Carbon and fAPAR assimilation within CCDAS

W. Knorr[1], N. Gobron[2], M. Scholze[3], P. Rayner[4], T. Kaminski[5], R. Giering[5], H. Widmann[1], and J. Kattge[1]

[1]*Max Planck Institut für Biogeochemie, Jena, Germany*
[2]*Global Vegetation Monitorin Unit, IES, EC Joint Research Centre, Ispra*
[3] *QUEST -- Dep.\ of Earth Sciences, University of Bristol, Bristol, UK*
[4] *Laboratoire des Sciences du Climat et de l'Environnement, France*
[5] *FastOpt, Hamburg, Germany*

Abstract

We present the Carbon Cycle Data Assimilation System (CCDAS), which is built around the Biosphere Energy Transfer HYdrology Scheme (BETHY), coupled to the atmospheric transport model TM2. In its current form, the system uses a two-step assimilation procedure to estimate 57 model parameters from both satellite observed vegetation activity and atmospheric carbon dioxide samples. Observational uncertainties are mapped back on parameter uncertainties. These parameter uncertainties are then projected forward to important diagnostics such as regional carbon budgets. CCDAS includes BETHY's adjoint, Jacobian, and Hessian codes, which are provided by automatic differentiation (AD) of the BETHY code. This automated procedure allows quick updates of CCDAS after modifications of BETHY. We also report on the preparation of fraction of Absorbed Photosynthetically Active Radiation (fAPAR) data provided by SeaWiFS for use in CCDAS. Analysis of this fAPAR data suggests a time lag against precipitation of one to several months. A reaction of the terrestrial biosphere at such timescales carries promises for medium-range to seasonal prediction.

1. Introduction

The Third Assessment Report of the IPCC [Prentice et al., 2001] has shown that, while the terrestrial biosphere appears to have been a significant sink for atmospheric CO_2 during the past, its further ability to take up atmospheric CO_2 could be reduced by the effect of climate change. The result can be a positive feedback effect with accelerated greenhouse warming [Cox et al, 2000]. The terrestrial biosphere models used in such studies, however, carry large uncertainties. Those uncertainties need to be (a) assessed in a systematic fashion, and (b) reduced through the provision of comprehensive data.

A major source of this uncertainty in predictions is the formulation of the underlying processes, including the values of process parameters [Friedlingstein et al., 2003]. This puts the following requirements on the design of a terrestrial data assimilation procedure for the climate time scale:

i) The system must take the uncertainty in process parameters into account by specifying them as control variables.

ii) The system must deliver a posteriori uncertainties on parameters to be used for quantification of the uncertainty range in predictions.

iii) The system must be quickly adaptable to changes in the representation of the underlying processes.

Requirement i) would be unusual in an NWP context, where, owing to short integration periods, only initial conditions are specified as control variables. Requirement iii) is best fulfilled by a system that can be updated in a highly automated process chain. This note presents a prototype of such a system, the Carbon Cycle Data Assimilation System (CCDAS, see http://CCDAS.org).

Section 2 gives an overview of CCDAS, and section 3 addresses details of its implementation. Section 4 sketches two examples of updates that are underway, the inclusion of an additional stream of observations and the upgrade of the satellite observations.

2. CCDAS

Figure 1 gives an overview on CCDAS. The box on the left hand side shows the stand level assimilation step which is not yet integrated into the system (see section 4). In its current form, CCDAS carries out the two separate assimilation steps depicted in the middle and right boxes. The first step (middle box) uses a variational approach to assimilate the observed fraction of Absorbed Photosynthetically Active Radiation (fAPAR) provided by AVHRR into the Biosphere Energy Transfer HYdrology Scheme (BETHY, Knorr [1997, 2000]). This step operates at each grid point separately with three control variables representing limitations by temperature, water, and resources not explicitly included in the model (e.g. nitrogen and land use). For details of the assimilation procedure see Knorr [1997] and Knorr and Schulz [2001].

