REMOTE SENSING INPUT FOR REGIONAL TO GLOBAL CO_2 FLUX MODELLING

Wolfgang Knorr[1], Thomas Kaminski[2], Marko Scholze[1], Nadine Gobron[3], Bernard Pinty[3], and Ralf Giering[2]

[1]*University of Bristol, Wills Memorial Building, BS8 1RJ Bristol, United Kingdom*
[2]*FastOpt, Schanzenstr. 36, 20357, Germany*
[3]*European Commission - Joint Research Centre, via E. Fermi, 2749, 21027 Ispra, Italy*

ABSTRACT

The current and future strength of the terrestrial carbon sink has a crucial influence on the expected climate warming on Earth. Usually, Earth Observation (EO) by its very nature focusses on diagnosing the current state of the planet. However, it is possible to use EO products in data assimilation systems to improve not only the diagnosis of the current state, but also the accuracy of future predictions.

This contribution reports from an on-going Eurpean Space Agency (ESA) funded study in which the MERIS FAPAR product is assimilated into a terrestrial biosphere model within the global Carbon Cycle Data Assimilation System (CCDAS). Results are presented from a range of selected sites spanning the major biomes of the globe, and show how the inclusion of MERIS land products results in improved accuracy of the site carbon flux estimates. They also show the uncertainty in the predicted carbon sink of those sites for selected climate scenarios until 2039.

Key words: Global carbon cycle; MERIS; FAPAR; evaluation of sensor concepts, mission design, OSSE.

1. INTRODUCTION

During the coming decades, the continuing rise of greenhouse gases (GHG) in the atmosphere is expected to lead to more and more pronounced climate warming. Of those GHGs, carbon dioxide is the most important one as its lifetime in the atmosphere is almost infinite. If nothing is done to drastically curb GHG emissions, severe disruptions of current weather patterns, ice shelves, river flows and coastal ecosystems will become a possibility [1].

As the European Union has committed itself to avoid dangerous levels of climate warming and is starting to set itself stringent reduction targets beyond the Kyoto Protocol, European researchers are called upon to better investigate the role of the global carbon cycle. This is because terrestrial vegetation emits and absorbs carbon at five to ten times the rate of the anthropogenic CO_2 emissions. As a result, changes in the activity of terrestrial plants can have a large impact on the predicted levels of CO_2 for a given emissions scenario. The main tools for investigating the relationships between emissions, climate change and the carbon cycle are now comprehensive Earth System Models (ESMs).

The scientific information required to quantify the relationship between CO_2 emissions and atmospheric concentrations is presently insufficient due to major uncertainties in the terrestrial climate-carbon cycle feedback that are still inherent the current generation of ESMs [2, 3, 4]. This calls for quantitative methods that are able to reduce the large margins of uncertainty for both the strength of the current carbon sink in the terrestrial biosphere and for its evolution into the future. For this reason, CCDAS have been developed that are able to assimilate additional observational information into ESMs. Such assimilation systems will allow for improving the prediction accuracy of the carbon sink strength of the terrestrial biosphere.

The sources of information assimilated into ESMs should ideally be of global, uniform coverage. An example of an ideal product for carbon cycle data assimilation at the global scale is the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) product from ESA's MERIS sensor on-board Envisat [5]. It documents the presence of vigorously photosynthesising vegetation at the global scale using a purely radiometric definition [6]. The Biosphere Energy-Transfer-HYdrology (BETHY) model [7] uses a two-stream radiative transfer model to compute light absorption within the canopy and explicitly simulates FAPAR. Thus, both the satellite product and the biosphere model can reproduce the same physical quantity FAPAR, which opens the possibility for assimilating FAPAR into BETHY using the techniques developed in Section 2.

