using 4,000,000 iterations for MA experiment E1 and E2. Different colors are used to distinguish between the chains.

results between CCDAS and the MA is due to the difference in the posterior parameter PDFs for some of the parameters derived by both methods as shown previously.

4. Discussion

4.1. Computational Expense and Performance

[48] One of the major advantages of the adjoint approach with regards to this study is its fast convergence. When starting from the prior parameter values we were able to reach the cost function minimum within 361 iterations. Within each iteration the gradient of the cost function and the cost function value itself are calculated with respect to all parameters using the adjoint of the BETHY model. In terms of

computational time, the evaluation of the adjoint is about two times more expensive than a forward model run. The 361 adjoint evaluations therefore correspond to roughly 800 forward model runs. On top of this, each iteration requires (multiple) BETHY model runs for the line search algorithm within the optimization scheme, so that the total number of forward model runs adds up to about 1300. The whole optimization process with CCDAS required less than one hour of computational running time on a standard PC with a 2.4 GHz processor. All CCDAS ensemble runs were performed in parallel on a computer cluster. The computation of the posterior uncertainties via the inverse Hessian and the propagation of the parameter uncertainties via the Jacobian also
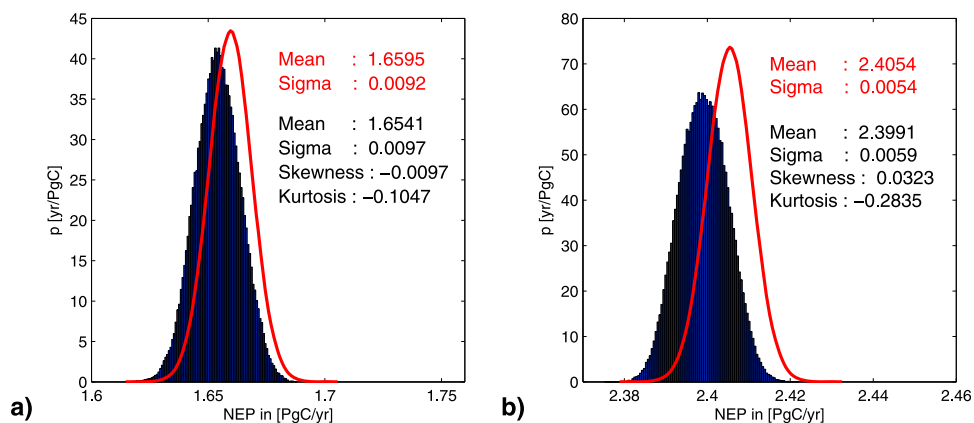

Figure 6. PDF for global NEP per year for the (a) 1980s and (b) 1990s estimated using CCDAS (red) and the MA experiment E1 (black).
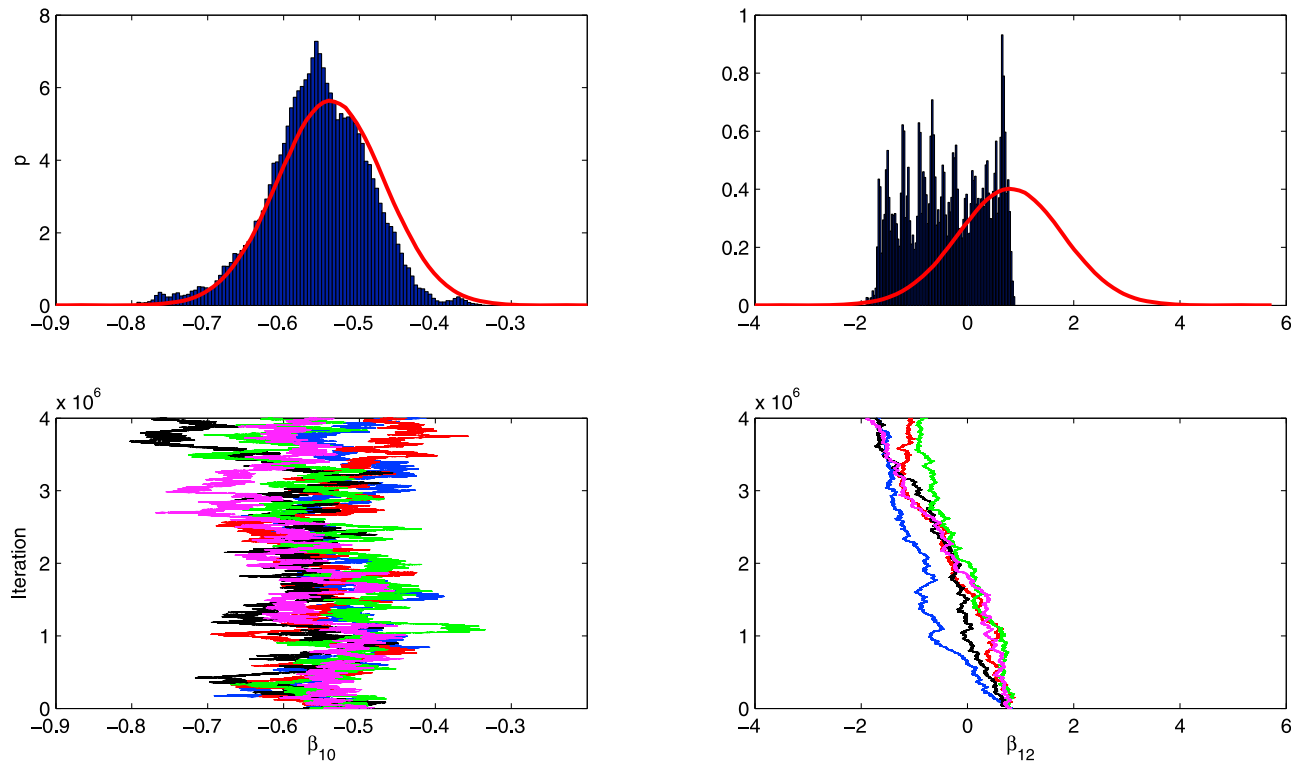
Figure 7. (top) PDF and (bottom) sampled parameter values for β_{10} and β_{12} using the MA experiment E1 over 4,000,000 iterations (note that number of iteration is on the y-axis and parameter value on the x-axis). Different colors are used to distinguish between the chains.

