

The BETHY/JSBACH Carbon Cycle Data Assimilation System: experiences and challenges

T. Kaminski,¹ W. Knorr,² G. Schürmann,³ M. Scholze,² P. J. Rayner,⁴ S. Zaehle,³ S. Blessing,¹ W. Dorigo,⁵ V. Gayler,⁶ R. Giering,¹ N. Gobron,⁷ J. P. Grant,² M. Heimann,³ A. Hooker-Stroud,⁸ S. Houweling,⁹ T. Kato,¹⁰ J. Kattge,³ D. Kelley,^{8,14} S. Kemp,⁸ E. N. Koffi,⁷ C. Köstler,³ P.-P. Mathieu,¹¹ B. Pinty,⁷ C. H. Reick,⁶ C. Rödenbeck,³ R. Schnur,⁶ K. Scipal,¹¹ C. Sebald,⁵ T. Stacke,⁶ A. Terwisscha van Scheltinga,⁸ M. Vossbeck,¹ H. Widmann,¹² and T. Ziehn¹³

Received 29 June 2013; revised 8 September 2013; accepted 11 September 2013.

[1] We present the concept of the Carbon Cycle Data Assimilation System and describe its evolution over the last two decades from an assimilation system around a simple diagnostic model of the terrestrial biosphere to a system for the calibration and initialization of the land component of a comprehensive Earth system model. We critically review the capability of this modeling framework to integrate multiple data streams, to assess their mutual consistency and with the model, to reduce uncertainties in the simulation of the terrestrial carbon cycle, to provide, in a traceable manner, reanalysis products with documented uncertainty, and to assist the design of the observational network. We highlight some of the challenges we met and experience we gained, give recommendations for operating the system, and suggest directions for future development.

Citation: Kaminski, T., et al. (2013), The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges, *J. Geophys. Res. Biogeosci.*, 118, doi:10.1002/jgrg.20118.

1. Introduction

[2] There is an ever increasing number of observations on the carbon cycle becoming available that describe particular

processes or features of the global carbon cycle at various spatial scales, ranging from detailed measurements on the leaf level to regional scale information about boundary layer air masses. It is a highly challenging task to combine this wealth of observational information into an integrated view on the carbon cycle and to assure consistency between the available data streams. Such an integrated view is strongly needed to understand current trends in the global carbon cycle [Peters *et al.*, 2012] and to reduce uncertainty in future projections of the global carbon cycle and its climate feedbacks [Arora *et al.*, 2013; Jones *et al.*, 2013].

[3] A variety of methods have been developed in recent years for assimilation of observations into terrestrial biosphere models. Barrett [2002] applied a genetic algorithm to a conceptual model at continental scale. At site level, Wang *et al.* [2001] applied a variational approach, Braswell *et al.* [2005] a Monte Carlo algorithm, Williams *et al.* [2005] embedded an ensemble Kalman filter around a box model in a variational scheme, and Medvigy *et al.* [2009] used the simulated annealing technique. Fox *et al.* [2011] are used preparing a sequential scheme for assimilating site level observations into the Community Land Model (CLM) [Lawrence *et al.*, 2011]. Trudinger *et al.* [2007] and Fox *et al.* [2009] provide a comparison of assimilation methods applied to a simplified test model at site scale, and the review of Montzka *et al.* [2012] provides a classification of assimilation approaches.

¹FastOpt, Hamburg, Germany.

²Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden.

³Max Planck Institute for Biogeochemistry, Jena, Germany.

⁴School of Earth Sciences, the University of Melbourne, Melbourne, Victoria, Australia.

⁵Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria.

⁶Max Planck Institute for Meteorology, Hamburg, Germany

⁷Institute for Environment and Sustainability, Joint Research Centre, European Commission, Ispra, Italy.

⁸School of Earth Sciences, University of Bristol, Bristol, UK.

⁹Netherlands Institute for Space Research, Utrecht, Netherlands.

¹⁰Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France.

¹¹European Space Agency, Frascati, Italy.

¹²German Climate Computing Center, Hamburg, Germany.

¹³CSIRO Marine and Atmospheric Research, Aspendale, Victoria, Australia.

¹⁴Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia.

Corresponding author: T. Kaminski, FastOpt, Lerchenstr. 28a, DE-22767 Hamburg, Germany. (Thomas.Kaminski@FastOpt.com)

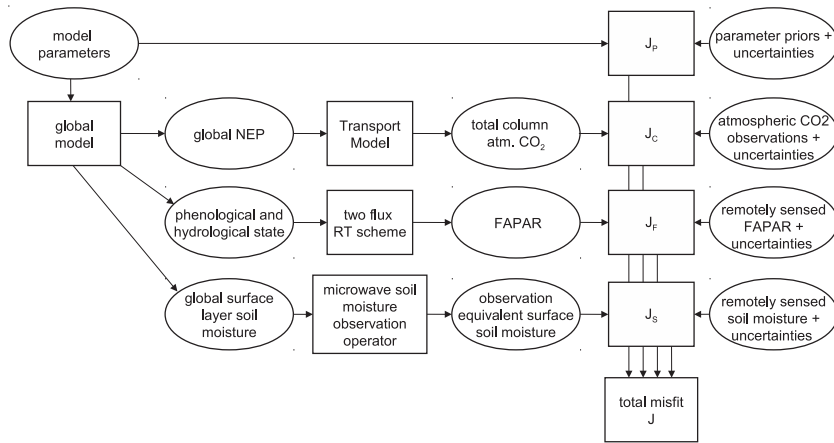


Figure 1. Flow graph of the cost function evaluation in the CarbonFlux study. Ovals denote data, rectangles processing. The total cost function is composed of contributions quantifying the misfit to individual data streams and the deviation from prior information.

[4] The Carbon Cycle Data Assimilation System (CCDAS) was designed as a modeling framework that uses observations related to the carbon cycle in a mathematically rigorous way to constrain simulations of the terrestrial biosphere. The combination of an advanced variational assimilation concept with a dynamical model allows us, in theory, to use an observation of a particular variable taken at a particular time and a particular location to help constrain another variable at a different time and location. It also allows us to assess whether a set of observations, possibly of different variables taken at different times and locations are statistically consistent with each other and the dynamics of the system. As pointed out by *Rayner* [2010], it was not clear whether these expectations could be met for a dynamical system as heterogeneous as the terrestrial biosphere.

[5] The beginnings of CCDAS date back to a study by *Knorr and Heimann* [1995] who employed high-precision flask samples of the atmospheric CO_2 concentration provided by a global network [*Conway et al.*, 1994] to constrain the Simple Diagnostic Biosphere Model (SDBM). Since this early study, significant progress has been made in various aspects, including complexity of the terrestrial model, the number of data streams, and the sophistication of the inversion strategy. After almost two decades of steady development by an ever-increasing team, the CCDAS has evolved into a complex system which can be illustrated by a flowchart (Figure 1), here adopted from the current CarbonFlux study (see <http://CarbonFlux.CCDAS.org>). This particular application of CCDAS aims at assimilating three Earth Observation (EO) data streams simultaneously, namely, soil moisture [*Bartalis et al.*, 2007; *Owe et al.*, 2008; *Naeimi et al.*, 2009], fraction of absorbed photosynthetically active radiation (FAPAR) [*Pinty et al.*, 2011], and column-integrated atmospheric carbon dioxide (XCO_2) [*Reuter et al.*, 2011]. These data are assimilated into two terrestrial biosphere models, which were integrated into CCDAS, namely, the Biosphere Energy-Transfer Hydrology (BETHY) model [*Knorr*, 2000], the Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH) [*Raddatz et al.*, 2007], the land surface scheme of the Max Planck Institute Earth System Model (MPI-ESM; M. A. Giorgetta et al., Climate change from 1850 to 2100 in

MPI-ESM simulations for the Coupled Model Intercomparison Project 5, submitted to *Journal of Advances of Modelling Earth Systems*, 2013).

[6] After giving a brief technical description of the current CCDAS methodology in section 2, the methodological progress of CCDAS will be described in section 3. Section 4 highlights some of the experience the team gained and challenges they had to face during the development and operation of CCDAS. Finally, section 5 draws various conclusions from this experience and should be of interest to the measurement, remote sensing, and modeling communities and those interested in exploiting large-scale data with complex models.

2. CCDAS Method

[7] CCDAS applies a variational data assimilation approach to estimate posterior process parameter values with their uncertainties, as well as posterior estimates of quantities we are interested in (target quantities), complete with uncertainties. The term “posterior” here stands for model simulations constrained by the observations. Potential target quantities are those that can be extracted from a model simulation such as carbon, water, and energy fluxes or stores, typically aggregated in space and time. If they can be extracted from a simulation over the observational period, we refer to them as diagnostic target quantities, and if they cover a period outside the observational period (either in the past or in the future), we refer to them as prognostic target quantities. These target quantities may reflect component processes not directly observed, but still the observations may be able to constrain them through the dynamics of the model. In this context, the inverse problem inherent in variational data assimilation consists of estimating a vector of parameters, x , that is linked to a vector of observations, d , via a function $d = M(x)$. The function M represents the terrestrial biosphere model including so-called observation operators that link the model state variables to observed quantities. For example in Figure 1, these include models of the radiative transfer within the canopy, dynamic calculation of the soil water balance, as well as a coupling to models of atmospheric transport.

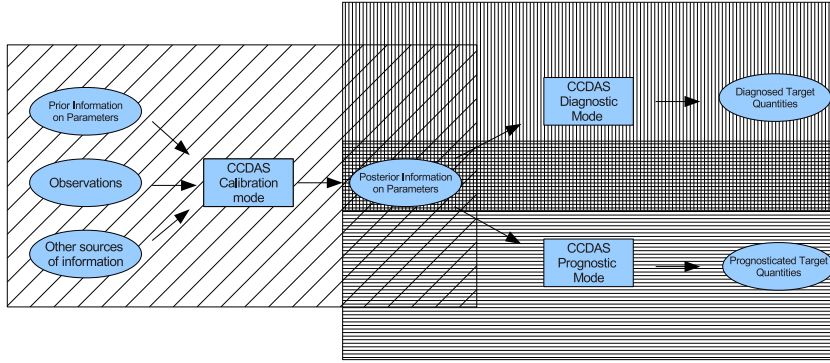


Figure 2. Schematic overview of two-step procedure for inferring diagnostic and prognostic target quantities from CCDAS. Rectangular boxes denote processes, and oval boxes denote data. The diagonally hatched box includes the inversion or calibration step, the vertically hatched box the diagnostic step, and the horizontally hatched box the prognostic step. Figure taken from *Scholze et al.* [2007].