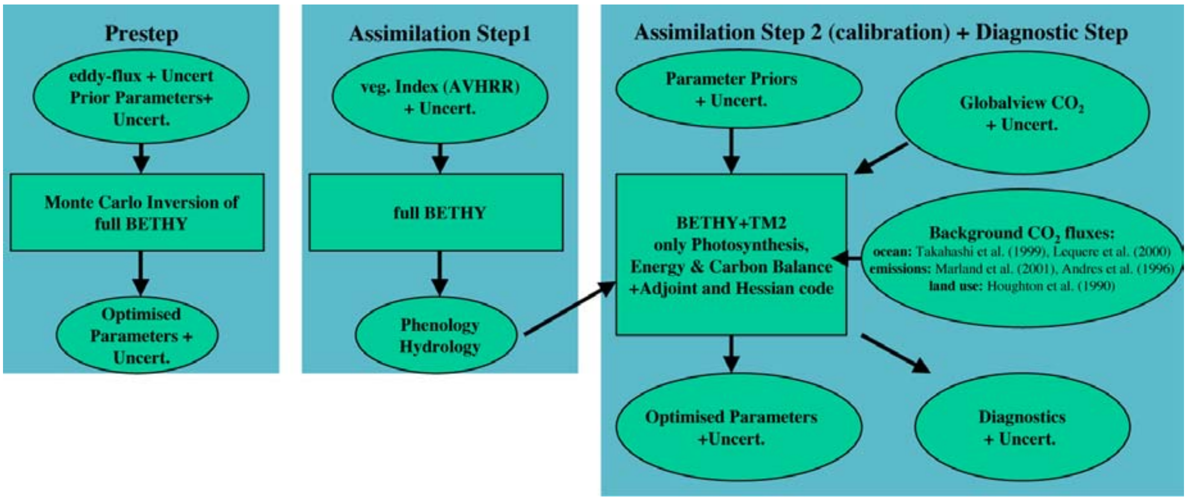

Figure 1: Schematic representation of CCDAS.

The second assimilation step uses a modified BETHY version, restricted to the simulation of photosynthesis, carbon and energy balance referred to as Carbon-BETHY (see also Scholze [2003] and Rayner et al. [2004]). At each grid point, time series of Leaf Area Index (LAI) quantifying phenology and plant available soil moisture quantifying hydrology are provided by the first assimilation step. To assimilate atmospheric carbon dioxide from the GLOBALVIEW flask sampling network [GLOBALVIEW- CO_2, 2001] Carbon-BETHY is coupled to the atmospheric transport model TM2 [Heimann, 1995]. TM2 is represented by its Jacobian matrix [Kaminski et al., 1999] (taking the role of an observational operator). This second assimilation step applies the adjoint method and estimates 56 process-parameters plus an initial condition, all specified as control variables. The calibration mode of CCDAS is completed by an evaluation of the Hessian of the cost function at its optimum. The inverse Hessian approximates the control variable's posterior uncertainties reflecting the sum of observational and model uncertainty.

The CCDAS diagnostic/prognostic mode consists of an integration with optimised control variables. To propagate the uncertainties in the control variables forward to estimate uncertainties in diagnosed/prognosed target quantities, the model's Jacobian matrix is evaluated. So far, CCDAS has only been used for diagnostic computations. As an example, figure 2 shows regional budgets and error bars from an older CCDAS version [Scholze et al., 2002]. For more CCDAS results we refer to Scholze [2003] and Rayner et al. [2004]. For details on the methodology consult Kaminski et al. [2002, 2003] and Rayner et al. [2004]. Details on the CCDAS consortium and news on CCDAS are available via http://CCDAS.org.