A method that is fully consistent with both the philosophy of inverse modelling and the general approach used when doing climate predictions is implemented in the Carbon Cycle Data Assimilation System (see http://www.CCDAS.org [8, 9, 10]) As shown in Figure 1, the system employs techniques from variational data assimilation: here, the current state of the system is estimated with the best possible accuracy using the

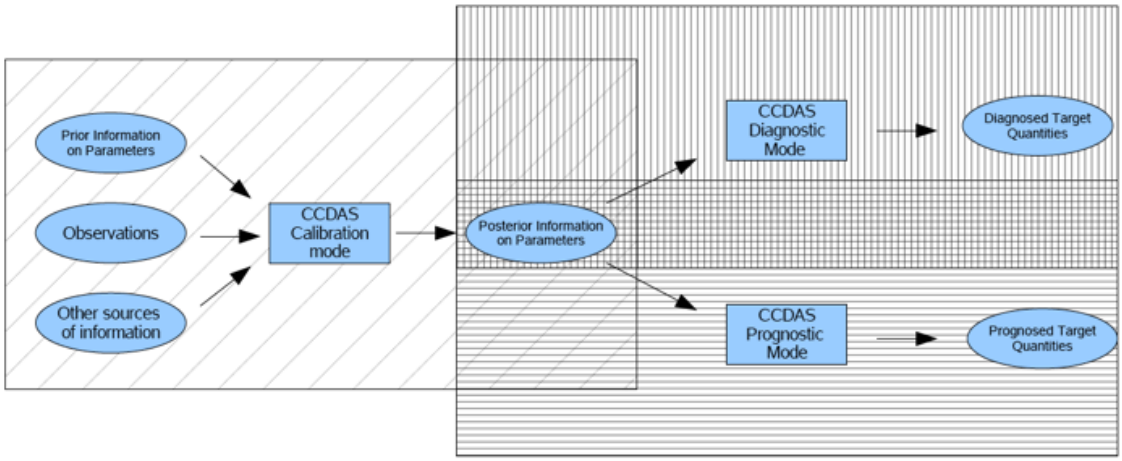

Figure 1. CCDAS two-step procedure: Inverse step followed by diagnostic or prognostic step.

best available observational constraints at the appropriate scale of the problem. This first step is used to estimate both process parameters and initial conditions for the subsequent prediction. For example, state variable estimation is important for the slowly evolving carbon pools, whose turnover time (parameter) and concentration at the beginning of the forecast (intitial condition) must be estimated with confidence [11]. The prediction or prognostic step then uses not only standard modelling techniques to arrive at a forecast but also techniques of uncertainty propagation to estimate uncertainty ranges for the prediction [10]. This latter is a significant advance over current modelling techniques.

A crucial aspect of CCDAS is the use of process parameters to bridge gaps in spatial and temporal scales. The use of generic, process-oriented models of carbon fluxes allows both the extrapolation in space (from sites to the globe) and across time (from the current observations to future predictions) [12]. The assumption of scale independence and universality of process parameters is at the core of mechanistic presentations of Earth system processes.

The present paper reports on a study that includes MERIS observations at site and global scale into the set of observations that are assimilated by CCDAS. Furthermore, the study demonstrates the applicaton of quantitative network design methods, also called Observing System Simulation Experiment (OSSE), to evaluate potential sensor concepts in support of the design of future EO missions. Section 2 introduces the formalism for assimilation and network design as employed in CCDAS. Section 3 shows results from simultaneous assimilation of different observables at six sites and Section 4 demonstrates potential sensor evaluation. Section 5 presents conclusions and an outlook.

2. CCDAS METHODOLOGY

This section gives a brief introduction into the formalism underlying Figure 1 and shows the fundamental equations. The presentation of the methodology closely follows previously published work [13].

It is convenient to formulate the inverse problem of quantifying the state of information on a specific physical quantity by a probability density function (PDF): the prior information is quantified by a PDF in the space of control variables (here: process parameters of BETHY and the initial concentration), and the observational information by a PDF in the space of observations. This probabilistic framework is described in detail by [14]; an introduction of the same framework with an exhaustive overview on applications to biogeochemistry is given by [15].