[8] The CCDAS inversion methodology is conveniently formulated in a probabilistic framework [Enting, 2002; Tarantola, 2005], which means that each independent data item and any prior information on the parameters available (e.g., from literature or laboratory experiments) are represented by probability density functions (PDFs). Combining the information in the PDFs with the numerical model yields a posterior PDF for the parameters, i.e., the solution of the inverse problem. Typically, we treat all these PDFs (if necessary after a transformation) as Gaussian. Operating in this Gaussian framework, the relevant PDFs are given by a vector representing the mean and a matrix representing the covariance of its uncertainty: d , C_d for the observations, and x_{pr} , C_{pr} and x_{po} , C_{po} for the prior and posterior information, respectively. More precisely, the data uncertainty C_d accounts for uncertainties in the observations C_{obs} as well as uncertainties in simulating their counterpart (model error) C_{mod} . Provided these errors are both Gaussian they can be summed [Tarantola, 2005, equation (1.74)]

$$C_d^2 = C_{obs}^2 + C_{mod}^2 \quad (1)$$

[9] If (in addition to Gaussian prior and data PDFs) the model is linear, the posterior parameter PDF is Gaussian as well, with posterior mean (being also the maximum likelihood estimate)

$$x_{po} = x_{pr} + C_{po} M^T C_d^{-1} (d - M(x)) \quad (2)$$

and

$$C_{po}^{-1} = (M^T C_d^{-1} M)^{-1} + C_{pr}^{-1} \quad (3)$$

where the linear model is $M(x) = \mathbf{M}x$ and $()^T$ denotes the transposed. An example is inversion of the atmospheric transport of CO_2 [Enting, 2002], in which case x represents the CO_2 surface fluxes and \mathbf{M} the transport model.

[10] The term on the right-hand side of equation (3) equals the Hessian (matrix of all partial second derivatives) of the cost function, i.e.,

$$J(x) = \frac{1}{2} ((M(x) - d)^T C_d^{-1} (M(x) - d) + (x - x_{pr})^T C_{pr}^{-1} (x - x_{pr})), \quad (4)$$

which has its minimum at x_{po} .

[11] If the model is nonlinear, as is the case in CCDAS, iterative minimization of J is used to determine x_{po} , and

the posterior parameter uncertainty is approximated by the inverse of the Hessian of the cost function, evaluated at x_{po} :

$$C_{po}^{-1} \approx \frac{\partial^2 J}{\partial x^2} \quad (5)$$

[12] Once the posterior parameter uncertainty has been derived, it can be propagated forward to uncertainty in a vector of target quantities, $y = N(x)$, where N denotes the model operator for simulation of the target quantity and \mathbf{N} its linearization around x_{po} . Using \mathbf{N} , the posterior uncertainty in y can be approximated by

$$C_y^2 = \mathbf{N} C_{po} \mathbf{N}^T + C_{mod}^2 \quad (6)$$

[13] We note that the above formalism also applies when the parameter vector x is extended to a more general control vector that also includes initial or boundary conditions. For example, in the above mentioned case of atmospheric transport inversion, the control vector contains the time-dependent surface fluxes, i.e., a boundary condition.

[14] The CCDAS procedure is illustrated in Figure 2. First, a calibration step (diagonally hatched) computes the posterior parameter PDF, i.e., it infers x_{po} through minimization of equation (4) and approximates C_{po} via equation (5). Second, a diagnostic or prognostic step uses equation (6) to propagate posterior parameter uncertainties to a diagnostic (vertically hatched area) or prognostic (horizontally hatched) target quantity.

[15] Computationally, the minimization of equation (4) is performed by an efficient algorithm that relies on the gradient of $J(x)$. Furthermore, the second derivative of $J(x)$ is used to evaluate equation (5) and the first derivative of $N(x)$ to evaluate Equation (6). All derivative information is provided with the same numerical accuracy as the original model in an efficient form via automatic differentiation of the model code by the automatic differentiation (AD) [Griewank, 1989] tool Transformation of Algorithms in Fortran (TAF) [Giering and Kaminski, 1998]. TAF offers so-called forward and reverse modes of AD, which respectively produce tangent and adjoint codes for evaluation of first derivatives. Both produce the same results. Tangent code is generally more efficient in evaluating the derivative of a function, when the number of its dependent variables (outputs) exceeds the number of its independent variables

(inputs). Adjoint code is generally more efficient when the number of independent variables exceeds the number of dependent variables, as is usually the case for $J(x)$. Efficient Hessian code is generated by reapplying TAF in forward mode to a previously generated adjoint code. Jacobians are efficiently evaluated in vector mode of AD, which simultaneously propagates all components of the derivative vector. Further computational aspects of CCDAS are discussed by *Kaminski et al.* [2003].

3. Evolution of CCDAS

3.1. The First CCDAS

[16] CCDAS started with the development of the previously mentioned SDBM [*Knorr and Heimann, 1995*], a diagnostic model of terrestrial primary productivity. In contrast to similar models [see, e.g., *Ruimy et al., 1994*], SDBM was specifically designed to exploit high-precision measurements of CO₂ concentration from a global network of flask samples. Because atmospheric CO₂ concentration is only indirectly related to NPP, the study used SDBM coupled to the atmospheric transport model TM2 [*Heimann, 1995*], taking the role of an observation operator.

[17] It is useful to present SDBM in some detail because it can illustrate several essential elements of the CCDAS approach. The model computes the seasonal cycle of net ecosystem exchange (NEE) on a global grid as the difference between net primary production (NPP) and heterotrophic respiration and assumes an annually balanced NEE [*Knorr and Heimann, 1995*]. It is driven by remotely sensed vegetation greenness, incoming solar radiation, and surface temperature. The model has only two parameters: a photosynthetic light use efficiency ϵ and a parameter Q_{10} that determines the temperature dependency of heterotrophic respiration. Simulated NEE depends on Q_{10} in a nonlinear way and on ϵ in a linear way, i.e., Q_{10} determines the shape of the seasonal cycle and ϵ its amplitude. This form of dependency on Q_{10} and ϵ extends to the seasonal cycles in atmospheric CO₂ simulated by the linear TM2. For some ad hoc prior value of ϵ , the authors computed the combined SDBM-TM2 atmospheric response for a plausible set of Q_{10} values. For each value of Q_{10} , they determined a global scaling factor that minimizes the difference between the simulated and observed seasonal cycle at five atmospheric CO₂ monitoring stations of the NOAA/CMDL flask sampling network [*Conway et al., 1994*]. The optimal Q_{10} is the value that achieves the best fit after scaling and the optimal ϵ the scaled value of the prior ϵ . The procedure also tested varying ways in which drought stress was incorporated into NPP or heterotrophic respiration. The SDBM modeling framework provided a powerful tool not only to estimate global NPP but also to elucidate how drought differentially impacts the two contributors to NEE (i.e., NPP and heterotrophic respiration). Despite its simplicity, simulated NPP with SDBM was very similar to a range of more complex models [*Cramer et al., 1999; Bondeau et al., 1999*]. It repeatedly served as a benchmark in tests of complex models [*Heimann et al., 1998; Kelley et al., 2013*].

[18] A limitation of the SDBM framework is that both parameters (Q_{10} and ϵ) need to be global, whereas from an ecophysiological viewpoint, one would expect a difference

in ϵ between major vegetation types with widely different adaptation strategies, e.g., grass, broad-leaved trees, and conifers. However, such further differentiation would have made the stepwise optimization approach of the SDBM-TM2 framework infeasible. The first modeling framework [*Kaminski et al., 2002*] to overcome this limitation used SDBM in a version where a set of plant functional types (PFTs) was each assigned a separate pair of values for ϵ and Q_{10} . It used an efficient gradient-based algorithm, suitable for optimization of higher-dimensional parameter vectors, with one parameter pair per PFT. It also built on parallel work by *Kaminski et al.* [1999] who, through the adjoint of TM2, derived an efficient and accurate matrix representation of the mapping from CO₂ fluxes to concentrations at atmospheric sampling locations. The matrix representation by the Jacobian of TM2 allowed them to represent the transport by a simple matrix multiplication. Another feature of this first CCDAS framework was the computation of posterior uncertainties of both parameters (using equation (3)) and fluxes (using equation (6)). This first CCDAS can thus be considered a combination of components and functionality of the *Knorr and Heimann* [1995] study and a set of transport inversion studies based on the Jacobian of TM2 [*Kaminski et al., 1999*].

[19] Moving from transport inversions to the first CCDAS meant a drastic reduction of the dimension of the control space. While transport inversions optimize the time-dependent surface fluxes on the full transport model grid, CCDAS only used 24 parameters, 2 parameters each for 12 PFTs. As with *Knorr and Heimann* [1995], CCDAS took into account the effect of CO₂ fluxes from processes not represented in SDBM, termed “background” fluxes. Fluxes from fossil fuel burning, land use change, and between ocean and atmosphere were represented by prescribed fields. Observations were provided by 41 flask sampling sites [*GLOBALVIEW-CO₂, 2000*]. A new feature in this CCDAS framework was the use of automatic differentiation software [*Giering and Kaminski, 1998*], which efficiently provided the tangent and adjoint codes of the combined biosphere/transport simulation, required for the optimization and uncertainty propagation.