require computational time, however in this case it was negligible.

[49] The MA requires only forward model runs, one per iteration. We constrained the maximum number of iterations to eight million per chain for the optimization (experiment E2). This is a rather arbitrary number and the limit was chosen simply because of computational limitations. All chains were run on a computer cluster in parallel which allowed up to one million iterations per month and chain. The overall running time was about eight months. The adjoint approach was thus by several orders of magnitude faster than the MA for the given set-up.

[50] Even though the adjoint approach is very fast in identifying the cost function minimum, there is a risk that the minimum is only a local one. We have demonstrated that if we start the optimization in the parameter priors or close to the priors (variation by maximal 1%) we always find the same minimum and that this minimum has the lowest cost function value within the physically possible parameters space. However, if we move the starting point further away from the prior (variation by up to 10% or 25%), then the chance increases that we identify a local (higher) minimum or a minimum with non-physical parameter values, or that the optimization does not converge. Ideally, we would like to find the global minimum independently from the starting point, which is not the case here, probably due to the non-linearity and high-dimensional parameter space of the BETHY model. This also demonstrates the importance of using appropriately chosen prior parameter values. For future applications, we might need to prevent parameters from assuming non-physical values, for example through the use

of a constrained optimization scheme or the use of a different parameter transformation method for certain parameters.

[51] We have not investigated here how the MA would cope with different starting points (i.e. starting each of the chains in a different point), something we considered beyond the scope of this paper. However, we feel that the choice of the starting point is only of minor importance for the MA in terms of convergence to a local minimum. In contrast, the initial guess of the step length within the MA scheme has a far bigger impact on the convergence. This is discussed in more detail at a later stage.

[52] Another issue with the MA is the enormous hard disk space requirement while running the algorithm. Parameter values and output files need to be stored for each iteration or at least for every sample. The BETHY model produces gridded output files for a number of diagnostic quantities which adds up to about 10 MB disk space per run for the current set-up. It is not possible to store all output files for all iterations while running the MA. One solution applied here is not to write any output files at all, saving computing time required for disk access, but to run the BETHY model again with one sub-sample of every 100th iteration for each of the chains of the recorded parameter values. These sub-sampled runs were used in order to compute the PDF of the chosen diagnostic quantities, in this case the global decadal mean NEP. This required additional computational running time.