For the case of a linear model where the input to the inverse problem can be characterised by Gaussian PDFs and the model error is Gaussian as well, the posterior information can be shown to be characterised by a Gaussian PDF (see [14]). The mean of this PDF is given by:

$$\mathbf{x} = \mathbf{x_0} + [\mathbf{M}^T \mathbf{C}(d)^{-1} \mathbf{M} + \mathbf{C}(x_0)^{-1}]^{-1} \mathbf{M}^T \mathbf{C}(d)^{-1} (\mathbf{d} - \mathbf{M}\mathbf{x_0}) \quad , \tag{1}$$

and the covariance of its uncertainty is given by:

$$\mathbf{C}(x)^{-1} = \mathbf{M}^T \mathbf{C}(d)^{-1} \mathbf{M} + \mathbf{C}(x_0)^{-1} \quad , \tag{2}$$

where $\mathbf{M}$ denotes the Jacobian matrix of the model (linking control variables to observations), x_0 and $\mathbf{C}(x_0)$ the mean and covariance of the prior information's PDF. $\mathbf{d}$ and $\mathbf{C}(d)$ denote the mean and the covariance of uncertainty of the observations.

In the inversion procedure, the corresponding PDF has to reflect errors in both the observational process and our ability to correctly model the observations. We achieve this via

$$\mathbf{C}(d) = \mathbf{C}(d_{\text{obs}}) + \mathbf{C}(d_{\text{mod}}) \tag{3}$$

and by subtracting the mean model and observational errors from $\mathbf{Mx_0}$ and d, respectively.

We also note that Eq. (2) can be reformulated to:

$$\mathbf{C}(x) = \mathbf{C}(x_0) - \mathbf{C}(x_0)\mathbf{M}^T(\mathbf{C}(d) + \mathbf{M}\mathbf{C}(x_0)\mathbf{M}^T)^{-1}\mathbf{M}^T\mathbf{C}(x_0) \quad . \tag{4}$$

One can easily verify that $\mathbf{x}$ also minimises the cost function (the exponent of the Gaussian posterior PDF) Eq. (1)

$$\begin{aligned} J(\tilde{\mathbf{x}}) = \quad & \tfrac{1}{2} \quad [(\mathbf{M}\tilde{\mathbf{x}} - \mathbf{d})^T\mathbf{C}(d)^{-1}(\mathbf{M}\tilde{\mathbf{x}} - \mathbf{d}) \\ & + (\tilde{\mathbf{x}} - \mathbf{x_0})^T\mathbf{C}(x_0)^{-1}(\tilde{\mathbf{x}} - \mathbf{x_0})] \end{aligned} \tag{5}$$

and that the Hessian matrix $\mathbf{H}(\tilde{x})$ of J, i.e. the matrix composed of its second partial derivatives $\frac{\partial^2 J}{\partial x_i \partial x_j}$, is constant and given by

$$\mathbf{C}(x)^{-1} = \mathbf{H}(\tilde{x}) \quad . \tag{6}$$

If the model (denoted by $\mathbf{M}(\tilde{\mathbf{x}})$) is not linear or any of the PDFs of the inputs are non Gaussian, an approximation of the posterior PDF is constructed via the minimum of the cost function J, which is

$$\begin{aligned} J(\tilde{\mathbf{x}}) = \quad & \tfrac{1}{2} \quad [(\mathbf{M}(\tilde{\mathbf{x}}) - \mathbf{d})^T\mathbf{C}(d)^{-1}(\mathbf{M}(\tilde{\mathbf{x}}) - \mathbf{d}) \\ & + (\tilde{\mathbf{x}} - \mathbf{x_0})^T\mathbf{C}(x_0)^{-1}(\tilde{\mathbf{x}} - \mathbf{x_0})] \end{aligned} \tag{7}$$

The posterior PDF is now characterised by an estimated mean given by the minimum of J, and an uncertainty given by the covariance matrix $\mathbf{C}$ calculated via Eq. (6) is .