3.2. Prognostic Model

[20] The potential of this first CCDAS framework was limited by SDBM’s focus on the seasonal cycle and its diagnostic nature: Due to the need for remote sensing data to drive its photosynthesis model, SDBM is not suited for prognostic, i.e., predictive simulations of the terrestrial carbon cycle. Recognizing this shortcoming, *Knorr* [1997, 2000] had developed BETHY, a prognostic model of the terrestrial carbon cycle. The model is structured into four compartments: (1) energy and water balance, (2) photosynthesis, (3) phenology, and (4) carbon balance. It is run on the basis of daily climate data, which is internally converted into hourly microclimate. BETHY decomposes the global terrestrial vegetation into 13 PFTs based on the specification by *Wilson and Henderson-Sellers* [1985]. Each grid cell contains up to three PFTs.

[21] The integration of BETHY into CCDAS was performed stepwise. A first series of studies [*Rayner et al., 2001; Scholze, 2003; Rayner et al., 2005a*] focused on the assimilation of atmospheric CO₂. For that purpose, it

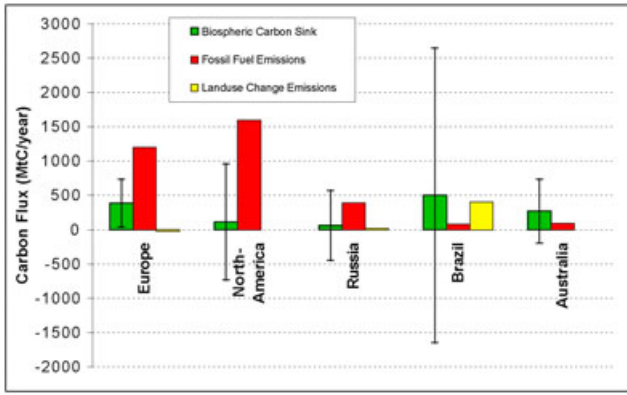


Figure 3. From 1980 to 1999 averages of NEE aggregated over five regions from Carbon-BETHY (with 1 sigma uncertainty range), fossil fuel emissions, and land use change in megaton C/yr.

was sufficient to operate CCDAS on Carbon-BETHY, a reduced version of BETHY that does not include a water balance nor a phenology scheme. A preceding run of the full BETHY scheme provided to Carbon-BETHY fields of leaf area index (LAI) and plant available soil moisture optimized against remotely sensed FAPAR. Carbon-BETHY could then simulate carbon fluxes as a function of 20 process parameters controlling photosynthesis and autotrophic and heterotrophic respiration. The parameter values are constant in time, and their spatial scope can be selected by the user in a flexible manner. Parameter values can, for example, be specific to grid cells, regions, PFTs, or be globally valid. In the CCDAS default configuration, there are three PFT-specific parameters, and the remaining 17 parameters are globally valid. With one extra global parameter for the initial atmospheric CO_2 concentration, this amounts to $3 \times 13 + 17 + 1 = 57$ parameters. The model uses two soil carbon pools, representing fast and slowly decomposing organic matter. One of the PFT-specific parameters, β , allows to scale heterotrophic respiration and implicitly determines the initial size of the slow carbon pool and thus avoids a spin up of the pool. Since the relative change in pool size is typically small, this change in pool size over the assimilation period is neglected [Rayner et al., 2005a].

[22] For this default configuration, atmospheric flask samples provided by 41 sites [GLOBALVIEW- CO_2 , 2004] over the period from 1980 to 1999 were used to calibrate the model [Rayner et al., 2001; Scholze, 2003; Rayner et al., 2005a]. As for SDBM, the atmospheric transport of the simulated fluxes plus a set of background fluxes to the observational sites was simulated by TM2. The system achieved a considerably improved fit to the observations. It produced parameter values including uncertainty ranges as well as net ecosystem production (NEP) with uncertainty ranges at the grid-scale level and also aggregated to regions as illustrated by Figure 3. The figure clearly shows the value of the uncertainty ranges in the assessment of regional carbon budgets. In technical terms, this generation of CCDAS is the first to derive posterior uncertainties from the full Hessian (equation (5)).

[23] The limitation of the above CCDAS setup through the use of prescribed background fluxes was tackled by a

further sequence of studies. Hooker-Stroud [2008] replaced the fossil fuel background flux through a model of the fossil fuel emissions which was calibrated jointly with Carbon-BETHY. Scholze et al. [2013] replaced the oceanic background flux by the MIT general circulation model of the ocean, including the dissolved inorganic carbon marine carbon cycle model [Dutkiewicz et al., 2006]. The ocean component adds 13 marine carbon parameters to be estimated jointly with the standard CCDAS parameters. A further example for an extension of the process model is provided by Kelley [2008] who included a diagnostic fire model into CCDAS and calibrated it jointly with Carbon-BETHY.

[24] Another strand of activities explored the CCDAS configuration options. Ziehn et al. [2009, 2011] investigated the sensitivity of CCDAS results with respect to the spatial resolution of parameters by introducing a differentiation by PFT and region for the key carbon storage parameter β . The studies of Koffi et al. [2012a, 2012b] investigated the sensitivity of CCDAS results with respect to the observational network and the atmospheric transport model. Koffi et al. [2012b] also quantify how atmospheric CO_2 sampling, an observation type sensitive to NEP, constrains gross primary productivity (GPP), a quantity further up the modeling chain of NEP.

[25] The prognostic capability of CCDAS was exploited by a further sequence of studies that used the network and assimilation period of Rayner et al. [2005a]. Scholze et al. [2007] simulated a prognostic period of only 4 years subsequent to the assimilation period and kept the approximation of a constant slow carbon pool. By contrast Rayner et al. [2005b, 2011] simulated a prognostic period of 50 and 90 years, respectively, and included the dynamics of the slow pool. Target quantities included prognostic fluxes [Scholze et al., 2007; Rayner et al., 2005b; Rayner et al., 2011] and concentrations [Scholze et al., 2007]. The uncertainty reduction relative to the prior was used to assess the impact of the observational constraint. All studies found considerable uncertainty reductions (about 90% for the 90 year prediction and above 95% for the 4 year prediction).

3.3. Inclusion of Phenology and Water Cycle

[26] The focus of Carbon-BETHY on the terrestrial carbon cycle was useful to simplify the setup and operation of CCDAS. However, this focus also restricted the type of observations that could be assimilated into the model, as well as the type of potential target quantities. An essential development step to reduce this restriction was the extension of CCDAS to the full BETHY model, i.e., the inclusion of its hydrology and phenology compartments.

[27] One of the associated challenges was the lack of a “smooth” (i.e., differentiable) response of the model state (and thus of the cost function of equation (4)) to changes in process parameters caused by BETHY’s phenology scheme [Knorr, 2000]. Not only is a high sensitivity of the cost function to very small parameter changes questionable from a biophysiological point of view, but it also hampers the use of variational assimilation techniques. This is because these techniques rely on the validity of the derivative information in a certain neighborhood of any point in parameter space. A smooth response to parameter changes was achieved [Knorr et al., 2010] through a newly developed phenology scheme and a revised soil evaporation model with a new shallow

surface bucket overlapping a deep-root zone bucket. The modified BETHY captures temperature- and water-driven leaf development. Taking spatial variability within a grid cell into account avoids hard switches between growth, senescent, and dormant vegetation states. The new components add another 8 parameters to the 20 parameters of Carbon-BETHY. In the new default setup, some of the parameters are globally valid, while others are shared between selected PFTs.

[28] *Knorr et al.* [2010] applied the extended CCDAS for simultaneous assimilation of 20 months of remotely sensed FAPAR over seven sites covering seven PFTs. The FAPAR product [*Gobron et al.*, 2007; *Gobron et al.*, 2008] was derived from the Medium Resolution Imaging Spectrometer on board the European Space Agency’s ENVISAT platform. In that multisite configuration, the model used 38 process parameters. The smoothness of the extended CCDAS was demonstrated by efficient minimization of the cost function (equation (4)) in some 30 iterations, with a gradient reduction by more than 7 orders of magnitude (a great improvement on previous versions). The robustness of the parameter estimate was assured by starting the iterative procedure from three different points in parameter space each on two different computers. The optimization improved the data fit at all seven sites. More importantly, it also reduced the root-mean-square (RMS) difference at a further site that was not included for assimilation (validation site) by more than 40%. The study also demonstrated how data assimilation can extend the temporal scope of the assimilated information by simulating NPP over a 48 months period and showed a substantial reduction of posterior uncertainty in mean annual NPP, which was 48% at the validation site.

[29] The first global-scale application [*Kaminski et al.*, 2012a] of the extended CCDAS simultaneously assimilated the FAPAR product and flask samples of atmospheric CO₂ at two sites on a coarse grid to estimate 70 terrestrial process parameters plus one initial atmospheric concentration. It demonstrated the same robustness of the minimization procedure, where four out of five runs from different starting points converged to the same minimum with gradient reductions by more than 8 orders of magnitude in about 150 iterations. This robustness was a direct consequence of the smoothness of the revised phenology scheme. The calibrated model showed an improved fit also at atmospheric flask sampling sites that were not included in the assimilation.

[30] *Kato et al.* [2013] used an extended site level setup for simultaneous assimilation of FAPAR and eddy covariance measurements [*Baldocchi et al.*, 2001] of latent heat flux at the FLUXNET site in Maun, Botswana [*Veenendaal et al.*, 2004]. They used a FAPAR product [*Gobron et al.*, 2006] derived from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) of the National Aeronautics and Space Administration (NASA). In their two-PFT configuration, they estimated 24 model parameters. Comparison against eddy covariance measurements of GPP showed that the calibration reduced the RMS difference by 16%. They also tested individual assimilation of the two data streams, which resulted in considerable differences in some of the parameter values and a substantial degradation of the fit to the nonassimilated data stream compared to the prior. It was the simultaneous assimilation of both data streams that achieved

the compromise between the two suboptimal states reached after assimilating only one data stream.