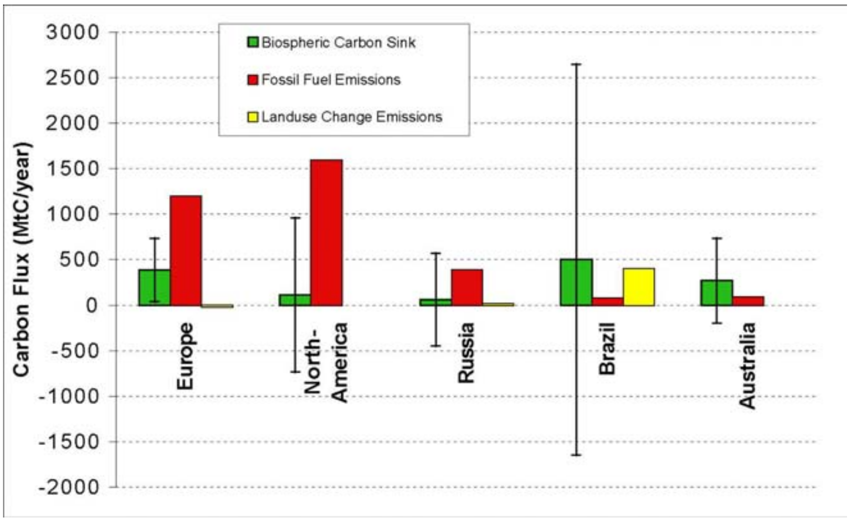

Figure 2: CCDAS estimates of regional budgets and their uncertainty.

3. Implementation via automatic differentiation

Section 1 has motivated the need for frequent CCDAS updates after each modification/improvement of the formulation of the underlying processes. To facilitate this update process, all derivative code (adjoint, Hessian, and Jacobian code) is generated from the code of the underlying model (roughly 5000 lines, excluding comments, of Fortran 90) by automatic differentiation (AD, Griewank [2000]). CCDAS uses the AD tool Transformation of Algorithms in Fortran (TAF, http://FastOpt.com, Giering and Kaminski [1998]; Giering et al. [2005]). The system contains a verification environment that allows to compare the generated adjoint and Jacobian codes against finite differences and the generated Hessian codes against finite differences of the adjoint.

The classical challenge for efficient adjoint coding is to provide required values, i.e. values from the model integration, in an efficient way to the adjoint code. (Required values are often referred to as trajectory, although they usually encompass many more values than the pure state.) As CCDAS simulates periods of a few decades at time steps between one hour and one month, depending on the respective process, it is most efficient to apply a so-called checkpointing strategy [Griewank, 1992]: During an initial integration all necessary restart information is saved at monthly intervals on disk. Each monthly interval in the adjoint integration is then preceded by a forward integration over the same interval, which is initialised with the restart information on disk. The forward integration stores all required values for the current interval that are not recomputed by the adjoint code in memory. This checkpointing procedure results in a considerable reduction of disk/memory space for storing required values at the cost of one additional model integration. The adjoint is still as fast as about 3.4 model integrations. For integration periods that are short enough to avoid the checkpointing procedure the adjoint would take the time of 2.4 model integrations.

Because of memory restrictions (2GB), the Hessian evaluation is split into groups of 12 columns each. The evaluation of 12 columns takes the time of about 50 model integrations. Often we are only interested in a few diagnostic quantities, and thus use a “vector valued adjoint” to evaluate the Jacobian. Performance tests show that computing an additional row of the Jacobian costs about 0.25 model integrations. For details on the derivative code within CCDAS and its performance we refer to Kaminski et al. [2003].

Initially, a number of modifications to the original BETHY code were necessary, to render the model TAF compliant. From this point on, BETHY has been developed further within CCDAS, allowing immediate update of the derivative code by TAF. This yields, at each development step, both sensitivity information and systematic comparison with observational data meaning that CCDAS is supporting model development. The data assimilation activities, in turn, benefit from using the current model version. AD is essential for this integrated modelling concept, as it avoids the usual delay from a release of a new model version to its use for applications that rely on derivatives such as assimilation or adjoint sensitivity studies.

4. Next steps

The current CCDAS prototype is being extended and improved in several ways. In the following we sketch two examples of such extensions, the inclusion of an additional stream of data and the upgrade of the satellite data.