4.2. Agreement Between the Results

[53] In the first experiment, E1, we started the MA in the global cost function minimum M_1 and the MA did not diverge from this minimum. In the second experiment, E2,
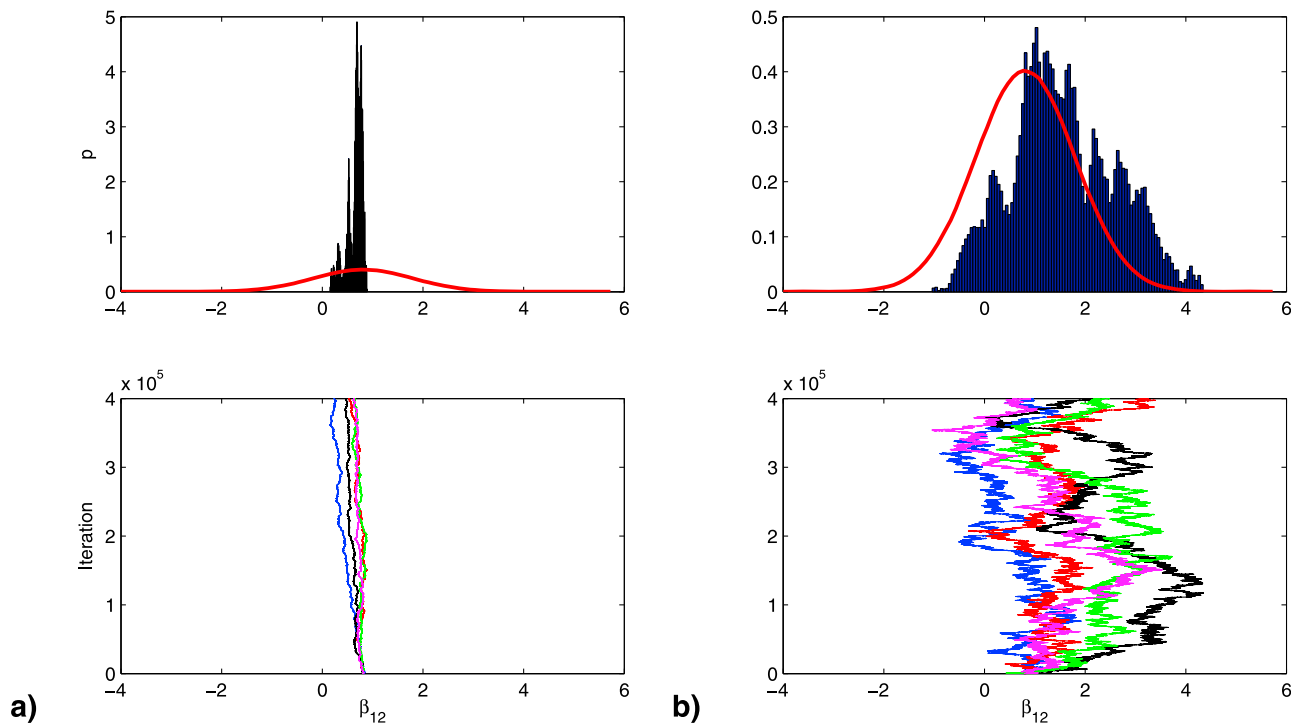
Figure 8. (top) PDF and (bottom) sampled parameter values for β_{12} using the MA experiment E1 over 400,000 iterations with different step lengths (note again number of iteration is on the y-axis and parameter value on the x-axis). (a) As implemented, (b) increased for test case. Different colors are used to distinguish between the chains.

we started the sampling from the prior parameter values and after a long burn-in time (four million iterations) the MA also converged to the global minimum M_1. Posterior mean and standard deviation for the parameters could be directly computed from the samples and we obtained similar results for both experiments for most of the parameters. Most of the results also agreed well with the mean and uncertainty derived by CCDAS and even the shape of the posterior parameter distribution was close to a Gaussian for most parameters.

[54] However, for some β parameters (β_3, β_4, β_7, β_{11} and β_{12} in particular), the mode of the PDF and the uncertainty did not agree with the values derived by CCDAS or even within the two MA experiments. We suspect that this is due to the choice of a single step length factor for all parameters. In some cases, the step length might have been too small and the PDF of the specific parameter was still under-sampled.

[55] Figure 7 shows the sampled parameter values and derived PDF for β_{10} for the MA experiment E1. For this parameter we obtain good agreement between the PDFs derived by both methods. We can see that sampled β_{10} values change significantly with the number of iterations, which allows the whole PDF to be covered over the total number of iterations. In contrast to this, the sampled values for β_{12} (also shown in Figure 7) change only slightly and more worryingly they follow a negative trend. We suspect that at least in this case, the chosen step length is too small. Figure 8 shows what would happen if we would increase the step length for β_{12}. Due to computational limitations we can only present this for a smaller number of iterations (400,000), but the difference is apparent already: a larger

part of the PDF is visited if the step length is larger and there is no negative trend. The parameter is now sampled around its mean value. We obtain similar results for the other β parameters, where we had a big discrepancy between the CCDAS and MA results. If we avoid the occurrence of such a negative trend for all β parameters, we suspect we would also avoid the strong "artificial" correlations between the β parameters as shown in Figure 4.