3.4. Earth System Model

[31] The studies of *Scholze et al.* [2007] and *Rayner et al.* [2005b, 2011] clearly demonstrated the potential of BETHY-CCDAS to reduce predictive uncertainty in terrestrial carbon cycle simulation. Since BETHY-CCDAS was designed as a model driven by observed or reconstructed meteorological forcing, its application to analyze global carbon cycle trajectories in the Earth system is limited. The next step in CCDAS development was thus to extend the framework to a land-surface model that is capable of simulating the land component of the global carbon cycle coupled to representations of the marine and atmospheric branches of the global carbon cycle. The logical choice of land-surface model is the terrestrial component of the Max Planck Society’s ESM (MPI-ESM) [*Jungclaus et al.*, 2010], called Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH) [*Raddatz et al.*, 2007], because the JSBACH development was originally based on BETHY. Despite this fact, there are considerable differences between the two models, both in terms of process representations and code structure. While the light absorption and photosynthesis representation follow BETHY, the soil energy and water balance calculations follow the scheme of the atmospheric model ECHAM5 [*Roeckner et al.*, 2003]. A complementary five-layer soil hydrology has been developed to improve simulation of the seasonal hydrological cycle [*Hagemann and Stacke*, 2013]. JSBACH further simulates the allocation of carbon assimilates to vegetation, litter and soil organic matter pools, represented by seven pools of different live times, and thus explicitly represents heterotrophic and autotrophic respiration processes, the latter based on the formulation in BETHY (a description is given in *Goll et al.* [2012]). JSBACH includes modules for land-cover change [*Reick et al.*, 2013], disturbance regimes and vegetation dynamics [*Brovkin et al.*, 2009], and nutrient cycles for nitrogen and phosphorus [*Goll et al.*, 2012]. JSBACH can be run as a component of the comprehensive Earth System Model (online), as a coupled atmosphere-land model with prescribed sea-surface properties, or standalone (off-line), driven with prescribed atmospheric forcing, with the same, identical code base.

[32] In the CCDAS setup, JSBACH is employed in off-line mode, driven by reconstructed, observed meteorology (meteorological reanalysis products) [*Schürmann et al.*, 2013]. The default observational operator for atmospheric concentrations is the TM3 atmospheric transport model [*Heimann and Körner*, 2003]. TM3 is available in various spatial resolutions (e.g., 8° by 10° to 2° by 2.5° horizontally) and, as TM2, offers precomputed Jacobians for many observational sites [*Rödenbeck et al.*, 2003]. As a difference to BETHY-CCDAS-TM2, TM3 is typically driven with interannually varying winds.

[33] To render the dependency of the cost function on the parameters as smooth as possible, the model’s default phenology scheme logro-P [*Raddatz et al.*, 2007] was replaced by the smooth scheme of *Knorr et al.* [2010]. While both the bucket and five-layer soil schemes as well as the carbon cycle of JSBACH are included in CCDAS, the recently developed nutrient cycles are not yet considered. Also, dynamic vegetation has been excluded and CCDAS

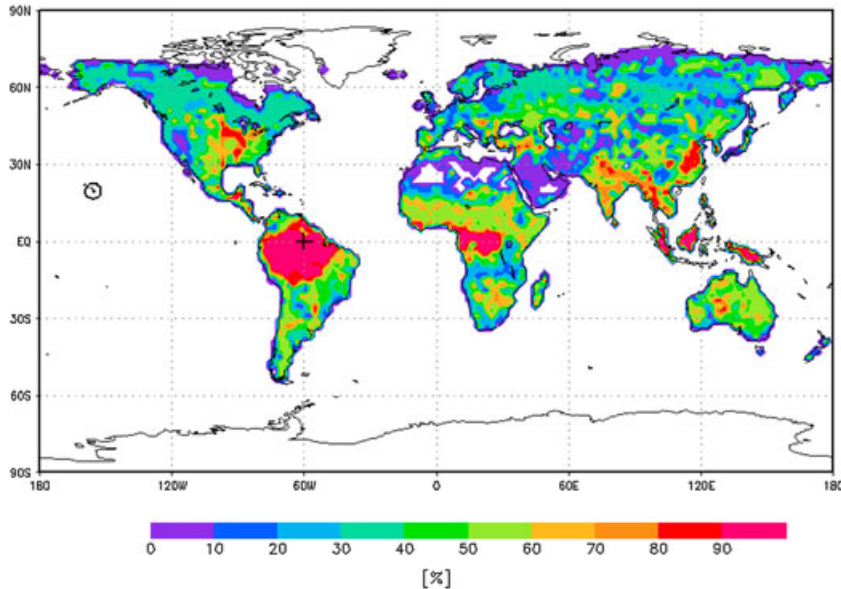


Figure 4. Uncertainty reduction in NEP for a network composed of one eddy covariance site (denoted by a cross) and a flask sampling site (denoted by a circle with a dot).

operates with prescribed static land-cover maps. Currently, tangent linear JSBACH codes are available in scalar and vector forms, while development of adjoint and Hessian codes is ongoing. First successful assimilation experiments have been carried out at site level and global scale [Schürmann *et al.*, 2013]. The site level setup can assimilate eddy covariance-based observations of NEE or latent heat fluxes and the global scale setup atmospheric CO₂ observed at long-term monitoring stations. Control variables are the model’s process parameters and initial conditions. With slight revisions in details of the process formulations (see section 4.3), the sensitivity to process parameters is in a reasonable range and the minimization of equation (4) proceeds efficiently without being terminated at discontinuities.

3.5. Quantitative Network Design

[34] One of the innovative aspects of the study by Kaminski *et al.* [2002] was that it tested the impact of a hypothetical direct flux observation on the posterior uncertainty within CCDAS-SDBM. Since the observation was hypothetical, one could not compute its impact on the posterior mean value that minimizes equation (4). It was possible, however, to use a plausible data uncertainty C_d together with the Jacobian \mathbf{M} that links the parameter vector to the hypothetical observation and then evaluate equation (3). The hypothetical flux observation was placed in the model’s broadleaf evergreen PFT, which was not well observed by the atmospheric network. The extra observation resulted in a substantial uncertainty reduction for both model parameters associated with this PFT, indicating the complementarity of the two data streams. The study by Kaminski *et al.* [2002] was the first application of quantitative network design (QND) to a terrestrial biosphere model, a technique that was originally introduced to biogeochemistry [Rayner *et al.*, 1996] in the context of atmospheric transport inversions. Its principle is the application of the CCDAS uncertainty propagation step (equation (3) or (5)) without a

preceding minimization. For details on the method, we refer to Kaminski and Rayner [2008].

[35] For Carbon-BETHY, an interactive QND tool (available at <http://IMECC.CCDAS.org>) was developed [Kaminski *et al.*, 2012b]. The tool is called Network Designer and handles three observational data streams, namely, direct CO₂ flux observations, continuous, and flask samples of atmospheric CO₂. The user can compose an observational network by selecting from a list of locations (for which Jacobians have been precomputed) and by specifying a corresponding data uncertainty. The Network Designer returns posterior uncertainties for a set of target quantities. These are regional integrals of long-term NEP and NPP, as well as NEP on the model grid. As an example, Figure 4 shows the uncertainty reduction in NEP relative to the prior for a simple network composed of an atmospheric flask sampling site at Mauna Loa (155.58°W, 19.53°N) and an eddy covariance site in the tropical rain forest (60°W, 0°). Kaminski *et al.* [2012b] applied the tool to assess the complementarity and redundancy of terrestrial and atmospheric networks. They also explored the sensitivity of network performance to the degree of vegetation heterogeneity by varying the globally available number of PFTs between 13 and 325.

[36] The QND study by Koffi *et al.* [2012a] applies CCDAS to assess the constraint of several combinations of terrestrial and atmospheric networks on Carbon-BETHY’s process parameters and tested the impact of the atmospheric transport model. Another QND study [Kaminski *et al.*, 2010] used the CCDAS around Carbon-BETHY to assist the design of a space mission. It assessed the benefit of sampling XCO₂ with a hypothetical active LIDAR instrument [Ehret *et al.*, 2008] for alternative mission layouts. For the assessment of hypothetical FAPAR observations from space with QND methods, Kaminski *et al.* [2012a] derived a Mission Benefit Analysis (MBA) tool from CCDAS around full BETHY. The tool quantifies the effect of data uncertainty and mission length and availability of complementary

atmospheric CO₂ observations on the constraint on carbon and water fluxes. The MBA tool quantified the separate and combined constraint of FAPAR and atmospheric CO₂ observations on carbon and water fluxes.

4. Challenges, Experiences, and Lessons Learnt

[37] This section provides a set of recommendations based on our experience with CCDAS development and operation.

4.1. Carefully Select the Control Vector

[38] The selection of control variables should be generally guided by the resolution of the large uncertainties in the terrestrial system. Sources of uncertainty in a simulation of the terrestrial biosphere are the following: the selection of the relevant processes to incorporate in the model, their formulation, the approximations made in their numerical implementation (structural uncertainty); the values of the parameters in the implementation of the processes (e.g., ϵ and Q_{10} in the case of SDBM) (parametric uncertainty); the initial state of the system; and the atmospheric forcing, in the case of an off-line simulation.

[39] Sources 2 to 4 can be directly addressed in the CCDAS by incorporating uncertain process parameters, initial conditions, and atmospheric forcing in the control vector of the CCDAS. To explore structural uncertainty (source 1), one can modify (within the CCDAS) process formulations and their implementation and study their performance. An example for this type of analysis is the impact of drought stress in SDBM discussed in section 3.1. To resolve the large uncertainties for a given model implementation, one would generally select those control variables that are deemed uncertain and have high impact on observables and target quantities. If a variable is excluded from the control vector, we have to take its effect into account in the model error contribution to the data uncertainty (equation (1)). An increased data uncertainty will reduce the weight of the data in the cost function (equation (4)) relative to the prior information, i.e., we can learn less from the data.

[40] In some cases, there are parameters that do not individually impact the model's fit to available observables, but only in a given combination with another parameter, for example, as a product. This renders the inversion underdetermined, because there are many pairs of the parameter values that yield identical values for the observables (a situation sometimes termed equifinality). If, in addition, the relevant target quantities are also only sensitive to the same parameter combination, we recommend solving for the combination of the two parameters. An example is the first implementation of the Farquhar photosynthesis model [Farquhar *et al.*, 1980] in Carbon-BETHY [Scholze, 2003], where the two parameters j_{mt} and j_{tv} exclusively act as a product. The two parameters describe, respectively, the ratio of J_{\max} to temperature and the ratio of J_{\max} to V_{\max} , where J_{\max} is the maximum electron transport rate and V_{\max} the maximum enzyme carboxylation rate. In later CCDAS implementations [Rayner *et al.*, 2005a], the parameter j_{mt} was absorbed by j_{tv} , i.e., a parameter $j_{tv} = j_{mt} \times j_{tv}$ was used to replace the individual parameters.