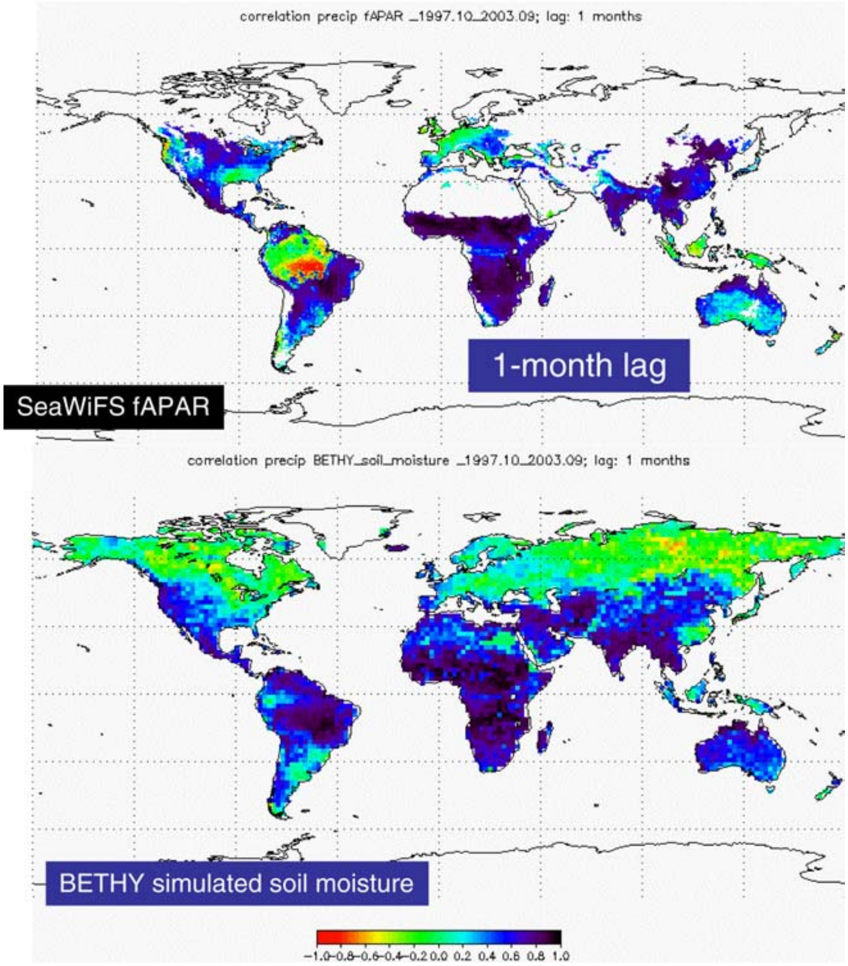

Figure 3: Correlation between precipitation and SeaWiFS fAPAR (upper panel) or BETHY simulated soil moisture (lower panel) at 1 month lag (positive means precipitation is leading).

In the current setup, prior information on process-parameters are a combination of values from the published literature and expert knowledge. A more rigorous approach of Knorr and Kattge [2004] (see figure 1) employs Monte Carlo sampling to probe the probability distribution of BETHY's process parameters, taking stand-level information from eddy-flux measurements into account. First results [Knorr and Kattge, 2004] suggest that stand-level data provide a constraint that is complementary to atmospheric carbon dioxide. Hence, it is planned to extend CCDAS by this assimilation step. A first link will be established by using the parameter estimates from the stand-level assimilation as prior information for the subsequent steps. Eventually it is planned to assimilate the stand-level observations jointly with the atmospheric carbon dioxide observations in the adjoint-based assimilation procedure.

Another future modification is to replace fAPAR observations from AVHRR by SeaWiFS data, which are available from September 1997 onwards. First analyses of SeaWiFS derived fAPAR data against station derived gridded precipitation [Chen et al., 2002] show large regions of the tropics and subtropics where vegetation activity is strongly correlated with water input from precipitation (see upper panel of figure 3). As shown in figure 4, the dominant lag is one month, a typical response time for the soil moisture/vegetation system. A very similar pattern and response time is found (see lower panel of figure 3) for simulations of soil moisture with BETHY, here used in its independent form without data assimilation. An exception are areas that are either not water limited (high latitudes), or have tropical evergreen vegetation.