[56] Parameters where the step length appears to be too small also turned out to be parameters with only a small reduction in the uncertainty according to the CCDAS inversion (see Table 1), which means that they cannot be constrained well by the atmospheric CO_2 data. Because the posterior PDF is larger for unconstrained parameters, they should have used a larger step length. Information about posterior uncertainties is not available when setting up the MA, but adapting the step length for individual parameters seems to be necessary in order to run the MA efficiently. This is a long iterative process which requires a large computational effort on top of the actual MA running time, which is not affordable for larger models, such as global TEMs.

5. Summary and Conclusions

[57] We investigated two different approaches on how to efficiently estimate parameters of the terrestrial ecosystem model BETHY. As expected, the adjoint approach was the computationally most efficient approach and outperformed the MA by several orders of magnitude of "real" compute time. Running an ensemble of optimizations with CCDAS by varying the starting point allowed us to test if the identified

minimum is only a local minimum. The ensemble runs can be performed in parallel so that no additional computational running time is required. For the set-up used in this study, we were able to identify the global minimum if we kept the starting point for the optimization close to the priors. This made the success of the adjoint approach somewhat dependent on the starting point. However, the ensemble runs identified only one other minimum in the physical parameter space with a larger cost function value than the one derived with the prior parameter values.

[58] The MA is computationally very expensive and the convergence rate also depends on the step length factor for each of the parameters. We demonstrated that if the step length for a parameter is too small, only a small part of the posterior parameter PDF will be visited, which will consequently lead to misleading results. In fact, we showed that for some parameters the MA also diverges if the step length is too small. To avoid this problem, we would need to run the MA several times first (tuning mode), where we adapt the step length for each parameter individually. This iterative process would require another large number of model runs.

[59] For most parameters (where the step length was chosen appropriately), the MA was able to confirm mean and standard deviation as derived with the adjoint approach. The MA did not converge into a different minimum, which strengthens our confidence of having found the global minimum with both approaches. Although the adjoint approach calculates only the mode of the posterior PDF, we were able to estimate the uncertainties using the inverse of the Hessian in the cost function minimum with only minimal additional computational effort. *Kaminski et al.* [2003] have shown that this computational effort scales nearly linearly with the number of parameters. This means that for higher dimensional parameter spaces the estimation of the posterior parameter uncertainties will add to the computing costs, however, the overall computing time will not change dramatically. In contrast, this is not the case for the MA. Here, an increase in the number of parameters will consequently lead to slower convergence (curse of dimensionality) which makes this method even less suitable for higher dimensional parameter spaces.

[60] The results from the MA confirmed that the approximation of the posterior parameter PDF by a normal distribution as used in CCDAS is reasonable for most parameters. We also obtained good agreement between diagnostic quantities (global mean NEP for the 1980s and 1990s) between CCDAS and the MA. Both these results as well as the fact that we have confirmed our adjoint-derived minimum by the MA have important implications for terrestrial ecosystem parameter estimations. They demonstrate that despite the high non-linearity of terrestrial ecosystem models local linearizations of these models (i.e. derivatives such as the gradient, Hessian and Jacobian) provide valuable information for constraining model process parameters against observations and for deriving posterior uncertainty estimates on parameters and diagnostic quantities. From this study it is clear that Monte Carlo methods alone are not suitable for model optimization and posterior uncertainty estimations with complex state-of-the-art terrestrial ecosystem models.

[61] In summary, the adjoint approach is the most efficient way of estimating process parameter and their uncertainties in complex terrestrial ecosystem models. We demonstrated

that it is also possible to combine the adjoint approach and the MA. Here, the adjoint approach was used to locate the cost function minimum and the MA was then used to derive the posterior parameter uncertainties by sampling the minimum. The computational effort is much larger than using the adjoint approach alone, but smaller than using the MA solely since it requires a large number of iterations to converge to the cost function minimum.

Appendix A: Metropolis Algorithm: Step by Step

[62] First, choose a starting point $\mathbf{x}^i$ ($i = 1$), e.g. $\mathbf{x}^1 = \mathbf{x}_0$ (model parameter priors).