[41] The CCDAS formalism described in section 2 assumes Gaussian parameter PDFs. This restriction is weakened by the implementation of a set of parameter transformations, which map a Gaussian parameter, x , from the space in which the inversion is formulated onto a parameter, p , in physical space in which the model operates. These include functions like $p = \exp(x)$ or $p = x^2$ which exclude negative values of p . Suitable choice of scaling factors assure that the Gaussian uncertainty range is mapped onto the desired range in physical space. For example, such transformations are typically applied to the Q_{10} parameters of slow and fast carbon pools [Rayner *et al.*, 2005a], in Ziehn *et al.* [2011] also to the β parameters.

[42] Even though such parameter transformations can achieve a wide range of PDF shapes in physical space, the inversion procedure can only adapt the mean and variance in x space, i.e., much of the PDF shape is prescribed a priori. Markov Chain Monte Carlo methods are a class of inversion algorithms that avoid any restriction on the posterior PDF shape through frequent sampling of the parameter space. Such algorithms are thus suitable for examining the severity of the Gaussian assumption in CCDAS. Due to the considerable computational requirements, which grow with the dimension of parameter space and model complexity, for BETHY, the method requires the combination of a fast model configuration and a low dimensional parameter space. Knorr and Kattge [2005] applied the Metropolis algorithm [Metropolis *et al.*, 1953] to calibrate a reduced version of BETHY in two different site level setups (with 14 and 23 parameters, respectively) against eddy covariance measurements of NEE and latent energy with an assimilation period of a week. One of their main findings was that the posterior parameter PDFs were close to Gaussian. Ziehn *et al.* [2012] compared the Metropolis algorithm and CCDAS for a reduced global setup of Carbon-BETHY. They calibrated 19 parameters against the same atmospheric CO₂ observations as Rayner *et al.* [2005a]. The agreement of posterior parameter values and uncertainties was generally good. Remaining differences were attributed to convergence problems of the Metropolis algorithm.

4.2. Take Biases Into Account

[43] The CCDAS formalism described in section 2 assumes that both data and model are unbiased. This means the mean values of their Gaussian PDFs correspond to the true value and to each other. An example of a biased model would be a model that systematically overestimates soil respiration. In the presence of biases the optimization will attempt a compensation through a biased posterior parameter estimate. In the above example, the optimization could adjust the parameters of the photosynthesis model such that enhanced GPP compensates the bias in soil respiration resulting in NEE that matches observed atmospheric CO₂. If the biases in model and observations were known, we could subtract them before computing the model-data mismatch in equation (4). For example, some data products are provided in bias-corrected form. For the model, we generally aim at avoiding biases through model improvements; the second best choice is to correct the biases.

[44] Absolute biases are difficult to assess, because the truth is generally not known. We can, however, use the

model-data differences to evaluate the difference in their biases and try to correct for this relative bias or at least guarantee that the statistical assumptions underlying the assimilation are respected. This is complicated by the fact that part of the bias can be caused by suboptimal parameter values. We only want to correct for the residual fraction of the bias with optimal parameter values. On the other hand, for an unbiased optimization, we need to remove the bias beforehand. An obvious solution is to construct a model of the bias that can be calibrated along with the process model.

[45] As an example, we use the setup of *Kaminski et al.* [2012a] who calibrate BETHY by simultaneous assimilation of atmospheric CO₂ and FAPAR data (see section 3.3). Based on site level analyses [*Knorr et al.*, 2010], we suspect a low bias due to cloud contamination of the FAPAR data (F_{obs}) and construct a model for the corresponding bias, B , to be added to F_{obs} . Given that observed FAPAR is low compared to a prior model simulation in some evergreen tropical rainforest areas with dense vegetation, but agreement is good in areas of no vegetation, we assume that the bias itself is proportional to F_{obs} . One consequence is that if the observed FAPAR is 0 then the bias-corrected FAPAR will always be 0 as well. We also assume that the bias is increasing with precipitation P (as a proxy for cloud cover), which results in

$$B = (a + bP)F_{obs}, \quad (7)$$

with two tunable parameters, a and b . A joint calibration with BETHY yielded parameters of the phenology model set such that the simulated LAI remained very small, with maximum LAI values globally well below 1, which is obviously unrealistic. The reason was that by reducing simulated LAI and thus FAPAR and at the same time reducing observed FAPAR via $B < 0$, the optimization achieved the biggest reduction in the cost function. Despite the corresponding low NPP value, the optimization achieved a good fit to the atmospheric CO₂ observations by slowing down soil respiration. After the failure of this integrated bias correction approach, *Kaminski et al.* [2012a] stepped back and applied a bias correction before the assimilation, despite the above mentioned disadvantages.

[46] Another data stream that typically requires bias correction is surface soil moisture as provided by several remote sensing instruments. State of the art schemes for bias correction attempt a scaling to match mean and variance [*Scipal et al.*, 2008a; *Dorigo et al.*, 2010] or a more complex transformation that matches the cumulative distribution functions (CDF matching) [*Drusch et al.*, 2005]. If a third independent data set is available, the so-called triple collocation analysis [*Scipal et al.*, 2008b] can be applied to estimate both biases and uncertainties. Another related topic is quality assurance of the observations, i.e., rejection of erroneous data before assimilation [see, e.g., *Hollingsworth et al.*, 1986]. Bias correction and quality assurance are areas where carbon cycle data assimilation should learn from the long experience in, for example, numerical weather prediction where complex bias correction and quality assurance schemes are the norm.

4.3. Be Prepared To Modify Your Model

[47] When integrating a model into the CCDAS framework, you need to be prepared for modifications. Some of these modifications address slight structural changes in the

code, e.g., to nominate control variables (see section 4.1) or to achieve compliance with automatic differentiation software (see below). These usually do not affect the model results. A further class of modifications are those which are useful to improve the model performance within CCDAS and beyond. An example is the newly developed phenology scheme that exhibits a smooth dependency of the vegetation state on the process parameters. The scheme is now implemented in BETHY (see section 3.3) and JSBACH (see section 3.4). Not only does the scheme considerably improve the model performance in the gradient-based optimization framework. The underlying assumption of subgrid variability made the model more realistic. In fact, one of the major uses of data assimilation in terrestrial modeling is to separate model deficiencies due to poor parameter choices from more fundamental structural problems. Only once the parameters have been optimized can we be sure they are not the cause.

[48] A further example is ongoing refinement of BETHY's soil scheme. This modification was triggered by a poor match of the statistical distributions of modeled and remotely sensed surface soil moisture, a new CCDAS data stream being assimilated within the above mentioned CarbonFlux project. Similarity of the statistical distribution of daily soil moisture values between the BETHY surface layer and remotely sensed surface soil moisture is a prerequisite for assimilating these data [*Drusch et al.*, 2005].

[49] Sometimes it is useful to modify low-level details in the process formulations that produce unrealistic sensitivities. In the following, we present three examples experienced in CCDAS development around JSBACH to highlight the nature of such changes. A first example is the use of the calculation of the fraction of snow cover [*Roesch et al.*, 2001, equation (7)], which includes the square root of the amount of snow. Unfortunately, the amount of snow often is zero which leads to a division by zero in the derivative code. A solution here is to define a lower limit to the value for which the square root is taken. Another example is the calculation of the specific humidity at saturation, which originally was a step function read from a look-up table that was precomputed in the MPI-ESM. A third typical case is the use of maxima or minima functions which lead to nondifferentiable points in otherwise continuous functions. These can be smoothed by defining a maximum or minimum function with an exponential transition from one value to the other. These three examples illustrate aspects of the model that only became apparent during CCDAS development. The corresponding modifications can be regarded as model improvements.

[50] The use of AD facilitates updates of CCDAS after modifications. This requires compliance of the model code with an AD tool. To some extent, the process of achieving compliance with an AD tool is similar to porting the model code to a new compiler and includes revision of particular code constructs not handled by the AD tool, without changing the model results. From our experience, the main effort is required to achieve compliance for an initial model version. From that point, we recommend development of the model within CCDAS. Adapting the derivative code to the day-to-day modification of the model then typically requires little effort. Of course, inclusion of new process models, such as the step from Carbon-BETHY to full BETHY requires a larger effort again. As programming standards

such as Fortran are continuously evolving, it is desirable that the AD tool is continuously maintained and evolves with the standards.

4.4. Attempt Consistent Assimilation

[51] An example of a setup for assimilation of multiple data streams is shown in Figure 1. A consistent view on the terrestrial carbon cycle can only be achieved by assimilating all data simultaneously (consistent assimilation). Stepwise procedures appear less demanding but consistency can only be ensured for linear models and requires that all uncertainties from one step be propagated to the next, negating the apparent advantage. Typically, the fit to a data item assimilated in an early step will be degraded by later assimilation steps. For non-linear systems (as the terrestrial biosphere) only joint assimilation can fully exploit the complementarity of the observational constraints. We also note that stepwise assimilation procedures based on a split of the entire assimilation period in subperiods is prohibitive when the control vector affects the initial state. This is because such a sequential assimilation procedure would result in the violation of mass conservation, a crucial requirement in our context. Section 3 has summarized some early examples of consistent assimilation. In *Knorr et al.* [2010] they were eddy covariance observations at multiple sites, in *Kaminski et al.* [2012a] FAPAR and atmospheric CO₂, and in *Kato et al.* [2013] FAPAR and eddy covariance measurements of latent heat flux at a site. These cases were all challenging and should be regarded as a first step toward consistent assimilation. They all led to improvements of the process model formulation.