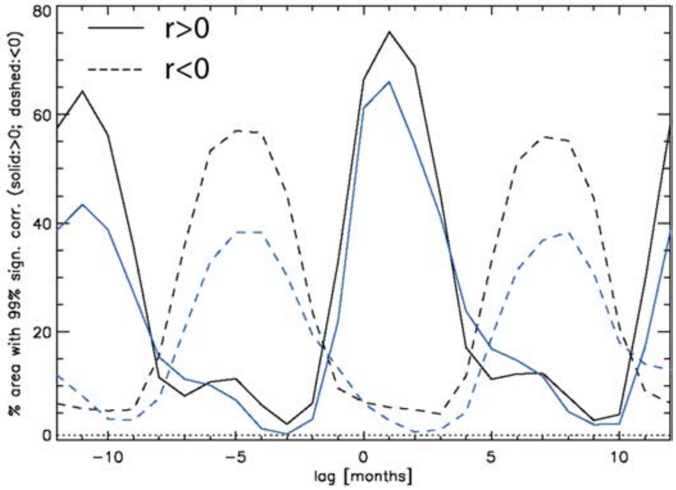

Figure 4: Percentage of valid land pixels with 99% significant correlation as a function of the lag.

Figure 5 shows both prior and posterior values of the soil-water content from a first run of CCDAS assimilation step 2 (middle box in figure 1) with fAPAR from SeaWiFS [Gobron et al., 2005]. These first results suggest that BETHY can simulate the dominant dynamics and time scales of the terrestrial water and vegetation processes realistically and that CCDAS promises to deliver important information on the land surface state. Such information could then be used in coupled atmosphere-land vegetation models, with the potential of improving longer-range forecasts.

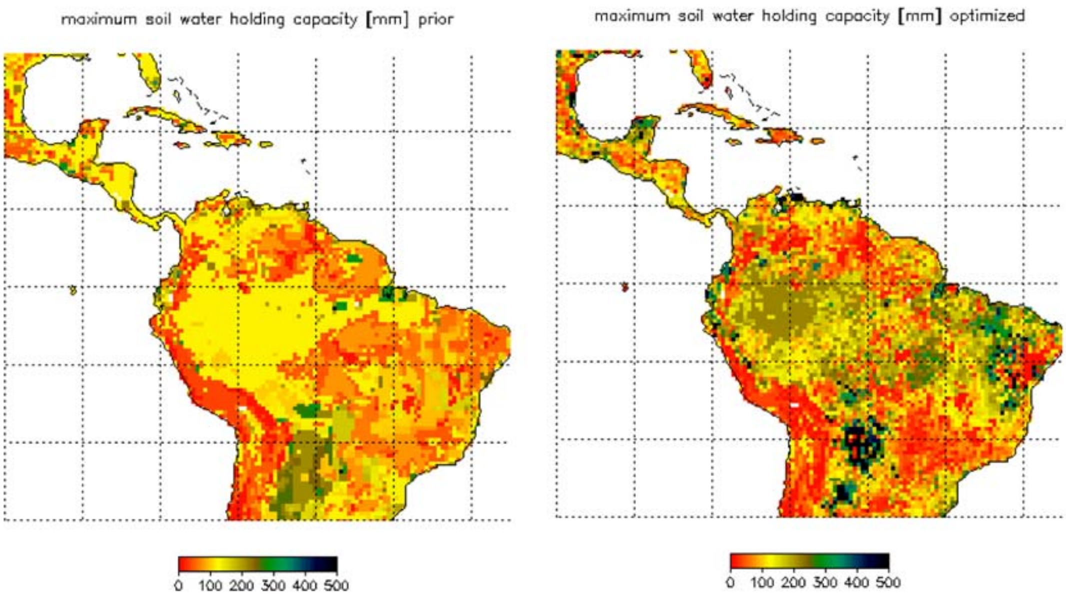

Figure 5: Soil water content prior (left panel) and posterior (right panel) values.

5. Acknowledgements

We thank Martin Heimann for his support of our work. Part of this work has been carried out in the project CAMELS, supported by the EU under contract no. EVK2-CT-2002-00151 within the 5th Framework Programme for Research and Technological Development.