[63] Second, generate a proposed subsequent value $\mathbf{x}*$ by varying all elements of the parameter vector $\mathbf{x}^i$ by some step length $\Delta\mathbf{x}$. The step length is set for each parameter separately and determined by a predefined step length factor s_i times a Gaussian distributed random number with zero mean and standard deviation according to the parameters' prior uncertainty.

[64] Third, test for acceptance or rejection of the proposed point $\mathbf{x}* = \mathbf{x}^i + \Delta\mathbf{x}$ using a random number with a uniform distribution $U(0, 1)$:

[65] 1. Accept $\mathbf{x}*$, if $p(\mathbf{x}^i + \Delta\mathbf{x})/p(\mathbf{x}^i) \geq U(0, 1)$

[66] 2. Reject $\mathbf{x}*$, if $p(\mathbf{x}^i + \Delta\mathbf{x})/p(\mathbf{x}^i) < U(0, 1)$: $\mathbf{x}^{i+1} = \mathbf{x}^i$

[67] Fourth, if $\mathbf{x}*$ is accepted, then evaluate the forward model and test:

[68] 1. Accept $\mathbf{x}*$, if $p(\mathbf{c}|\mathbf{x}^i + \Delta\mathbf{x})/p(\mathbf{c}|\mathbf{x}^i) \geq U(0, 1)$: $\mathbf{x}^{i+1} = \mathbf{x}*$

[69] 2. Reject $\mathbf{x}*$, if $p(\mathbf{c}|\mathbf{x}^i + \Delta\mathbf{x})/p(\mathbf{c}|\mathbf{x}^i) < U(0, 1)$: $\mathbf{x}^{i+1} = \mathbf{x}^i$

[70] Fifth, repeat steps 2–4 until maximum number of iterations is reached.

[71] Note that only the second test (step 4) requires the evaluation of the model.

References

Best, M. J., et al. (2011), The Joint UK Land Environment Simulator (JULES), Model description—Part 1: Energy and water fluxes, *Geosci. Model Dev. Discuss.*, 4, 595–640, doi:10.5194/gmdd-4-595-2011.

Clark, D. B., et al. (2011), The Joint UK Land Environment Simulator (JULES), Model description—Part 2: Carbon fluxes and vegetation, *Geosci. Model Dev. Discuss.*, 4, 641–688, doi:10.5194/gmdd-4-641-2011.

Fletcher, R., and M. J. D. Powell (1963), A rapidly convergent descent method for minimization, *Comput. J.*, 6, 163–168.

Fox, A., et al. (2009), The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, *Agric. For. Meteorol.*, 149, 1597–1615.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003), Bayesian Data Analysis, Chapman and Hall/CRC Press, Boca Raton, Fla.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Software*, 24, 437–474.

GLOBALVIEW-CO$_2$ (2008), *Cooperative Atmospheric Data Integration Project—Carbon Dioxide* [CD-ROM], NOAA, Boulder, Colo.

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

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

Kaminski, T., W. Knorr, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P.-P. Mathieu (2010), Assimilation of MERIS FAPAR into a terrestrial vegetation model and mission design, paper presented at Living Planet Symposium, Eur. Space Agency, Bergen, Norway.

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

Knorr, W., and M. Heimann (2001), Uncertainties in global terrestrial biosphere modeling: 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, *Global Biogeochem. Cycles*, *15*, 207–225, doi:10.1029/1998GB001059.

Knorr, W., and J. P. Schulz (2001), Using satellite data assimilation to infer global soil moisture status and vegetation feedback to climate, in *Remote Sensing and Climate Modeling: Synergies and Limitations*, *Adv. Global Change Res.*, vol. 7, edited by M. Beniston and M. Verstraete, pp. 273–306, Springer, New York.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*, 12,431–12,447.

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.

Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, *40*(3), 1009, doi:10.1029/2000RG000089.

Scholze, M. (2003), Model studies on the response of the terrestrial carbon cycle on climate change and variability, PhD 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.

Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328.

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. Appl. and Math., Philadelphia, Pa.

Trudinger, C. M., et al. (2007), OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, *J. Geophys. Res.*, *112*, G02027, doi:10.1029/2006JG000367.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, Cambridge, UK.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius (2005), An improved analysis of forest carbon dynamics using data assimilation, *Global Change Biol.*, *11*, 89–105, doi:10.1111/j.1365-2486.2004.00891.x.

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

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