[52] The relative strengths of the constraints through the individual data streams on the prior depend on the data and prior uncertainties in equation (4). A too low uncertainty on a data stream will overemphasize the importance of that data stream in the integrated CCDAS view.

[53] Assessment of prior and data uncertainties is difficult. Observational products are constantly improving, and there is a tendency toward provision of uncertainty ranges with the products [*Pinty et al.*, 2011; *Reuter et al.*, 2011]. The dependence of uncertainties of eddy covariance fluxes on the flux magnitude itself and their autocorrelation structure have been described [*Richardson et al.*, 2006; *Lasslop et al.*, 2008] and research is still ongoing to further characterize random and systematic uncertainties [e.g., *Richardson et al.*, 2012; *Mauder et al.*, 2013].

[54] For the model error contribution to data uncertainty, there is little guidance. We usually assume prior and data uncertainty to our best knowledge and apply CCDAS to assess the consequences of these assumptions. A test of these assumptions is the value of the cost function at the optimum, which multiplied by 2 and normalized by the dimension of the data space should be around 1 [*Tarantola*, 2005]. For example, *Rayner et al.* [2005a] calculate a value of 2.76. To achieve a value of 1, they could have increased the standard deviations in their diagonal prior and data uncertainties by a factor of $\sqrt{2.76}$, i.e., by about 66%. Correlations within the data uncertainty also have a high impact on data weight. If, for example, the uncertainty for an entire data stream is fully correlated, it takes the same weight as a single observation with the same data variance. Some of these problems can be diagnosed by careful consideration of the statistical

behavior of the model fit [e.g., *Michalak et al.*, 2005; *Kuppel et al.*, 2012a].

[55] The inclusion of processes is so far only represented by prescribed background flux fields as those mentioned in section 3.2 are modifications that make further data streams accessible to CCDAS. At the same time, this type of modification can help to resolve uncertainty that before had to be accounted for in the model error contribution to the data uncertainty (equation (1)).

4.5. Handling an ESM

[56] Rather than including the entire MPI-ESM into the CCDAS, we have, as a first step, chosen to include the JSBACH as a standalone model, driven by observed or reconstructed meteorology. This allows to identify potential model structural defaults of the land-surface model in a better constrained setup, rather than attempting to reduce model-data mismatches by optimizing atmospheric and terrestrial processes simultaneously. The standalone model can be operated at the site level, driven with observed meteorological forcing, which allows to assimilate observations representative for a particular PFT as observed at a set of specific sites representing this PFT (e.g., eddy covariance observations) and potentially identify model structural failures. The use of interannually varying global carbon cycle observations in the JSBACH-CCDAS, such as the global atmospheric CO₂ monitoring network, requires that the meteorological forcing of JSBACH represents not only the climatological mean correctly but also the correct timing and magnitude of seasonal and interannual anomalies, e.g., the occurrence of meteorological teleconnections to the El Niño–Southern Oscillation phenomenon. For this reason, forcing data are taken from reanalyses or interpolated observations [e.g., *Weedon et al.*, 2011] rather than a free run of the MPI-ESM, in which timing and magnitude may differ substantially from observed patterns. In addition, a standalone JSBACH run is computationally much faster than a full ESM run and the amount of code is considerably reduced. Even though (*Blessing, S. et al.*, Testing variational estimation of process parameters and initial conditions of an Earth System Model, submitted to *Tellus A*, 2013) recently applied TAF to generate efficient tangent and adjoint codes of an entire ESM, the limited code size of JSBACH's off-line version clearly facilitates automatic generation of derivative code.

[57] Nevertheless, since the ultimate purpose of JSBACH is to correctly represent the land carbon cycle in the Earth system model, it needs to be thoroughly tested to what degree the offline calibration and initialization of the JSBACH affect the global carbon cycle as simulated by the MPI-ESM. In configurations with compensating errors, the immediate effect of improving a source of error may be a degradation in the performance of the entire system. *Dalmonech et al.* [2013] show that climate biases of the MPI-ESM can lead to substantial differences between the JSBACH's global carbon cycle projections when driven with observed or MPI-ESM-generated climatic forcing fields. Future work will need to establish how model calibration with observed climate will affect the carbon cycle trajectories within the fully coupled Earth system. Another aspect related to assimilation into an ESM is the enhanced nonlinearity of the model as demonstrated, e.g., by *Lorenz* [1963].

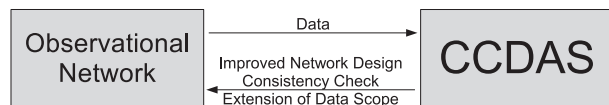


Figure 5. Mutual benefit of CCDAS and observations.

With a control vector composed of dynamical atmospheric and oceanic state, the length of the feasible assimilation period is limited to a time span over which the derivative provides a useful approximation (see, e.g., Blessing, S. et al., submitted manuscript, 2013). For parameters controlling the terrestrial carbon cycle, one might argue that their influence on atmosphere and ocean dynamics acts on long enough time scales (decades and longer) [Gregory et al., 2009] to avoid a highly nonlinear response of our cost function. However, the interaction path through the hydrological cycle is faster. The effect of these interactions on the data assimilation procedure requires further investigation.

5. Conclusions

[58] From our CCDAS experience, we conclude that integrated use of observations and models in a CCDAS is beneficial for experimentalists, remote sensing experts, and modelers. As illustrated by Figure 5, the modelers benefit from the constraints provided by the observational data on their process formulations. On the other hand, CCDAS is beneficial for observationalists in several respects. First, CCDAS provides a consistency check among the data (types) that are assimilated and between the data and the process formulation in the model. Consistency means that the model can simultaneously match all observational data streams within their uncertainty ranges. This was illustrated by simultaneous assimilation of FAPAR observations over multiple sites [Knorr et al., 2010] or simultaneous assimilation of atmospheric CO₂ and FAPAR [Kaminski et al., 2012a]. Second, CCDAS allows us to extend the information contained in the data in time and space, as well as to build bridges between the available observations and hitherto unobserved quantities. This means it can use observational data to constrain a model-based simulation of quantities other than those being observed, beyond the observational period, and beyond the observational domain. For example, Scholze et al. [2007] inferred CO₂ surface fluxes for the period from 1980 to 2003 from atmospheric CO₂ observations from 1980 to 1999 and Knorr et al. [2010] constrained NPP over a 48 month period, through assimilation of 20 months of FAPAR observations.

[59] CCDAS can thus be regarded as an instrument that allows us to enhance the observational information and to derive higher-level products. These reanalysis products combine the observational information and the process understanding in an integrated view on the terrestrial carbon cycle. The cost is a limited view of phenomena consistent with the (imperfect) model dynamics. This is why we have stressed the importance of proper validation. The data flow through the CCDAS processing chain (Figures 1 and 2) is traceable and documented, from the input observations with their uncertainties to the simulated reanalysis products with their posterior uncertainty. This is supported by keeping CCDAS model codes under version control. Third, a CCDAS can help to improve the design of the obser-

vatational network, i.e., it can suggest which quantities to observe when and where in order to extract the maximum of information on a given aspect of the simulation.

[60] As illustrated by some of the studies reported, operation of a CCDAS is sometimes challenging. To extract the maximum benefit from a CCDAS requires the combined expertise of observationalists and modelers. The success of the concept is manifested by its application to further terrestrial models, namely, the Joint UK Land Environment Simulation [Clark and Harris, 2007] and ORCHIDEE [Krinner et al., 2005], for which Luke [2011] and Kuppel et al. [2012b], respectively, present site-scale applications. The Earth Observation Land Data Assimilation System [Lewis et al., 2012] is another effort that uses the CCDAS concept with focus on EO data. It applies a weak constraint variational assimilation approach that allows for deviations from the dynamics of a simplified, highly flexible land surface model.

[61] **Acknowledgements.** CCDAS work was supported by the Max Planck Society; the Commonwealth Scientific and Industrial Research Organisation; the European Community within the 5th, 6th, and 7th Framework Programmes for Research and Technological Development under contracts EVK2-CT-2002-00151, FP6-511176, FP6-026188, FP7-283080, and FP7-264879; the European Space Agency and the Natural Environment Research Council, UK, through its QUEST Programme; the National Centre for Earth Observations and the advanced research fellowship to M. Scholze; and Germany's Federal Ministry of Education and Research through the research programme MiKlip (FKZ 01LP1108A/B). Rayner is in receipt of an Australian Professorial Fellowship (DP1096309).

References

- Arora, V. K., et al. (2013), Carbon-concentration and carbon-climate feedbacks in CMIP5 Earth system models, *J. Clim.*, 26, 5289–5314.
- Baldocchi, D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, 82(11), 2415–2434.
- Barrett, D. J. (2002), Steady state turnover time of carbon in the Australian terrestrial biosphere, *Global Biogeochem. Cycles*, 16(4), 55-1–55-21, doi:10.1029/2002GB001860.
- Bartalis, Z., W. Wagner, V. Naeimi, S. Hasenauer, K. Scipal, H. Bonekamp, J. Figa, and C. Anderson (2007), Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), *Geophys. Res. Lett.*, 34, L20401, doi:10.1029/2007GL031088.
- Bondeau, A., D. W. Kicklighter, J. Kaduk, and the participants of the Potsdam NPP model intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Importance of vegetation structure on seasonal NPP estimates, *Global Change Biol.*, 5, 35–45.
- Braswell, B. H., W. J. Sacks, E. Linder, and D. S. Schimel (2005), Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, *Global Change Biol.*, 11(2), 335–355.
- Brovkin, V., T. Raddatz, C. H. Reick, M. Claussen, and V. Gayler (2009), Global biogeophysical interactions between forest and climate, *Geophys. Res. Lett.*, 36, L07405, doi:10.1029/2009GL037543.
- Clark, D., and P. Harris (2007), Joint UK Land Environment Simulator (JULES) version 2.0 user manual, *Tech. rep.*, NERC/Centre for Ecology & Hydrology, Wallingford, U. K.
- Conway, T. J., P. P. Tans, and L. S. Waterman (1994), Atmospheric CO₂ records from sites in the NOAA/CMDL air sampling network, in *Trends '93: A Compendium of Data on Global Change*, edited by Boden, T., D. Kaiser, R. Sepanski, and F. Stoss, p. 41–119, Carbon Dioxide Information Analysis Center, Oak Ridge, U.S.A.
- Cramer, W., D. W. Kicklighter, A. Bondeau, B. M. III, G. Churkina, B. Nemry, A. Ruimy, A. L. Schloss, and the participants of the Potsdam NPP model intercomparison (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, 5, 1–15.
- Dalmonech, D., S. Zaehle, G. Schrmann, V. Brovkin, C. Reick, and R. Schnur (2013), A systematic benchmark of carbon variability in an uncoupled and coupled land biosphere model JSBACH as case study, *J. Clim.*, in preparation.
- Dorigo, W., K. Scipal, R. Parinussa, Y. Liu, W. Wagner, R. de Jeu, and V. Naeimi (2010), Error characterisation of global active and