Any variational data assimilation system, be it in operational numerical weather prediction or oceanography, is based on Eq. (7). The optimisation mode of CCDAS uses an iterative procedure to minimise the cost function of Eq. (7), which yields $\mathbf{x}$, and computes $\mathbf{C}(x)$ via Eq. (6).

As long as the uncertainties in the individual data streams are independent, the contribution of each of them to the right hand side of Eq. (7), and, hence, to Eq. (6), can be quantified by a separate term in the sum. In this formalism the contribution of a synthetic data set is to be handled as follows:

- The mean value is generated with the model itself, i.e. the equation $\mathbf{d} = \mathbf{M}(\mathbf{x})$ is applied.
- The covariance of uncertainties is specified such as to reflect the expected characteristics of the observational products generated by the instrument.

Consequently, the minimum of the cost function of Eq. (7) is independent of the synthetic data: Two experiments, one with and one without synthetic data, will yield the same $\mathbf{x}$. Eq. (6), however, takes full account of the specified uncertainty in the synthetic data. The effect of the synthetic data will show up as a reduction of posterior parameter uncertainty.

For the second step (cf. Figure 1), i.e. the estimation of a diagnostic or prognostic target quantity y, its PDF can be approximated by a Gaussian with mean

$$y = \mathbf{N}(\mathbf{x}) \tag{8}$$

and covariance

$$\mathbf{C}(y) = \mathbf{D}(\mathbf{N})\mathbf{C}(x)\mathbf{D}(\mathbf{N})^T + \mathbf{C}(y_{mod}) \quad , \tag{9}$$

where $\mathbf{N}$ is the model (in Figure 1 denoted by diagnostic/prognostic model) which maps the control variables onto the target quantity, $\mathbf{D}(\mathbf{N})$ is its linearisation around the mean of the posterior PDF of the control variables, also denoted by the Jacobian matrix of $\mathbf{N}$, and $\mathbf{C}(y_{mod})$ is the uncertainty in the model result resulting from errors in the model. Only if y coincides with one of the observations used in the inversion step, this uncertainty is already accounted for in $\mathbf{C}(x)$, and we omit the $\mathbf{C}(y_{mod})$ contribution. If $\mathbf{N}$ is linear and the posterior PDF of the control variables Gaussian, then the PDF of the target quantity is Gaussian as well and completely described by Eq. (8) and Eq. (9).

Following the formalisms developed above, it can be seen that the essential information about the value of the synthetic data that are used to evaluate new data streams is contained in the covariance matrix of uncertainties. Mean values do not enter Eq. (6). Once the synthetic data can be replaced by real data, the only modification is that this resulting additional data term may shift $\mathbf{x}$, and thus modify the linearisation $\mathbf{D}(\mathbf{N})$ and the curvature of J, i.e. the Hessian, in Eq. (6).

3. ASSIMILATION OF MERIS MEDIUM RESOLUTION PRODUCT OVER SIX SITES

We assimilate FAPAR data from the MERIS medium resolution product at the sites listed in Table 1. The satellite scenes with a resolution of approximately 1.2 km encompass 15 by 15 pixels, i.e. have a scene size of around 18 by 18 km^2. Approximate plant functional type (PFT) distributions were estimated for the scenes based on information from the sites combined with images obtained from Google Earth. These estimates were used as priors within CCDAS.

The study's base experiment consists of the assimilation of the MERIS reduced-resolution FAPAR product. The assimilation procedure minimises Eq. (5). To specify the data uncertainty, $\mathbf{C_d}$, we use an uncorrelated data uncertainty of 0.1 representing the combined effect of observational error and error in the terrestrial biosphere model. Prior values of process parameters and their uncertainties are taken from [10].