References

Chen, M., P. Xie, J. Janowiak, and P. Arkin, Global land precipitation: a 50-yr monthly analysis based on gauge observations, *J. Hydrometeorol*., **3**, 249–266, 2002.

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

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

Giering, R., and T. Kaminski, Recipes for Adjoint Code Construction, *ACM Trans. Math. Software*, **24**, 437–474, 1998.

Giering, R., T. Kaminski, and T. Slawig, Applying TAF to a Navier-Stokes solver that simulates an Euler flow around an airfoil, To appear in *Future Generation Computer Systems*, 2005.

GLOBALVIEW- CO_2, *Cooperative Atmospheric Data Integration Project - Carbon Dioxide*, CD-ROM, NOAA CMDL, Boulder, Colorado, 2001, [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].

Gobron, N., B. Pinty, F. Mélin, M. Taberner, M. M. Verstraete, A. Belward, T. Lavergne, and J.-L. Widlowski, The state of vegetation in Europe following the 2003 drought, to appear in *Int. J. Remote Sensing*, 2005.

Griewank, A., Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation, *Optimization Methods and Software*, **1**, 35–54, 1992.

Griewank, A., Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, no. 19 in *Frontiers in Appl. Math*., SIAM, Philadelphia, 2000.

Heimann, M., The global atmospheric tracer model TM2, *Technical Report No. 10*, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 1995.

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

Kaminski, T., W. Knorr, P. Rayner, and M. Heimann, Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochemical Cycles*, **16**, 14–1–14–16, 2002.

Kaminski, T., R. Giering, M. Scholze, P. Rayner, and W. Knorr, An example of an automatic differentiation-based modelling system, in *Computational Science – ICCSA 2003, International Conference Montreal, Canada, May 2003, Proceedings, Part II*, edited by V. Kumar, L. Gavrilova, C. J. K. Tan, and P. L'Ecuyer, vol. 2668 of Lecture Notes in Computer Science, pp. 95–104, Springer, Berlin, 2003.

Knorr, W., Satellitengest¨utzte Fernerkundung und Modellierung des Globalen CO_2 -Austauschs der Landvegetation: Eine Synthese, Ph.D. thesis, Max-Planck-Institut f¨ur Meteorologie, Hamburg, Germany, 1997.

Knorr,W., Annual and interannual CO_2 exchanges of the terrestrial biosphere: process based simulations and uncertainties, *Glob. Ecol. and Biogeogr.*, **9**, 225–252, 2000.

Knorr, W., and J. Kattge, Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using Monte Carlo sampling, submitted to *Global Change Biology*, 2004.

Knorr, W., and J.-P. Schulz, Using satellite data assimilation to infer global soil moisture and vegetation feedback to climate, in *Remote Sensing and Climate Modeling: Synergies and Limitations*, edited by M. Beniston and M. Verstraete, Advances in Global Change Research, pp. 273–306, Kluwer Academic Publishers, Dordrecht, the Netherlands, 2001.

Prentice, I. C., G. D. Farquhar, M. J. R. Fasham, M. L. Goulden, M. Heimann, V. J. Jaramillo, H. S. Kheshgi, and C. Le Quéré, The carbon cycle and atmospheric carbon dioxide, chap. 3, pp. 183–237, Published for the IPCC by CUP, Cambridge, UK, 2001.

Rayner, P., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann, Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS), Accepted for publication in *Global Biogeochemical Cycles*, 2004.

Scholze, M., Model studies on the response of the terrestrial carbon cycle on climate change and variability, Examensarbeit, Max-Planck-Institut f¨ur Meteorologie, Hamburg, Germany, 2003.

Scholze, M., P. Rayner, W. Knorr, T. Kaminski, and R. Giering, A prototype Carbon Cycle Data Assimilation System (CCDAS): Inferring interannual variations of vegetation-atmosphere CO_2 fluxes. Abstract CG62A-05, *Eos Trans. AGU*, 83, 2002.