- passive microwave soil moisture data sets, *Hydrol. Earth Syst. Sci.*, *14*, 2605–2616.
- Drusch, M., E. Wood, and H. Gao (2005), Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture, *Geophys. Res. Lett.*, *32*, L15403, doi:10.1029/2005GL023623.
- Dutkiewicz, S., M. Follows, P. Heimbach, and J. Marshall (2006), Controls on ocean productivity and air-sea carbon flux: An adjoint model sensitivity study, *Geophys. Res. Lett.*, *33*, L02603, doi:10.1029/2005GL024987.
- Ehret, G., C. Kiemle, M. Wirth, A. Amediek, A. Fix, and S. Houweling (2008), Space-borne remote sensing of CO₂, CH₄, and N₂O by integrated path differential absorption lidar: A sensitivity analysis, *Appl. Phys. B.*, *90*, 593–608, doi:10.1007/s00340-007-2892-3.
- Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge University Press, Cambridge.
- Farquhar, G., S. v. v. Caemmerer, and J. Berry (1980), A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species, *Planta*, *149*(1), 78–90.
- 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*(10), 1597–1615, doi:10.1016/j.agrformet.2009.05.002.
- Fox, A. M., T. J. Hoar, D. J. Moore, S. J. Berukoff, and D. Schimel (2011), *A Model-Data fusion Approach to Integrate National Ecological Observatory Network Observations Into an Earth System Model*, pp. D520, AGU Fall Meeting Abstracts, Washington D.C.
- Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Software*, *24*(4), 437–474, doi:10.1145/293686.293695.
- GLOBALVIEW-CO₂ (2000), *Cooperative Atmospheric Data Integration Project—Carbon Dioxide, CD-ROM*, NOAA CMDL, Boulder, Colorado, [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].
- GLOBALVIEW-CO₂ (2004), *Cooperative Atmospheric Data Integration Project—Carbon Dioxide, CD-ROM*, NOAA CMDL, Boulder, Colorado, [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].
- Gobron, N., et al. (2006), Evaluation of fraction of absorbed photosynthetically active radiation products for different canopy radiation transfer regimes: Methodology and results using joint research center products derived from SeaWiFS against ground-based estimations, *J. Geophys. Res.*, *D13110*, doi:10.1029/2005JD006511.
- Gobron, N., B. Pinty, F. Melin, M. Taberner, M. M. Verstraete, M. Robustelli, and J.-L. Widowski (2007), Evaluation of the MERIS/ENVISAT FAPAR product, *Adv. Space Res.*, *39*, 105–115.
- Gobron, N., B. Pinty, O. Ausedat, M. Taberner, O. Faber, F. Mlin, T. Lavergne, M. Robustelli, and P. Snoeij (2008), Uncertainty estimates for the FAPAR operational products derived from MERIS—Impact of top-of-atmosphere radiance uncertainties and validation with field data, *Remote Sens. Environ.*, *112*, 1871–1883.
- Goll, D. S., V. Brovkin, B. R. Parida, C. H. Reick, J. Kattge, P. B. Reich, P. M. van Bodegom, and U. Niinemets (2012), Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling, *Biogeosciences*, *9*(9), 3547–3569, doi:10.5194/bg-9-3547-2012.
- Gregory, J. M., C. Jones, P. Cadule, and P. Friedlingstein (2009), Quantifying carbon cycle feedbacks, *J. Clim.*, *22*(19), 5232–5250.
- Griewank, A. (1989), *On automatic differentiation*, in *Mathematical Programming: Recent Developments and Applications*, edited by Iri, M., and K. Tanabe, p. 83–108, Kluwer Academic Publishers, Dordrecht.
- Hagemann, S., and T. Stacke (2013), Impact of the soil hydrology scheme on simulated soil moisture memory in a GCM, *Geophys. Res. Abstr.*, *15*, EGU2013–2784.
- Heimann, M., (1995), The global atmospheric tracer model TM2, *Technical Report No. 10*, Max-Planck-Institut für Meteorologie, Hamburg, Germany.
- Heimann, M., and S. Körner, (2003), The global atmospheric tracer model TM3, *Tech. Rep. 5*, Max-Planck-Institut für Biogeochemie, Jena, Germany.
- Heimann, M., et al. (1998), Evaluation of terrestrial carbon cycle models through simulations of the seasonal cycle of atmospheric CO₂: First results of a model intercomparison study, *Global Biogeochem. Cycles*, *12*, 1–24.
- Hollingsworth, A., D. Shaw, P. Lönnberg, L. Illari, K. Arpe, and A. Simmons (1986), Monitoring of observation and analysis quality by a data assimilation system, *Mon. Weather Rev.*, *114*(5), 861–879.
- Hooker-Stroud, A. (2008), Anthropogenic CO₂: Seasonal fossil fuel emissions in CCDAS, Master's thesis, University of Bristol, U. K.
- Jones, C., et al. (2013), 21st century compatible CO₂ emissions and airborne fraction simulated by CMIP5 Earth System models under four Representative Concentration Pathways, *J. Clim.*, *26*, 4398–4413.
- Junglaeus, J. H., et al. (2010), Climate and carbon-cycle variability over the last millennium, *Clim. Past Discuss.*, *6*(3), 1009–1044, doi:10.5194/cpd-6-1009-2010.
- Kaminski, T., and P. J. Rayner (2008), Assimilation and network design, in *Observing the Continental Scale Greenhouse Gas Balance of Europe*, Ecological Studies, chap. 3, edited by Dolman, H., A. Freibauer, and R. Valentini, p. 33–52, Springer-Verlag, New York, doi:10.1007/978-0-387-76570-9-3.
- Kaminski, T., M. Heimann, and R. Giering (1999), A coarse grid three dimensional global inverse model of the atmospheric transport, 1. Adjoint model and Jacobian matrix, *J. Geophys. Res.*, *104*(D15), 18,535–18,553.
- Kaminski, T., W. Knorr, P. Rayner, and M. Heimann (2002), Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochem. Cycles*, *16*(4), 14-1–14-16, doi:10.1029/2001GB001463.
- Kaminski, T., R. Giering, M. Scholze, P. Rayner, and W. Knorr (2003), An example of an automatic differentiation-based modelling system, in *Computational Science—ICCSA 2003, International Conference Montreal, Canada, May 2003, Proceedings, Part II, Lecture Notes in Computer Science*, vol. 2668, edited by Kumar, V., L. Gavrilova, C. J. K. Tan, and P. L'Ecuyer, p. 95–104, Springer, Berlin.
- Kaminski, T., M. Scholze, and S. Houweling (2010), Quantifying the benefit of A-SCOPE data for reducing uncertainties in terrestrial carbon fluxes in CCDAS, *Tellus B*, *62*(5), 784–796, doi:10.1111/j.1600-0889.2010.00483.x.
- Kaminski, T., W. Knorr, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P.-P. Mathieu (2012a), Consistent assimilation of MERIS FAPAR and atmospheric CO₂ into a terrestrial vegetation model and interactive mission benefit analysis, *Biogeosciences*, *9*(8), 3173–3184, doi:10.5194/bg-9-3173-2012.
- Kaminski, T., P. J. Rayner, M. Voßbeck, M. Scholze, and E. Koffi (2012b), Observing the continental-scale carbon balance: Assessment of sampling complementarity and redundancy in a terrestrial assimilation system by means of quantitative network design, *Atmos. Chem. Phys.*, *12*(16), 7867–7879, doi:10.5194/acp-12-7867-2012.
- Kato, T., W. Knorr, M. Scholze, E. Veenendaal, T. Kaminski, J. Kattge, and N. Gobron (2013), Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana, *Biogeosciences*, *10*(2), 789–802, doi:10.5194/bg-10-789-2013.
- Kelley, D. (2008), Wildfires as part of the global carbon cycle: Quantitative analysis using data assimilation, Master's thesis, U. K.
- Kelley, D. I., I. C. Prentice, S. P. Harrison, H. Wang, M. Simard, J. B. Fisher, and K. O. Willis (2013), A comprehensive benchmarking system for evaluating global vegetation models, *Biogeosciences*, *10*(5), 3313–3340, doi:10.5194/bg-10-3313-2013.
- Knorr, W. (1997), Satellitengestützte Fernerkundung und Modellierung des Globalen CO₂-Austauschs der Landvegetation: Eine Synthese, PhD thesis, Max-Planck-Institut für Meteorologie, Hamburg, Germany.
- Knorr, W. (2000), Annual and interannual CO₂ exchanges of the terrestrial biosphere: Process-based simulations and uncertainties, *Glob. Ecol. Biogeogr.*, *9*(3), 225–252.
- Knorr, W., and M. Heimann (1995), Impact of drought stress and other factors on seasonal land biosphere CO₂ exchange studied through an atmospheric tracer transport model, *Tellus Ser. B*, *47*(4), 471–489.
- Knorr, W., and J. Kattge (2005), Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using Monte Carlo sampling, *Global Change Biol.*, *11*, 1333–1351.
- Knorr, W., T. Kaminski, M. Scholze, N. Gobron, B. Pinty, R. Giering, and P.-P. Mathieu (2010), Carbon cycle data assimilation with a generic phenology model, *J. Geophys. Res.*, *115*, G04017, doi:10.1029/2009JG001119.
- Koffi, E., P. Rayner, M. Scholze, F. Chevallier, and T. Kaminski (2012a), Quantifying the constraint of biospheric process parameters by CO₂ concentration and flux measurement networks through a carbon cycle data assimilation system, *Atmos. Chem. Phys. Discuss.*, *12*(9), 24,131–24,172, doi:10.5194/acpd-12-24131-2012.
- Koffi, E., P. J. Rayner, M. Scholze, and C. Beer (2012b), Atmospheric constraints on gross primary productivity and net flux: Results from a carbon-cycle data assimilation system, *Global Biogeochem. Cycles*, *26*, GB1024, doi:10.1029/2010GB003900.
- Krinner, G., N. Viovy, N. de Noblet-Ducoudr, J. Oge, 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.
- Kuppel, S., F. Chevallier, and P. Peylin (2012a), Quantifying the model structural error in carbon cycle data assimilation systems, *Geoscientific Model Dev. Discuss.*, *5*(3), 2259–2288, doi:10.5194/gmdd-5-2259-2012.