Table 1. List of sites for assimilation.

Site	Latitude	Longitude	Description
Sodankylä (Finland)	67° 21' N	26° 38' E	Boreal evergreen forest
Zotino (Russia)	60° 45' N	89° 38' E	Boreal mixed forest
Loobos (Netherlands)	52° 10' N	5° 44' E	Temperate pine forest
Hainich (Germany)	51° 06' N	10° 54' E	Temperate deciduous forest
Manaus (Brazil)	2° 35' S	60° 06' W	Tropical rainforest, largely undisturbed
Maun (Botswana)	19° 54' S	23° 33' E	Tropical savanna

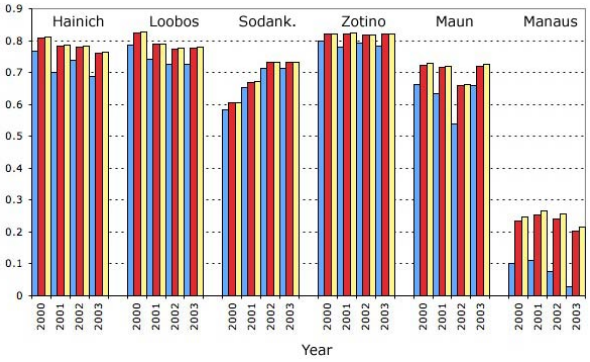

Figure 2. Uncertainty reduction for diagnostic NPP for 6 sites with MERIS FAPAR (blue), MERIS FAPAR + CO_2 fluxes (red), and MERIS FAPAR + CO_2 and water fluxes (yellow).

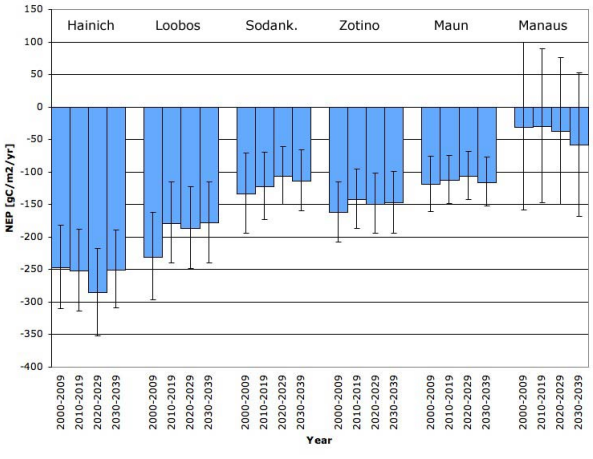

Figure 3. Prognostic NEP in gram carbon per m^2 and year (negative values are terrestrial uptake) including uncertainties for 6 sites with MERIS FAPAR only.

During the minimisation procedure, the cost function reduces to less than 10% of its prior value. The calibrated model, i.e. the model with the optimised parameters, can now be used to compute biosphere atmosphere exchange fluxes.

Applying Eq. (9) to the posterior parameter uncertainties yields diagnostic net primary productivity (NPP) for the years 2000 to 2003 on an annual basis (Figure 2), and prognostic net ecosystem productivity (NEP) from 2000 to 2039, averaged for each decade (Figure 3).

The procedure just described is applied to the beforementioned base case, in which only MERIS FAPAR data are assimilated, and to two more cases with additional data provided by eddy covariance measurements of CO_2 and water vapour fluxes. Figure 2 shows that for all sites except Manaus, most of the uncertainty reduction is already achieved with FAPAR data, information that is available from space continuously and over large areas. By contrast, eddy covariance data is only available locally at specific sites.

4. POTENTIAL SENSOR EVALUATION

This section describes two experiments used to evaluate different sensor concepts. The first sensor concept is that of a synthetic sensor with higher spatial resolution than the base case (see Section 3). The second is a synthetic sensor with ideal resolution, meaning that the spatial resolution is high enough to separate out and identify each vegetation type (PFT) within the scene.