- Kuppel, S., P. Peylin, F. Chevallier, C. Bacour, F. Maignan, and A. D. Richardson (2012b), Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosciences*, 9(10), 3757–3776, doi:10.5194/bg-9-3757-2012.
- Lasslop, G., M. Reichstein, J. Kattge, and D. Papale (2008), Influences of observation errors in eddy flux data on inverse model parameter estimation, *Biogeosciences*, 5(5), 1311–1324, doi:10.5194/bg-5-1311-2008.
- Lawrence, D. M., et al. (2011), Parameterization improvements and functional and structural advances in version 4 of the community land model, *J. Adv. Model. Earth Syst.*, 3(3), M03001, doi:10.1029/2011MS000045.
- Lewis, P. E., J. Gomez-Dans, T. Kaminski, J. Settle, T. Quaife, N. Gobron, J. Styles, and M. Berger (2012), An Earth Observation Land Data Assimilation System (EO-LDAS), *Remote Sens. Environ.*, 120, 219–235.
- Lorenz, E. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20(2), 130–141.
- Luke, C. M. (2011), Modelling aspects of land-atmosphere interaction: Thermal instability in peatland soils and land parameter estimation through data assimilation, PhD thesis, University of Exeter, U. K.
- Mauder, M., M. Cuntz, C. Drüe, A. Graf, C. Rebmann, H. P. Schmid, M. Schmidt, and R. Steinbrecher (2013), A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, *Agric. For. Meteorol.*, 169, 122–135.
- Medvigy, D., S. Wofsy, J. Munger, D. Hollinger, and P. Moorcroft (2009), Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem demography model version 2, *J. Geophys. Res.*, 114, G01002, doi:10.1029/2008JG000812.
- Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087–1092.
- Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for bayesian atmospheric trace gas surface flux inversions, *J. Geophys. Res.*, 110, D24107, doi:10.1029/2005JD005970.
- Montzka, C., V. R. N. Pauwels, H.-J. H. Franssen, X. Han, and H. Vereecken (2012), Multivariate and multiscale data assimilation in terrestrial systems: A review, *Sensors*, 12(12), 16,291–16,333, doi:10.3390/s121216291.
- Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner (2009), An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations, *IEEE Trans. Geosci. Remote Sens.*, 47(7), 1999–2013.
- Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res.*, 113, F01002, doi:10.1029/2007JF000769.
- Peters, G. P., R. M. Andrew, T. Boden, J. G. Canadell, P. Ciais, C. Le Quére, G. Marland, M. R. Raupach, and C. Wilson (2012), The challenge to keep global warming below 2°C, *Nat. Clim. Change*, 3, 4–6.
- Pinty, B., M. Clerici, I. Andreadakis, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J. L. Widowski (2011), Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 2. Fractions of transmitted and absorbed fluxes in the vegetation and soil layers, *J. Geophys. Res.*, 116, D09106, doi:10.1029/2010JD015373.
- Raddatz, T., C. Reick, W. Knorr, J. Kattge, E. Roeckner, R. Schnur, K.-G. Schnitzler, P. Wetzel, and J. Jungclauss (2007), Will the tropical land biosphere dominate the climate carbon cycle feedback during the twenty-first century?, *Clim. Dyn.*, 29(6), 565–574, doi:10.1007/s00382-007-0247-8.
- Rayner, P., W. Knorr, M. Scholze, R. Giering, T. Kaminski, M. Heimann, and C. Le Quere (2001), Inferring terrestrial biosphere carbon fluxes from combined inversions of atmospheric transport and process-based terrestrial ecosystem models, in *Proceedings of 6th Carbon Dioxide Conference at Sendai*, p. 1015–1017.
- Rayner, P., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005a), Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS), *GB2026*, 19, doi:10.1029/2004GB002254.
- Rayner, P., M. Scholze, P. Friedlingstein, J.-L. Dufresne, T. Kaminski, W. Knorr, R. Giering, and H. Widmann (2005b), The Fate of Terrestrial Carbon Under Climate Change: Results from a CCDAS, Poster Presentation at 7th Carbon Dioxide Conference at Broomfield, Colorado.
- Rayner, P., E. Koffi, M. Scholze, T. Kaminski, and J. Dufresne (2011), Constraining predictions of the carbon cycle using data, *Philos. Trans. R. Soc. London, Ser. A*, 369(1943), 1955–1966, doi:10.1098/rsta.2010.0378.
- Rayner, P. J. (2010), The current state of carbon-cycle data assimilation, *Curr. Opin. Environ. Sustainability*, 2(4), 289–296.
- Rayner, P. J., I. G. Enting, and C. M. Trudinger (1996), Optimizing the CO₂ observing network for constraining sources and sinks, *Tellus*, 48B, 433–444.
- Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler (2013), The representation of natural and anthropogenic land cover change in MPI-ESM, *J. Adv. Model. Earth Syst.*, 5, doi:10.1002/jame.20022.
- Reuter, M., et al. (2011), Retrieval of atmospheric CO₂ with enhanced accuracy and precision from SCLAMACHY: Validation with FTS measurements and comparison with model results, *J. Geophys. Res.*, 116, D04301, doi:10.1029/2010JD015047.
- Richardson, A. D., et al. (2006), A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agric. For. Meteorol.*, 136(1), 1–18.
- Richardson, A. D., M. Aubinet, A. G. Barr, D. Y. Hollinger, A. Ibrom, G. Lasslop, and M. Reichstein (2012), Uncertainty quantification, in *Eddy Covariance*, chap. 7, edited by M. Aubinet, T. Vesala, and D. Papale, pp. 173–209, Springer, New York.
- Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003), Time-dependent atmospheric CO₂ inversions based on interannually varying tracer transport, *Tellus Ser. B-Chem. Phys. Meteorol.*, 55(2), 488–497.
- Roeckner, E., et al. (2003), The atmospheric general circulation model ECHAM5: Part 1: Model description, *Report No. 349*, Max-Planck-Institut für Meteorologie, Hamburg, Germany.
- Roesch, A., M. Wild, H. Gilgen, and A. Ohmura (2001), A new snow cover fraction parameterization for the ECHAM4 GCM, *Clim. Dyn.*, 17(12), 933–946, doi:10.1007/s003820100153.
- Ruimy, A., G. Dedieu, and B. Saugier (1994), Methodology for the estimation of terrestrial net primary production from remotely sensed data, *J. Geophys. Res.*, 99, 5263–5283.
- Scholze, M. (2003), Model studies on the response of the terrestrial carbon cycle on climate change and variability, Examensarbeit, Max-Planck-Institut für Meteorologie, Hamburg, Germany.
- Scholze, M., T. Kaminski, P. Rayner, W. Knorr, and R. Giering (2007), Propagating uncertainty through prognostic CCDAS simulations, *J. Geophys. Res.*, 112, D17305, doi:10.1029/2007JD008642.
- Scholze, M., et al. (2013), Simultaneous optimisation of process parameters in a terrestrial and marine carbon cycle model using atmospheric CO₂ concentrations, in *Proceedings of 9th Carbon Dioxide Conference in Beijing*.
- Schürmann, G., et al. (2013), Assimilation of NEE and CO₂-concentrations into the land-surface scheme of the MPI Earth System Model, Poster Presentation at EGU, Vienna.
- Scipal, K., M. Drusch, and W. Wagner (2008a), Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system, *Adv. Water Resour.*, 31(8), 1101–1112.
- Scipal, K., T. Holmes, R. De Jeu, V. Naeimi, and W. Wagner (2008b), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, 35, L24403, doi:10.1029/2008GL035599.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia.
- 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.
- Veenendaal, E. M., O. Kolle, and J. Lloyd (2004), Seasonal variation in energy fluxes and carbon dioxide exchange for a broad-leaved semi-arid savanna (Mopane woodland) in southern Africa, *Global Change Biol.*, 10(3), 318–328.
- Wang, Y. P., R. Leuning, H. Cleugh, and P. A. Coppin (2001), Parameter estimation in surface exchange models using non-linear inversion: How many parameters can we estimate and which measurements are most useful?, *Glob. Change Biol.*, 7, 495–510.
- Weedon, G., et al. (2011), Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *J. Hydrometeorol.*, 12(5), 823–848.
- 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(1), 89–105.
- Wilson, M. F., and A. Henderson-Sellers (1985), A global archive of land cover and soils data for use in general-circulation climate models, *J. Climatol.*, 5(2), 119–143.
- Ziehn, T., M. Scholze, and W. Knorr (2009), Regionalization of the key carbon storage parameter within the Carbon Cycle Data Assimilation System (CCDAS), in *Proceedings of 8th Carbon Dioxide Conference at Jena*.
- Ziehn, T., W. Knorr, and M. Scholze (2011), Investigating spatial differentiation of model parameters in a carbon cycle data assimilation system, *Global Biogeochemical Cycles*, 25(2), GB2021.
- Ziehn, T., M. Scholze, and W. Knorr (2012), On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models, *Global Biogeochem. Cycles*, 26, GB3025, doi:10.1029/2011GB004185.