Since such sensors do not exist, we follow the methodology described in Section 2 for the evaluation of synthetic data. As pseudo-data we use the MERIS FAPAR data set from our base case, while changing the data uncertainty given by $\mathbf{C_d}$. We characterise the sensor with higher resolution by a reduction of the data uncertainty for FAPAR, which is assumed to go from 0.1 to 0.05. For the sensor with ideal resolution, we additionally assume that it provides the information necessary to fully constrain the relative fractions of plant functional types (PFT) within the scene. These PFT fractions therefore do not need to be optimised, nor do they longer enter the calculation of

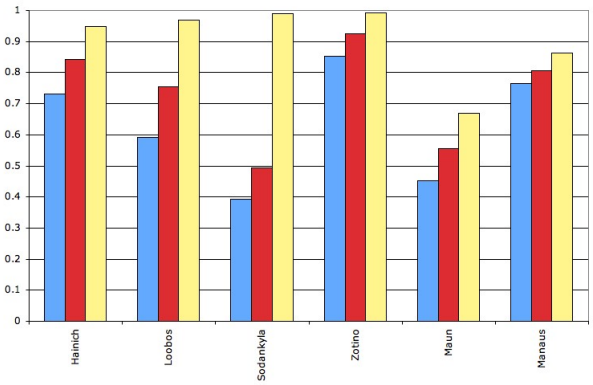

Figure 4. Reduction in uncertainty in diagnostic mean NPP from 2000-2003 for the base case (blue), the higher resolution sensor (red) and the ideal resolution sensor (yellow).

uncertainties. We reflect this in our inversion setup by excluding the PFT fractions from the list of unknown parameters. The constraint provided by the FAPAR observations can then act in a reduced parameter space and, thus, achieve a better reduction of uncertainty. From the parameter uncertainties, we compute uncertainty reductions for diagnostic mean NPP from 2000 to 2003 for the base case (blue), the higher resolution sensor (red) and the ideal resolution sensor (yellow) (see Figure 4).

5. CONCLUSIONS AND OUTLOOK

We have extended CCDAS in such a way that it is capable of performing simulations for a wide range of scenes at the lcoal scale. First, we have included the assimilation and uncertainty propagation of FAPAR data into the CCDAS core. We have then included the assimilation of synthetic (pseudo-)fluxes of CO_2 and latent heat (water vapour) at the site scale.

The extended system has been used to simultaneously assimilate the 1.2 km MERIS FAPAR product over six sites. This observational constraint yields a remarkable reduction in the uncertainty in process parameters. Projecting this uncertainty to CO_2 fluxes to the atmosphere, the observational constraint achieves an uncertainty reduction of up to 80%. The uncertainty reduction is not much improved by adding monthly averaged eddy flux observations of the carbon and water balance to the assimilation data stream.

We have further demonstrated the concept of potential sensor evaluation through data assimilation systems using, as examples, two hypothetical sensors: one with higher resolution, and one with an ideal resolution that makes it capable of unequivocally identifying the distribution of plant functional types.

The next step will be to include a global version of BETHY into the data assimilation system, which will allow extending the range of observations to be assimilated to two major data sets: The GLOBALVIEW [16] database of atmospheric CO_2 concentrations, and the MERIS FAPAR product at the full global scale. This will further allow to demonstrate the effect of the simultaneous assimilation of both of these global data sets in terms of ucnertainty reduction in target quantities of interest. Furthermore, this study will allow the evaluation of global-coverage synthetic sensor concepts as demonstrated for the local scale in Section 4. The methodology developed in this study is implemented within an Interactive Uncertainty Evaluation Tool for easy and rapid evaluation of new sensor concepts targeting the carbon cycle.

In conclusion, data assimilation offers exciting prospects for semi-automated sensor design evaluation and OSSE procedures for designated target quantities that may be of interest to certain costumers of EO products. The present study applies this concept to the broad field of land surface processes, and could as such be extended with ease to other sensors providing information on the state of the land surface. Ideal candidates are land surface skin temperature products from (A)ATSR, and skin layer soil moisture from SMOS.

ACKNOWLEDGMENTS

The study RS-CCDAS is supported by the European Space Agency under ESRIN/Contract No. 20595/07/I-EC. MERIS data over the sites have been downloaded through the MERCI system. The authors thank Monica Robustelli from JRC as well as Olivier Colin and Philippe Goryl of ESRIN for their support.

REFERENCES

[1] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, KB Averyt, M. Tignor, and HL Miller. IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007.

[2] P. D. Cox, 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.

[3] I. C. Prentice, G. D. Farquhar, M. J. R. Fasham, M. L. Goulden, M. Heimann, V. J. Jaramillo, H. S. Kheshgi, C. Le Quéré, R. J. Scholes, and D. W. R. Wallace. The carbon cycle and atmospheric carbon dioxide. In J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, editors, *Climate Change 2001: The Scientific basis*, chapter 3, pages 183–237. Cambridge University Press, Cambridge, U.K., 2001.

[4] P. Friedlingstein, P. Cox, R. Betts, L. Bopp, W. von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, G. Bala, J. John, C. Jones, F. Joos, T. Kato, M. Kawamiya, W. Knorr, K. Lindsay, H. D. Matthews, T. Raddatz, P. Rayner, C. Reick, E. Roeckner, K.-G. Schnitzler, R. Schnur, K. Strassmann, A. J. Weaver, C. Yoshikawa, and N. Zeng. Climate-Carbon Cycle Feedback Analysis: Results from the C^4MIP Model Intercomparison. *Journal of Climate*, 19:3337–3353, 2006.

[5] N. Gobron, B. Pinty, O. Aussedat, M. Taberner, O. Faber, F. Melin, T. Lavergne, M. Robustelli, and P. Snoeij. Uncertainty estimates for the FAPAR operational products derived from MERIS - Impact of top-of-atmosphere radiance uncertainties and validation with field data. *Remote Sensing of Environment*, 112:1871–1883, 2008.

[6] N. Gobron, B. Pinty, and MM Verstraete. Theoretical limits to the estimation of the leaf area index on thebasis of visible and near-infrared remote sensing data. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(6):1438–1445, 1997.

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

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

[9] P. Rayner, 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). *Global Biogeochemical Cycles*, 19:doi:10.1029/2004GB002254, 2005.

[10] M. Scholze, T. Kaminski, P. Rayner, W. Knorr, and R. Giering. Propagating uncertainty through prognostic CCDAS simulations. *J. Geophys. Res.*, 112:doi:10.1029//2007JD008642, 2007.

[11] W. Knorr, N. Gobron, M. Scholze, T. Kaminski, and B. Pinty. Global-Scale Drought Caused Atmospheric CO_2 Increase. *EOS, Transactions*, 86(18):178 & 181, 2005.

[12] W. Knorr and J. Kattge. Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using monte carlo sampling. *Global Change Biology*, 11:1333–1351, 2005.

[13] T. Kaminski and P. J. Rayner. Assimilation and network design. In H. Dolman, A. Freibauer, and R. Valentini, editors, *to appear in Observing the continental scale Greenhouse Gas Balance of Europe*, Ecological Studies, chapter 3. Springer-Verlag, New York, 2008.

[14] A. Tarantola. *Inverse Problem Theory - Methods for Data Fitting and Model Parameter Estimation*. Elsevier Sci., New York, 1987.

[15] I. G. Enting. *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press, 2002.

[16] GLOBALVIEW-CO_2. Cooperative Atmospheric Data Integration Project - Carbon Dioxide. CD-ROM, NOAA CMDL, Boulder, Colorado, 2004. [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].